Standard Setting Technical Report

Kaiapuni Assessment of Educational Outcomes (KĀʻEO)

Science

Grade 4

September 5, 2017

Prepared by Karla L. Egan, Ph.D.

EDMETRIC LLC

EXECUTIVE SUMMARY

A Bookmark Standard Setting Procedure (BSSP) was conducted in Honolulu, Hawaii on July 17-19, 2017. Dr. Karla Egan, EdMetric, LLC, designed the standard setting workshop collaboratively with Dr. Pohai Kukea Shultz, from the University of Hawaii at Manoa. Panelists engaged in content-based discussions to recommend three cut scores that separated four achievement levels – Beginning (Level 1), Developing (Level 2), Proficient (Level 3), and Distinguished (Level 4).

Staff from the University of Hawaii at Manoa helped answer content-related questions during the workshop. Dr. Egan answered all process-related questions for the workshop.

Staff from the University of Hawaii at Manoa recruited ten panelists to recommend cut scores in Grade 4 science. Panelists completed three days of work to arrive at recommended cut scores for the KĀʻEO science assessment. Day 1 of the standard setting event included an overview of the Hawaiian Immersion Assessment Project and the KĀʻEO science assessment, a discussion of policy definitions and range achievement level descriptors (ALDs), drafting of threshold ALDs, detailed examination of the operational test form, and an introduction to the BSSP.

On Day 2, panelists studied the Ordered Item Booklet (OIB) and completed the item map, participated in Bookmark training, and engaged in two rounds of Bookmark placement and discussion. On the final day of the standard setting event, panelists reviewed Round 2 impact data, participated in a large group discussion, set their final Bookmarks, and reviewed the resulting impact data. Panelists were asked, based upon the final impact data, if they would like to complete a fourth round of Bookmark placements; however, the group decided against another round of judgments. Panelists finished the workshop by creating draft reporting ALDs.

Table ES.1 shows the final recommended cut scores from the standard setting.

TABLE ES.1: Final Recommended Cut Scores for Grade 4 KĀʻEO Science Assessment

Achievement Level	Recommended Cut Score
Developing	480
Proficient	519
Distinguished	562

Table ES.2 shows the impact data associated with the final recommended cut scores for the Grade 4 KĀʻEO science assessment, along with the distribution of student scores on the Grade 4 Hawaii State Assessment (HSA) in science, which was administered to the general

Hawaiian student population in 2017. The final row of the table shows the percent of students scoring at or above Proficient for each test.

TABLE ES.2: Impact Data for Grade 4 KĀʻEO Science Assessment and 2017 Grade 4 HSA Science Assessment

Achievement Level	KĀʻEO Science - Percent of Students
Beginning (Level 1)	35.50%
Developing (Level 2)	27.40%
Proficient (Level 3)	28.10%
Distinguished (Level 4)	9%
Combined Proficient & Distinguished (Level 3 & Level 4)	37.10%

In general, panelists' evaluations of the workshop were positive, with panelists unanimously either agreeing or strongly agreeing with all evaluation statements. Panelists 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:

"Mahalo to Karla, Pōhai, Pono and all of the other limahana for all that you folks did to help make this workshop successful!"

"Mahalo for helping us through this process!"

"Mahalo piha i kā ʻoukou hana nui e hoʻomākaukau i nā mea e pono ai!"

"Mahalo nui!"

"The work is huge - but my understanding of the process is becoming clearer. I appreciate the time and effort put into having the information available to all in attendance. When the ALDs are finalized, how can we access a copy?"

"This was very valuable :)"

"Please ask me to come back. This was fun!"

"Mahalo nui no ka ho'onui 'ike.....a me ka mea'ai 'ono loa kekahi!"

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.

CONTENTS

CHAPTER 1. INTRODUCTION

In compliance with the federal Every Student Succeeds Act (ESSA) (and previously the reauthorization of the Elementary and Secondary Education Act – or ESEA – known as No Child Left Behind), HIDOE annually administers statewide assessments to students in grades 3 through 8 and high school in English language arts/literacy (ELA/L), mathematics, and science. All students, with the exception of students with the most significant cognitive disabilities, are required to participate in statewide assessments. The complete assessment program includes the Smarter Balanced Assessment Consortium's summative assessments in ELA/L and mathematics, and state developed science assessments - the Hawaii State Assessments (HSA) in science. The HSA science assessments are administered to students in Grades 4 and 8, and a Biology End-of-Course assessment is administered to high school students.

In 2015, Hawaii began an expansion of its statewide assessment program to address the needs of 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 Hawaii Department of Education (HIDOE), the Hawaiian Language Immersion Project, and the University of Hawaii at Manoa's College of Education have worked collaboratively to develop assessments that are aligned to the standards in place in Kaiapuni schools. The assessments are presented entirely in Hawaiian language, in the same way that instruction is delivered to students in grades 3 through 5. These assessments, the Kaiapuni Assessments of Educational Outcomes (KĀʻEO), were administered operationally in Hawaiian Language Arts (HLA) and mathematics for the first time in Spring 2016 to students in grades 3 and 4. An operational science assessment for students in Grade 4 was added to the KĀʻEO suite in Spring 2017.

Using the operational data from the Spring 2017 test administration, a Bookmark Standard Setting Procedure (BSSP) was held at the University of Hawaii at Manoa (UHM) from July 17-19, 2017. Through the BSSP, educators from Kaiapuni schools recommended three cut scores that resulted in four achievement levels: Beginning (Level 1), Developing (Level 2), Proficient (Level 3), and Distinguished (Level 4). An achievement level, along with specific descriptions of the knowledge, skills, and processes a student at that performance level demonstrates, will be reported for each individual student. In addition to providing information regarding individual students' performance in science, KĀʻEO assessment results will be aggregated and will provide the basis for each Kaiapuni school's Strive HI scores.

The KĀʻEO science standard setting is particularly notable because it marks the first time performance levels and content associated with statewide science 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 the KĀʻEO Grade 4 science assessment. The first chapter of the report provides background and organizational information. The second chapter of the report describes all steps leading up to the standard setting event, including the development of the standard setting design; identification of participants in the standard setting event; preparation of materials for standard setting; and selection and pre-training of table facilitators. The third 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 fourth chapter of the report provides a description and summary of results of participant evaluations completed during the standard setting event. The final chapter of the report addresses the contribution of standard setting to the overall validity argument for the KĀʻEO science 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/L – English language arts/literacy
ELL – English Language Learner
ESSA – Every Student Succeeds Act
HSA – Hawaii State Assessment
HIDOE – Hawaii Department of Education
IRT – Item Response Theory
KĀʻEO – Kaiapuni Assessment of Educational Outcomes
HLA – Hawaiian Language Arts
NDA – Non-Disclosure Agreement
OIB – Ordered Item Booklet
SEM – Standard Error of Measurement
SWD – Students with Disabilities
UHM – University of Hawaii at Manoa
USED – United States Department of Education

RELEVANT DEFINITIONS

The following definitions will apply throughout this report:

Achievement Level Descriptor (ALD) – the knowledge, skills and processes 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 processes 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 State Assessment Program – Hawaii's statewide system of assessments used to measure student progress toward the state's academic content standards.

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

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.

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.

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

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 panelists for the workshop.

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 (Seneca Consulting) and Dr. Pohai Kukea Shultz.

The standard setting plan provided a detailed implementation of the Bookmark Standard Setting Procedure (BSSP) for the KĀʻEO science assessment. The BSSP was recommended based on the technical characteristics of the KĀʻEO and its intended uses. The KĀʻEO science assessment includes mixed item types and relies upon item-response theory models for scoring; therefore, it is well-suited to the BSSP. This was also the standard setting methodology used to determine cut points for the KĀʻEO Hawaiian language arts (HLA) and mathematics assessments. In the proposed design, panelists would examine the Kaiapuni content standards for science and their connection to the KĀʻEO science assessment items to establish three cut points distinguishing among four levels of performance: Beginning (Level 1), Developing (Level 2), Proficient (Level 3), and Distinguished (Level 4). In addition to recommending achievement level cutpoints for the KĀʻEO science assessment, panelists would also be tasked with reviewing range achievement level descriptors (ALDs), and creating threshold and reporting ALDs.

The standard setting plan for the KĀʻEO called for empaneling a group of 12 content experts to complete three rounds of judgments pertaining to assessment items and content. The group would be divided into two tables for the purposes of making judgments and participating in small group discussions. Table level discussions would be facilitated by a designated table leader. To maintain continuity with the KĀʻEO HLA and mathematics assessments, panelists would rely upon the policy definitions developed prior to the KĀʻEO HLA and mathematics standard setting workshop in July 2016. The standard setting plan may be found in Appendix A.

DEVELOPMENT OF POLICY DEFINITIONS

The KĀʻEO development team developed policy definitions prior to setting standards for the KĀʻEO HLA and mathematics assessment. These same policy definitions provided the foundation for KĀʻEO science standard setting. They provide a high-level definition of student performance at each level used for state and federal reporting purposes. KĀʻEO policy definitions are shown in Table 2.1.

Table 2.1 Policy Level Definitions

Achievement Level	Policy Level Definitions

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.

PANELIST SELECTION

UHM staff selected 10 panelists to participate in the KĀʻEO science standard setting workshop. The KĀʻEO assessment was developed specifically for students in Kaiapuni schools, so panelists were selected from the small pool of teachers and content experts currently working in Kaiapuni schools. Because the pool of Kaiapuni educators 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. Although the standard setting plan called for selection of 12 panelists, due to limited resources in the Kaiapuni schools, only 10 panelists were confirmed; however, the number of panelists was deemed sufficient to provide reliable data for the standard setting process.

All panelists were classroom teachers, with the exception of one, who was a Curriculum Coordinator and Data Specialist. All panelists were familiar with the needs of Hawaiian language learners, and four had expertise working with English language learners (ELLs). One panelist had expertise working with students with disabilities (SWDs), and one panelist had experience with Title I Reading. Table 2.2 shows the geographic representation and the Kaiapuni school representation across the panelists.

Table 2.2. Panelists by Island and School

Island and School	Number of Panelists
Hawaii	**1**
Ke kula ʻo ʻEhunuikaimalino	1
Kauai	**1**
Kawaikini New Century Public Charter School	1
Maui	**2**
Ke Kula Kaiapuni ʻo Pāʻia	1
Kula Kaiapuni ʻo Nāhiʻenaʻena	1
Molokai	**1**
Ke Kula Kaiapuni ʻo Kualapuʻu	1
Oahu	**5**
Ke Kula Kaiapuni ʻo Waiau	1
Kula Kaiapuni ʻo Ānuenue	2
Ke Kula Kaiapuni o Pūʻōhala	1
Ke Kula Kaiapuni o Hauʻula	1
Total	**10**

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

Table 2.3 shows the distribution of panelists at each table by school, island, and position/area of expertise.

Table 2.3. Distribution of Panelists Per Table by School, Island, and Grade Level(s)

School	Island	Position/Expertise
Table 1		
Ke Kula Kaiapuni ʻ o Ānuenue	Oʻ ahu	Curriculum Coordinator and Data Specialist
Ke Kula Kaiapuni ʻo Pāʻia	Maui	4
Ke Kula Kaiapuni ʻo Pūʻōhala	Oʻ ahu	4
Ke Kula ʻo ʻEhunuikaimalino	Hawaiʻ i	3
Ke Kula Kaiapuni ʻ o Hauʻ ula	Oʻ ahu	5-6
Table 2		
Kawaikini New Century Public Charter School	Kauaʻ i	4
Ke Kula Kaiapuni ʻ o Ānuenue	Oʻ ahu	3
Ke Kula Kaiapuni ʻ o Nāhīʻ enaʻ ena	Maui	1
Ke Kula Kaiapuni ʻ o Waiau	Oʻ ahu	4
Ke Kula Kaiapuni ʻo Kualapuʻu	Molokaʻ i	K-2

After panelists were divided into two table groups, a table facilitator for each table was identified. Additionally, all table facilitators were classroom teachers or curriculum coordinators who had participated in the KĀʻEO development process as item writers, standards developers, standard setting panelists and/or alignment study panelists.

PREPARATION OF MATERIALS

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

Presentations. The Opening Session presentation overviewed KĀʻEO science development, discussed the standard setting process, and provided a description of the different types of achievement level descriptors (ALDs). Bookmark Training presentations included three parts – 1) an overview of the BSSP, including how to study the OIB and complete the item map, and 2) a detailed description of how to place a Bookmark, and 3) a discussion of how to create reporting ALDs. Appendix B includes the presentation slide decks.

Achievement Level Descriptor Review Guidelines. A set of guidelines was created to guide panelists in their review and refinement of ALDs. This document is included in Appendix B.

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

Surveys and Evaluations. EdMetric developed readiness surveys to administer following the opening session and after Bookmark training. A final workshop evaluation was administered at the end of the workshop. These surveys are presented in Appendix B.

OIB. EdMetric prepared the OIB for the assessment. Since the KĀʻEO science assessment is a single fixed form, it was reasonable to include all items from the assessment in the OIB. Within the 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 based on a .50 response probability. The item location was estimated using Spring 2017 operational data. The OIB was created in hard-copy format. Stimuli for the items were printed in a separate stimulus booklet.

Item Maps. EdMetric also created an item map based on the OIB. The item map 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 map 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?" The item map was prepared in electronic format as an Excel spreadsheet.

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

Item Map - Science Group T

Papa 4

Order of Difficulty (OIB Page Number)	Location	Score Point	Test Item #	Problem ID	Stimulus	Code	Standard/ 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	425	1	4	289	Ke Anilā	1.2	4.LL.4	FTG		N/A
2	428	1	10	284	Nā Wai ʻEhā	2.1	4.HL.1	SC		
3	430	1	23	314	Kumulipo	2.4	4.HL.11	SC		
4	445	1 of 3	32	324	Nā Papa O Ke Kai	2.4	4.HL.10	ER		
5	449	1	37	293	Ola Kino	2.3	4.HL.7	FTG		
6	452	1	1	305	Papahānaumoku	2.1	4.HL.2	STV		
7	452	1 of 2	45	318	Wai A Kāne	2.2	4.HL.5	ER		

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 rating associated with Level 2 was lower than the rating for Level 3, etc. Ratings were also flagged if an entry was left blank. Workshop facilitators monitored panelist ratings in real time. If the ratings were flagged for possible incorrect entry or incomplete entry, then the workshop facilitator asked the panelist to confirm or correct the entry. Panelists also completed all surveys and evaluations using the online tool.

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

Figure 2.2 shows the Control Panel.

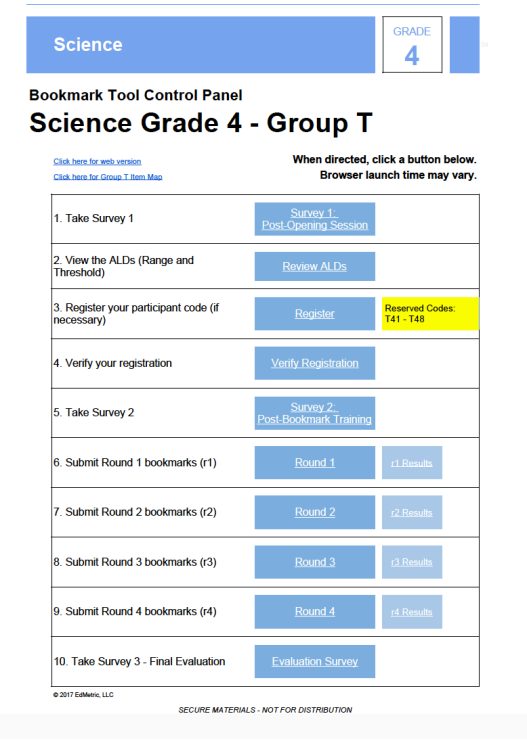


Figure 2.2. Standard Setting Control Panel PDF

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 were automatically inputted 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

Although the table facilitators were not new to the standard setting process and somewhat familiar with their roles, Dr. Egan developed a two-part Table Facilitator Training presentation which all Table Facilitators accessed via recorded webinar prior to the beginning of the standard setting event. Part 1 of Table Facilitator Training discussed the different types of ALDs and their role in the standard setting process. Part 2 overviewed the BSSP, presented the agenda for the standard setting event, and discussed the role of the Table Facilitator in guiding panelists through the process of setting Bookmarks. 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 17-19, 2017, at the University of Hawaii at Manoa. As shown in the High-Level Agenda (Appendix B), panelists arrived at 9:00 a.m. on July 17th, and completed three days of work to arrive at recommended cut scores for the KĀʻEO science assessment. 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, an overview of the Kaiapuni science content standards, discussion and revision of the range ALDs, development of threshold ALDs, and a detailed examination of the operational test form. On Day 2, panelists studied the OIB, completed the item map, completed Bookmark training, and engaged in two rounds of ratings. On the final day of the standard setting event, panelists completed a third round of ratings, reviewed final cut scores and impact data, and drafted reporting ALDs.

FACILITIES AND SECURITY OF MATERIALS

The KĀʻEO standard setting event was held at the University of Hawaii at Manoa. One large room with two tables was reserved for the event. The two tables were 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. 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, content facilitator, table facilitators, 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, ensuring that table facilitators and panelists follow the agenda; and attending to other needs as they arise.

Content Facilitator. Pono Fernandez, KĀʻEO Project Director, 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 OIB, discussing impact data, recommending final cut scores for the assessment, and drafting threshold and reporting ALDs.

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.
Content Facilitator	Pono Fernandez	Ms. Fernandez has a Master of Arts degree in Hawaiian language and is a fluent Native Hawaiian speaker. She has led many of the KĀʻEO content development efforts.

TABLE FACILITATOR TRAINING

Table facilitators' training occurred the week prior to the standard setting event via a two-part recorded webinar (see Appendix C for the slide presentation). Part 1 of the webinar focused on the inter-related system of ALDs that frame the KĀʻEO science assessment, the specific stakeholder audiences that use them, and their role in the standard-setting process. Table facilitators were instructed that they would be provided with policy definitions and range ALDs that were developed prior to the workshop, and that they, along with the other panelists, would develop threshold ALDs and reporting ALDs during the workshop.

During Part 2 of the 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 facilitators 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 map. 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. Table facilitators were provided email contact information if they had questions prior to the workshop.

STANDARD SETTING EVENT DAY 1

Table Facilitators' Meeting

Prior to the opening session on Day 1 of the standard setting event, Dr. Egan met with table facilitators 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. Dr. Pohai Kukea Shultz, Principal Investigator of the Hawaiian Immersion Assessment Program, welcomed participants and provided an overview of the developmental history of the KĀʻEO to provide context for those who had not been involved in the test development process. Participants were provided with and required to sign non-disclosure agreements before they could participate in the standard setting meeting.

Standard Setting Orientation And ALD Training

Following Dr. Kukea Shultz's presentation, Dr. Egan provided an orientation to the standard setting process and how it would be implemented during the three-day workshop. The slide deck "Training 1 KĀʻEO Opening Session" (Appendix B) accompanied this training. Training included a detailed discussion of ALDs, describing the three different types of ALDs – policy, range, threshold, and reporting – their intended audiences, and the relationships among them. This portion of the training established the ALDs as the framework for standard setting and prepared panelists to examine policy definitions, review and revise range ALDs, and create threshold and reporting ALDs.

Panelists were instructed that their first task would be to study the range ALDs with the goal of creating threshold ALDs for Developing (Level 2) and Proficient (Level 3). These threshold ALDs describe the knowledge, skills, and processes a student must demonstrate in order to enter each achievement level. Specifically, panelists were instructed to study the range ALDs for Beginning and Developing to determine the threshold ALDs for Developing, and then to study the range ALDs for Developing and Proficient to determine the threshold ALDs for Proficient.

Panelists were given an opportunity to ask questions before proceeding to a readiness survey.

Readiness Survey

All panelists completed a readiness survey immediately following standard setting orientation and ALD training. The purpose of the survey was to assess panelists' perceived understanding of the purpose of the standard-setting workshop and the purpose of the ALDs. Panelists completed and submitted the survey online. Workshop facilitators had real-time access to survey results, enabling them to address questions individually as necessary. Table 3.2 shows the results of the readiness survey. 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

Readiness Survey Question	% who Agree or Strongly Agree (n=10)
1. The orientation session provided a clear overview of the standard setting process	100%
2. I understand the goals of the standard setting workshop.	100%
3. I understand my role in the standard setting workshop.	100%
4. I understand the purpose of the Range ALDs	100%
5. I understand the purpose of the Threshold ALDs.	100%
6. The training on achievement level descriptors was helpful to me.	100%
7. I understand the steps necessary to begin working on Threshold ALDs.	100%
8. I understand that I will receive additional training throughout the workshop.	100%
The percentage of panelists who answered "yes" is reported for the following questions:	
9. Before I begin work on the Range & Threshold ALDs, I would like additional training on achievement level descriptors.	0.0%
10. I have additional questions on materials presented during the opening session that I would like answered before I begin the next task.	0.0%

Discussion of Range ALDs, Threshold ALDs, and Content Standards

Following completion of the readiness survey, Ms. Pono Fernandez led panelists through a discussion of the range and threshold ALDs. The group recommended minor revisions to the range ALDs. Then, the group created threshold ALDs as a subset of the range ALDs. Per Dr. Egan's training, they began by reviewing the range ALDs for Beginning and Developing to determine reasonable threshold ALDs for Developing. Then, they reviewed the range ALDs for Developing and Proficient to determine reasonable threshold ALDs for Proficient. These threshold ALDs would provide the foundation for their discussion of the threshold or target student in the Bookmark placement process. Ms. Fernandez closed this portion of the workshop with a question and answer session regarding the Kaiapuni science content standards.

Taking the Operational Assessment

Following the discussion of the content standards, participants took the operational KĀʻEO science assessment online, replicating the administration of the assessment to a student. Panelists used their personal laptops to access the assessments, using credentials provided by the workshop facilitators and UHM staff.

Round 1

Bookmark Training

After completing the operational assessment, participants worked in their small groups to begin Round 1 activities. Dr. Egan provided an overview of the BSSP to all participants using the slide deck "Training 2 Bookmark Overview" (Appendix B). She explained the Bookmark materials, including the OIB, the online Item Map, and the scoring rubrics. She discussed the process of studying each item in the OIB, instructing panelists to answer the questions: 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; and 2) What makes this item more difficult than preceding items. She also explained the process of completing the item map. Ms. Fernandez led the group through a discussion of the first five items in the OIB, demonstrating how each item would be discussed and how the item map would be completed.

Following the training activity and discussion, the group adjourned for the day.

STANDARD SETTING EVENT DAY 2

Round 1 (continued)

Study of OIB and Completion of Item Map

Panelists reconvened at 9:00 a.m. on Day 2 of the standard setting event to continue Round 1 activities. Dr. Egan opened the session by welcoming the group and asking if anyone had questions before getting started. No questions were asked, and the group moved forward with studying the OIB and completing item maps according to training provided on Day 1. Table facilitators assigned numbered packets of secure materials including the OIB,

stimulus booklet, and item map to each panelist. Additionally, table facilitators identified a scribe to take notes for the group during table discussions.

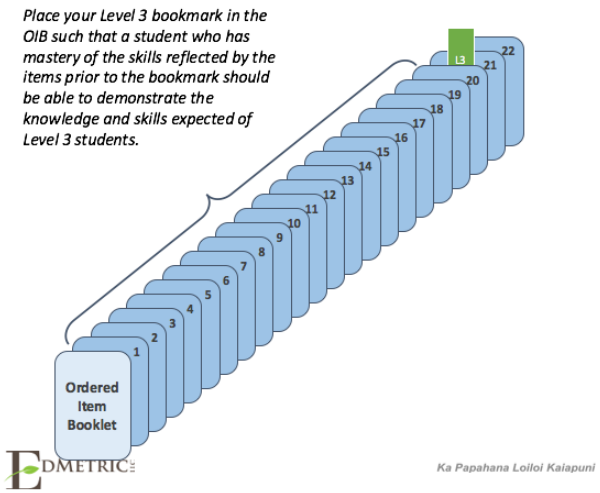
Table facilitators then facilitated a discussion of the OIB, beginning with the first (easiest) item in the OIB and progressing through the last (most difficult). Table facilitators led the panelists at the table through a discussion of two questions for each item:

- 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 via 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.

When both small groups had finished studying the OIB and completing the item map, Dr. Egan trained the group on the meaning of the Bookmark placement and how to place a Bookmark using the slide deck "Training 3 KĀʻEO Bookmark Training" (Appendix B). This training focused on the connection between the threshold ALDs developed on Day 1, the definition of the "target student," and how panelists would use the Control Panel to place their Bookmarks. Dr. Egan illustrated the meaning of "setting a Bookmark" as shown 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). Finally, Dr. Egan demonstrated the process by which panelists would electronically "set" their Bookmarks. She directed them to their Table Links on Google Drive and reminded them that Bookmark placement is an individual activity.

Readiness Survey

Immediately following the Bookmark Training activity, and before placing their first Bookmarks, all panelists completed an online readiness survey. The purpose of the survey was to determine panelists' readiness for making their first round of Bookmark placements and participating in a discussion of the results. Again, workshop facilitators were able to view survey results in real-time, enabling them to address panelists' concerns or questions. Table 3.3 shows the results of the readiness survey.

Table 3.3. Results from Bookmark Readiness Survey

Readiness Survey Question	% who Agree or Strongly Agree (n=10)
1. I participated in bookmark training and had an opportunity to as questions and discuss the meaning of the bookmarks.	100%
2. I understood how to study items in the ordered item booklet.	100%
3. I understand how to place my bookmarks.	100%
4. I understand I will have opportunities to change my bookmarks in Round 2.	100%
The percentage of panelists who answered "yes" is reported for the following questions:	
5. I would like additional training on placing my bookmarks for Round 1.	10.0%
6. I have additional questions that I would like to ask before placing my Round 1 bookmarks.	0%

The results of the readiness survey indicated that all panelists understood how to study the OIB and how to place their Bookmarks. One panelists indicated a need for additional training prior to placing a Round 1 Bookmark. Workshop facilitators provided additional guidance to this panelist prior to moving forward with Bookmark placement.

Bookmark Placement

After completing the readiness survey, all panelists set their Round 1 Bookmarks. 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.

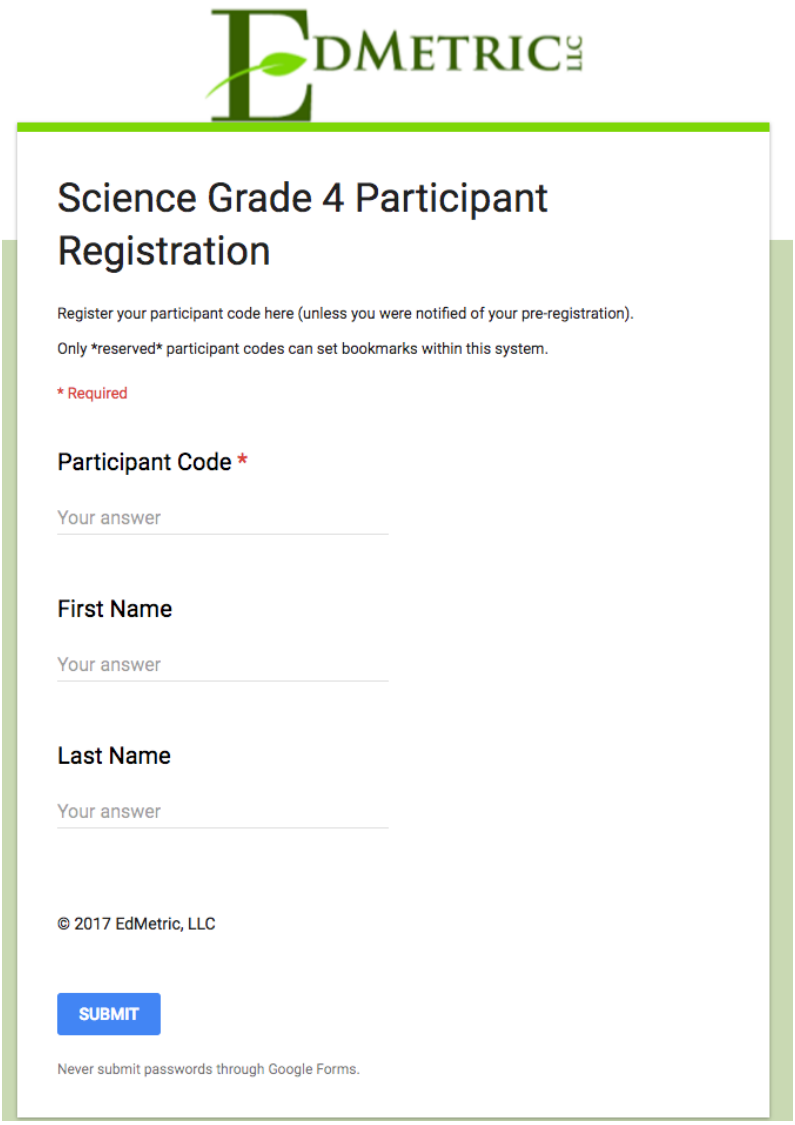


Figure 3.2. Online Panelist Registration

Panelist registration was specific to table, in order to allow median Bookmark placements and impact data for to be easily determined for each table. Following registration, each panelist accessed the system via the appropriate links in their Control Panel (a PDF form provided via Google 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.

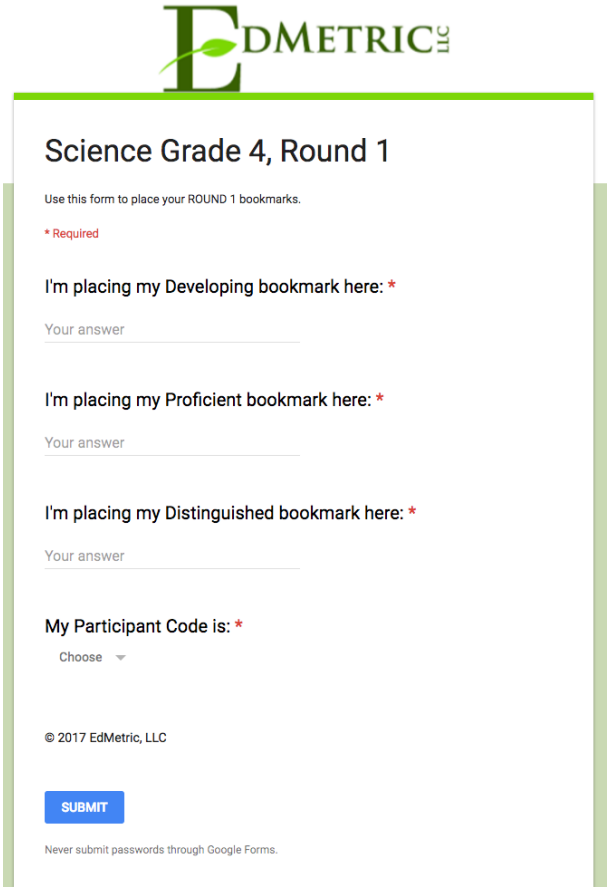


Figure 3.3 Round 1 Bookmark Placement Form

Panelists were instructed to first place their Proficient Bookmarks, followed by Developing and Distinguished. 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 from Google Drive 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 Developing, Proficient, and Distinguished.

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 and table. 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 cut scores associated with Round 1 judgments. Detailed judgments may be found in Appendix D. Round 1 cut scores were not shared with panelists prior to moving to Round 2.

Table 3.4. KĀʻEO Science Round 1 Cut Scores

	Level	Round 1 Cut Score
KĀʻEO Science	Level 2 (Developing)	478
	Level 3 (Proficient)	515
	Level 4 (Distinguished)	547

Round 2

Dr. Egan opened Round 2 with a brief training. She reminded panelists to discuss why they placed their Round 1 Bookmarks where they did, and to listen and reflect on the opinions of others. She also reminded panelists that there is not a right or a wrong Bookmark placement. Table facilitators then led small group discussions of the 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 in the online system via the Control Panel link for Round 2 activities.

Table 3.5 shows the median cut scores associated with Round 2 judgments. Detailed judgments may be found in Appendix D.

Table 3.5. Round 2 Cut Scores

	Level	Round 2 Cut Score
KĀʻEO Science	Level 2 (Developing)	478
	Level 3 (Proficient)	519
	Level 4 (Distinguished)	547

Day 2 adjourned following placement of Round 2 Bookmarks. (Round 2 cut scores were not presented to panelists until Day 3.)

Standard Setting Event Day 3

Round 3

Dr. Egan opened Day 3 of the standard setting event by presenting the results of Round 2 to the panelists. She showed the group the median Bookmarks for each table, the median Bookmark 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 Bookmark.

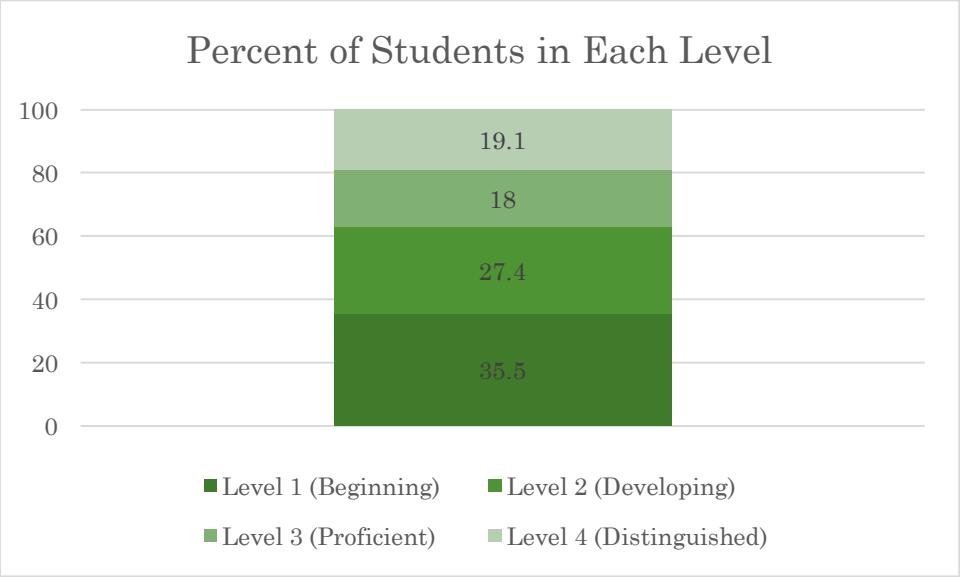


Figure 3.4. KĀʻEO Science Round 2 Impact Data

Following Dr. Egan's review of Round 2 results, Ms. Fernandez led a large-group discussion prior to panelists returning to their small groups to set their Round 3 Bookmarks. After all panelists had set their Round 3 Bookmarks, Dr. Egan again shared results. Table 3.5 shows the Round 3 cut scores; Figure 3.5 shows Round 3 impact data. Detailed judgments may be found in Appendix D.

Table 3.5. Round 3 Cut Scores

	Level	Round 3 Cut Score
KĀʻEO Science	Level 2 (Developing)	480
	Level 3 (Proficient)	519
	Level 4 (Distinguished)	562

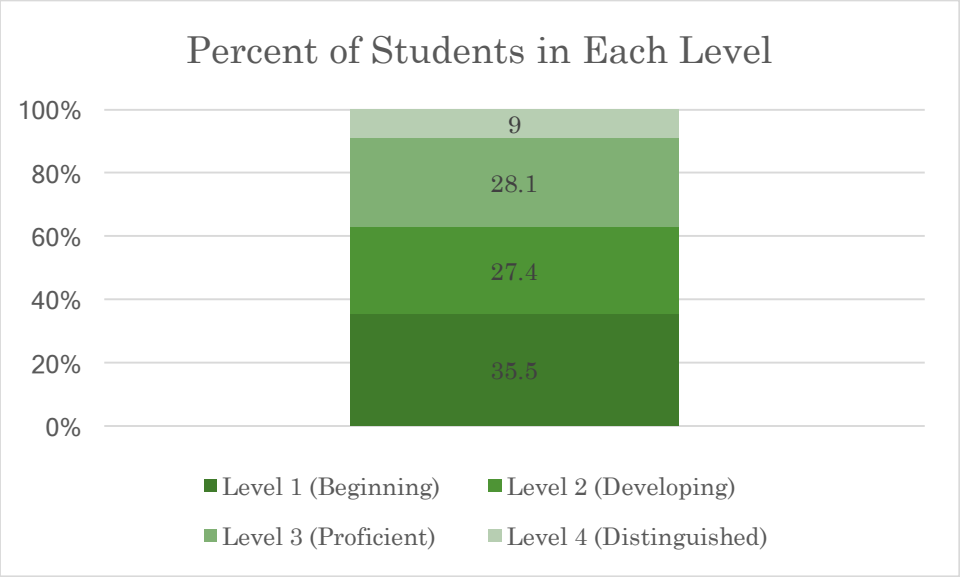


Figure 3.5 KĀʻEO Science Round 3 Impact Data

Dr. Egan noted that the Distinguished Bookmark changed by only a single item, but the impact data changed by several points. She explained that there was a large difference in the location value of the Round 2 Bookmark and the Round 3 Bookmark. The group was offered the opportunity to participate in a fourth round of judgments; however, the consensus was to recommend the cut scores associated with the Round 3 Bookmarks.

Reporting ALDs

Following the third round of Bookmark placements, Dr. Egan trained the group on reporting ALDs using the slide deck "Training 4 KĀʻEO Bookmark Training" (Appendix B). Training described the process of moving from threshold ALDs to reporting ALDs, focusing on item content and the description of the target student. To create reporting ALDs, the group was divided into three small groups. Group 1 drafted reporting ALDs for Beginning and Developing; Group 2 drafted reporting ALDs for Proficient; and Group 3 drafted reporting ALDs for Distinguished. When each group had completed their assigned reporting ALD drafts, the groups came together and reviewed the language across all achievement levels to ensure a clear progression from one level to the next.

Final Workshop Evaluation

Before the workshop adjourned, all participants completed a final workshop evaluation. Detailed results of the evaluation are discussed in Chapter 4.

CHAPTER 4. PARTICIPANT EVALUATION OF KĀʻEO STANDARD SETTING

Participant feedback was obtained in several ways during the KĀʻEO science standard setting. First, an ongoing feedback loop allowed table leaders and panelists to communicate with workshop facilitators and UH staff throughout the workshop. Second, all panelists completed standardized evaluations after the opening session of the standard setting event (Table 3.2) and again after Bookmark training (Table 3.3). The purpose of the first evaluation was to determine panelists' understanding of the standard setting process, their role as participants, and the purposes and uses of achievement level descriptors. The purpose of the second evaluation was to determine panelists' preparedness to begin implementing the BSSP. Results of these readiness surveys are discussed in detail in Chapter 3.

Finally, panelists submitted evaluation forms following the completion of all 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 Google Drive. 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.

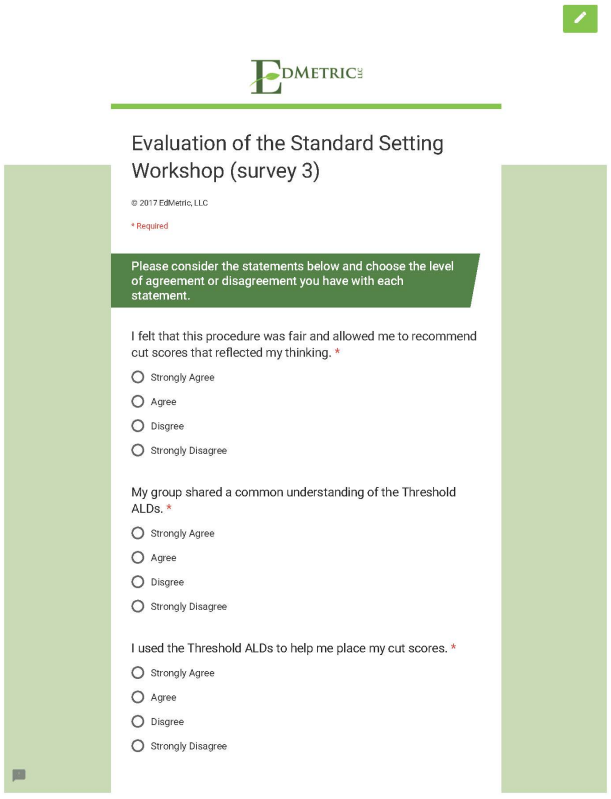


Figure 4.1. Panelist Evaluation Sample

Table 4.1 shows the complete list of statements to which panelists responded. For questions 1 through 26, panelists indicated whether they Strongly Agreed, Agreed, Disagreed, or Strongly disagreed with the statement.

Table 4.1. Panelist Evaluation Questions

Panelist Evaluation Questions
I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking.
My group shared a common understanding of the Threshold ALDs.
I used the Threshold ALDs to help me place my cut scores.
During Round 1, I placed my cut scores independently.
I had enough time to consider the placement of my cut scores.
I am satisfied with the Threshold ALDs.
I am satisfied with the Range PLDs.
The policy definitions were clearly communicated.
I understood how to place my cut scores.
I had enough time to consider the placement of my cut scores.
I am satisfied with our draft Reporting ALDs.
I understand the Reporting ALDs will be finalized after the workshop.
I feel the recommended cut scores that resulted from this process are reasonable.
I would defend the panel's recommended Level 3 cut scores against criticism that they are too high.
I would be able to defend the panel's recommended Level 3 cut scores against criticism that they are too low.
I would be able to defend the panel's recommended Level 4 cut scores against criticism that they are too high.
I would be able to defend the panel's recommended Level 4 cut scores against criticism that they are too low.
I would be able to defend the panel's recommended Level 2 cut scores against criticism that they are too high.
I would be able to defend the panel's recommended Level 2 cut scores against criticism that they are too low.
Overall, I believe that my opinions were considered and valued by my group.
Overall, I valued the workshop as a professional development experience.
This experience will help me target instruction for the students in my classroom.
Participating in the workshop increased my understanding of the KĀʻEO assessments.
The food and service at the facility met my expectations.
The work space had accommodations appropriate to facilitate our work.
The workshop was well organized.

Panelist Evaluation Questions
Which of the following best describes your current position? (Classroom Teacher, School Administrator; Non-classroom teacher; Curricular and/or Instructional Facilitator; Other)
How many years have you been in your current profession?
Please check all of the following in which you have experience (Special Education; English Language Learner; Hawaiian Language Learner; Title 1 Reading; Vocational Education; Adult Education; Other)
What is your gender? (Female; Male; Prefer not to say; Other)
Are you of Hispanic origin? (Yes; No)
What is your race? (Asian/Pacific Islander; Black/African American; American Indian; White; Multi-racial; Prefer not to say)
Your turn. Do you have any additional comments or thoughts about the workshop?

In general, panelists' evaluations of the workshop were positive, with the all panelists either agreeing or strongly agreeing with all 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 Understanding, Review, and Revision of ALDs; Bookmark Standard Setting Implementation; Confidence in Cut Scores; and Overall Impressions. Results are based on the responses of 10 panelists. The tables that follow show the percent of panelists that indicated they "Strongly Agree" or "Agree" with the evaluation statement. Complete evaluation results along with panelists' comments, are included in Appendix E.

UNDERSTANDING, REVIEW, AND REVISION OF ALDS

Table 4.2 shows the evaluation questions and panelists' responses pertaining to Understanding, Review, and Revision of ALDs. Responses to each question were submitted by all ten panelists.

Table 4.2. Evaluation – Understanding, Review, and Revision of ALDs

Evaluation Question	% of Panelists Answering "Agree" or "Strongly Agree"	% of Panelists Answering "Disagree" or "Strongly Disagree"
My group shared a common understanding of the Threshold ALDs	100	0
I am satisfied with the Threshold ALDs.	100	0
I am satisfied with the Range ALDs.	100	0
The policy definitions were clearly communicated.	100	0

Evaluation Question	% of Panelists Answering "Agree" or "Strongly Agree"	% of Panelists Answering "Disagree" or "Strongly Disagree"
I am satisfied with our draft Reporting ALDs	100	0
I understand the Reporting ALDs will be finalized after the workshop.	100	0

Clear understanding of the policy definitions and threshold ALDs is foundational to effective standard setting. Panelists unanimously responded that they clearly understood both the policy definitions and threshold ALDs, indicating that workshop facilitators effectively communicated the content within those types of ALDs. Furthermore, panelists unanimously responded that they were satisfied with their work in suggesting revisions to the threshold and range ALDs, as well as their draft of the reporting ALDs, indicating their clear understanding of the content represented within the assessment items and its connection to student performance.

BOOKMARK STANDARD SETTING IMPLEMENTATION

Table 4.3 shows panelists' responses to evaluation questions specifically related to the implementation of the BSSP. Responses were submitted by all ten panelists.

Table 4.3. Evaluation – Bookmark Standard Setting Implementation

Evaluation Question	% of Panelists Answering "Agree" or "Strongly Agree"	% of Panelists Answering "Disagree" or "Strongly Disagree"
I used the Threshold ALDs to help me place my cut scores.	100	0
During Round 1, I placed my cut scores independently.	100	0
I had enough time to consider the placement of my cut scores.	100	0
I understood how to place my cut scores.	100	0

Panelists' responses to evaluation questions indicated that the BSSP was implemented with fidelity, beginning with appropriate consideration of threshold ALDs in placing cut scores and carrying through independent judgments by panelists. All panelists expressed that they understood how to place a Bookmark, and that they were allocated sufficient time to do so thoughtfully.

CONFIDENCE IN CUT SCORES

Table 4.4 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 4.4. Evaluation – Confidence in Cut Scores

Evaluation Question	% of Panelists Answering "Agree" or "Strongly Agree"	% of Panelists Answering "Disagree" or "Strongly Disagree"
I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking.	100	0
I feel the recommended cut scores that resulted from this process are reasonable.	100	0
I would be able to defend the panel's recommended Level 3 cut scores against criticism that they are too high.	100	0
I would be able to defend the panel's recommended Level 3 cut scores against criticism that they are too low.	100	0
I would be able to defend the panel's recommended Level 4 cut scores against criticism that they are too high.	100	0
I would be able to defend the recommended Level 4 cut scores against criticism that they are too low.	100	0
I would be able to defend the panel's recommended Level 2 cut scores against criticism that they are too high.	100	0
I would be able to defend the recommended Level 2 cut scores against criticism that they are too low.	100	0
Overall, I believe that my opinions were considered and valued by my group.	100	0

Panelists 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. 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.

OVERALL IMPRESSIONS

Panelists' overall impressions of the standard setting event were overwhelmingly positive, as shown in Table 4.5, and reflected in panelists' comments on the evaluation.

Table 4.5. Evaluation – Overall Impressions

Evaluation Question	% of Panelists Answering "Agree" or "Strongly Agree"	% of Panelists Answering "Disagree" or "Strongly Disagree"
Overall, I valued the workshop as a professional development experience.	100	0
This experience will help me target instruction in my classroom.	100	0
Participating in the workshop increased my understanding of the KĀʻEO assessments.	100	0
The food and service at the facility met my expectations.	100	0
The work space had accommodations appropriate to facilitate our work.	100	0
The workshop was well organized.	100	0

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:

- *"Mahalo to Karla, Pōhai, Pono and all of the other limahana for all that you folks did to help make this workshop successful!"*
- *"Mahalo for helping us through this process!"*
- *"Mahalo piha i kā ʻoukou hana nui e hoʻomākaukau i nā mea e pono ai!"*
- *"Mahalo nui!"*

- *"The work is huge - but my understanding of the process is becoming clearer. I appreciate the time and effort put into having the information available to all in attendance. When the ALDs are finalized, how can we access a copy?"*

- *"This was very valuable :)"*

- *"Please ask me to come back. This was fun!"*

- *"Mahalo nui no ka ho'onui 'ike.....a me ka mea'ai 'ono loa kekahi!"*

CHAPTER 5. CONTRIBUTIONS TO THE VALIDITY ARGUMENT

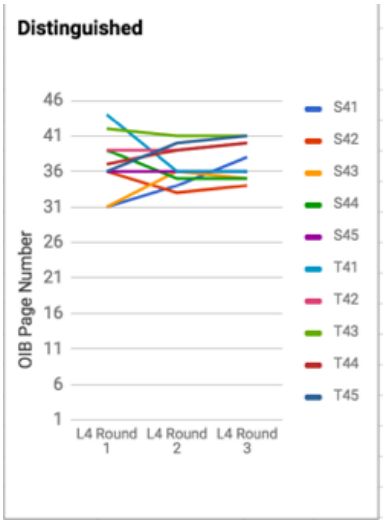
Technically sound standard setting procedures are a critical piece in establishing the validity of an assessment. As such, the standard setting plan and methodology, the standard setting workshop itself, the recommended cut scores and corresponding impact data, and participants' evaluation responses 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 field 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. Figure 5.1 shows the convergence plots for the KĀʻEO science standard setting across rounds for each achievement level. In general, there was greater agreement in Round 3 compared to Round 1 for all cut scores, and little movement from Round 2 to Round 3.



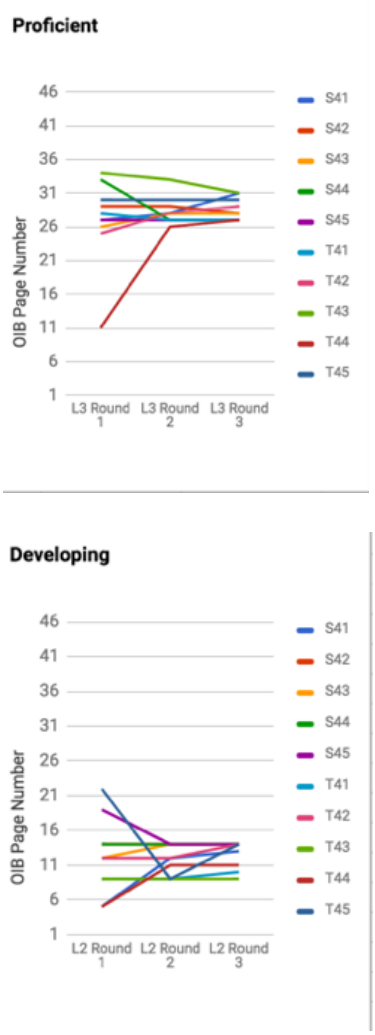


Figure 5.1. Internal convergence plots, KĀʻEO science

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). Panelists unanimously 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, and 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 9 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 specifically in Hawaiian language immersion schools), and all were classroom teachers. Additionally, the group had panelists who worked with special education students and with ELL 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 is 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; however, this recommendation does not consider the scope of the testing program. There are only 14 language immersion schools. There is not a large population of teachers from which to recruit potential standard setting panelists. The 10 panelists represented 9 schools. As such, the size of the panel seems appropriate given the size of the population.

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. For this standard setting, the total group of panelists were split into two small groups.

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., constructed response). 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 3.3). 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 4.3).

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 ALDs, (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 ALDs); 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 UH KĀʻEO development team explained the purpose of the test, and 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. All panelists indicated the opening session provided an adequate background on the Kaiapuni program, purpose of the meeting, and their role at the standard setting event (see Table 3.1).

Once panelists had studied their OIBs, Dr. Egan overviewed the threshold ALDs and the target student, and the specifics of Bookmark placement. Panelists indicated their understanding of the threshold ALDs, (see Table 3.2)

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, Dr. Egan reminded panelists to place Bookmarks based on the threshold ALDs (i.e., the target students). All panelists indicated that the threshold ALDs 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.

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 4.4 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. Review of the data allows panelists to see how their recommendations will play out in the real world. Impact data were presented after Round 2, and panelists had the opportunity to consider this information prior to placing their Bookmarks for Round 3.

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 3-day period. 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.

A detailed agenda was created with step-by-step instructions which described how the implementation would proceed.

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 2017 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).

APPENDIX A. STANDARD SETTING PLAN

SECTION 1. OVERVIEW

The University of Hawai'i (UH) is working collaboratively with the Hawai'i Department of Education's Office of Hawaiian Education to develop academic content standards and a suite of Native Hawaiian assessments – the Kaiapuni Assessments of Educational Outcomes (KĀʻEO) – for students enrolled in Hawaiian Language Immersion Programs. Students in Hawaiian Language Immersion Programs receive all instruction exclusively in the Hawaiian language until they reach Grade 5, at which time English is taught as a content area. Both the Kaiapuni academic content standards and the aligned assessments are reflective of the nuances of Hawaiian language and culture.

Assessment development and standard setting have been completed for Grades 3 and 4 Language Arts and Mathematics, and these tests are currently administered operationally as part of Hawaii's statewide assessment system in compliance with the federal Every Student Succeeds Act (ESSA). All students enrolled in Hawaiian Language Immersion Programs take the KĀʻEO assessments in lieu of the Smarter Balanced English language arts/literacy and Mathematics assessments, which are part of Hawaii's general state assessment program, and student performance on these assessments provides the basis for accountability (Hawaii's Strive HI program) for Hawaiian Language Immersion schools and classrooms.

Concurrent with the initial administration of the KĀʻEO Language Arts and Mathematics assessments, the UH KĀʻEO development team began working with teachers in the Hawaiian Language Immersion Program to create an assessment blueprint, item specifications, and eventually test items, for a KĀʻEO Science assessment for students in Grade 4. The KĀʻEO Science assessment was field tested in Spring 2016 with all Grade 4 KĀʻEO students participating. The test was administered operationally for the first time in Spring 2017. Like the KĀʻEO Language Arts and Mathematics assessments, the KĀʻEO Science assessment consists of multiple item types administered to students in an online environment.

In order for the KĀʻEO Science assessment to be incorporated meaningfully into Strive HI accountability determinations, the UH KĀʻEO development team needs to establish performance standards and achievement level descriptors that are reflective of the content standards being assessed, and consistent with the rigor of the standards previously established for the KĀʻEO Language Arts and Mathematics Assessments.

The purpose of this paper is to propose a design for setting performance standards for the Grade 4 KĀʻEO Science assessment using a technically sound, content-based procedure that will result in meaningful communication of student performance to a variety of stakeholders. This section provided background information about the development of the KĀʻEO assessments. The second section describes the proposed process for establishing performance standards – the Bookmark Standard Setting Procedure (BSSP) – and developing the accompanying Performance Level Descriptors (PLDs). Section 3 describes the materials that will be necessary for implementation of the workshop, and identifies the entities that will be responsible for developing and providing the materials. Section 4 describes the participants in the standard setting workshop. Section 5 describes the workshop implementation, including facilities, workshop design, high-level daily agenda, and security of materials. Section 6 discusses workshop documentation.

SECTION 2. METHODOLOGY – BOOKMARK STANDARD SETTING PROCEDURE

EdMetric proposes the use of the Bookmark Standard Setting Procedure (BSSP) to determine performance standards for the Grade 4 KĀʻEO Science assessment. The BSSP engages panelists with content expertise in multiple rounds of discussions of assessment content (and its connection to the applicable content standards) to determine the appropriate placement of cut points to delineate between performance levels. The BSSP is appropriate for assessments that include mixed item types and rely upon item-response theory models for scaling. The KĀʻEO Science assessment meets both of these criteria. Additionally, the BSSP was used to establish performance standards for the KĀʻEO Language Arts and Mathematics assessments.

For the Grade 4 KĀʻEO Science assessment, three cut points will be identified to establish four levels of performance: Level 1 (Ho'omaka), Level 2 (Holomua), Level 3 (Makaukau), and Level 4 (Kelakela). This scale is consistent with the scale established for the KĀʻEO Language Arts and Mathematics assessments.

Implementation of the BSSP begins with panelists experiencing the assessment just as a student would experience it. Panelists then study assessment content using an Ordered Item Booklet (OIB), which consists of operational assessment items arranged in ascending order of difficulty. For the KĀʻEO Science assessment, items will be ordered according using a Response Probability (RP) criterion of .50, or the location on the test scale where students have a 50/50 chance of answering the item correctly. This is consistent with the RP criterion used for standard setting for the KĀʻEO Language Arts and Mathematics assessments. The OIB is accompanied by an item map that specifies the order of difficulty, the scale location, the item number on the operational test, the scoring key, and the content standard the item measures. Panelists use the item map to guide their study of items in the OIB, answering two questions for each item:

- 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 than the items preceding it?

Panelists then participate in a discussion of the "Threshold Student," or the student demonstrating just enough content knowledge, skills, and abilities to be proficient (Level 3 on the KĀʻEO assessments). This discussion informs their placement of "Bookmarks" within the OIB to delineate the "proficient" cut score, and subsequently cut scores to delineate the other designated performance levels.

A system of interrelated Performance Level Descriptors (PLDs) also play an integral role in effective implementation of the BSSP and meaningful interpretation of test scores:

- ***Policy PLDs –*** Policy PLDs articulate policymakers' vision of the goals and rigor for the final performance standards.
- ***Range PLDs –*** Range PLDs are grade/content specific descriptors that may be used by item writers to describe the cognitive and content rigor that is encompassed within particular performance levels.

- ***Threshold PLDs –*** Threshold PLDs are used by standard setting panelists and are a subset of the range PLDs. Threshold PLDs represent the minimal knowledge, skills, and abilities that a student needs in order to enter a particular achievement level.
- ***Reporting PLDs –*** Reporting PLDs are also a subset of the Range PLDs. Reporting PLDs are descriptions of the content within each performance level that appear on reports of student performance for a variety of stakeholder groups.

Policy PLDs are typically presented to panelists early in the standard setting process to set the tone for the discussion of assessment content and the rigor that should be expected at each performance level. For the KĀʻEO Science assessment, the Policy PLDs will be those developed for the KĀʻEO Language Arts and Mathematics assessments to ensure consistency in the rigor expected within the program. These PLDs, shown in Table 1, are the same as those adopted by the Smarter Balanced Assessment Consortium.

Table 1: Policy PLDs

Achievement Level	Policy Level Definitions
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.

Threshold PLDs and Range PLDs are content-specific and will be developed based on the Science standards to which the assessment is aligned. The Range PLDs will be drafted by the UH KĀʻEO development team, and they will be revised as part of the standard setting process (see Section 5, Workshop Implementation). Reporting PLDs are typically created following recommendation of cut scores, and are based on the content that emerges within each performance level. Again, these PLDs may be drafted as part of the standard setting process, or may be created by content experts following standard setting. For the KĀʻEO Science standard setting, EdMetric recommends engaging panelists in the process of creating draft Reporting PLDs (see Section 5, Workshop Implementation).

SECTION 3. WORKSHOP MATERIALS

COMPLETE KĀʻEO SCIENCE ASSESSMENT

The UH KĀʻEO development team will provide the operational form of the KĀʻEO Science assessment to panelists via a secure link. Panelists will access the assessment using their personal laptop computers or tablets. The UH KĀʻEO development team will be responsible for ensuring proper item rendering for panelists.

ORDERED ITEM BOOKLET

EdMetric staff will create OIBs consisting of items that were administered operationally in Spring 2017 using the .50 RP criterion. EdMetric will provide a print copy of the OIB for each panelist. Each single point item will be presented on a single page of the OIB. Items that have multiple score points will be presented with each score point on a single page. A separate stimulus booklet will also be provided. Each stimulus will be clearly labeled with the corresponding item number(s) and page(s) in the OIB.

ITEM MAP

EdMetric will create the item map that corresponds to the OIB. EdMetric will provide a print copy of the item map for each panelist. Additionally, each panelist will be provided electronic access to the item map via a "control panel" with a unique access code.

KAIAPUNI SCIENCE CONTENT STANDARDS AND POLICY PLDS

Panelists will need to reference the content standards to which the test is aligned throughout the standard setting process. UH staff should provide a copy of the standards for each panelist. Although the standards are likely to be available electronically, it will be preferable for panelists to have a print copy to reference. UH should also provide panelists with a print copy of the Policy PLDs.

TRAINING MATERIALS

EdMetric will develop the slide deck for training specific to standard setting and how panelists will implement the BSSP. UH should develop a slide deck (or presentation outline) to describe the history of the KĀʻEO assessments, the overall goals of the standard setting workshop, and the Policy PLDs.

MATERIALS FOR BOOKMARK PLACEMENT

EdMetric will tailor an electronic system – a Bookmark Processor – for panelists to place their Bookmarks after each round of discussion. Panelists will access the online system via a "control" panel with a unique access code for each participant. The system will allow workshop facilitators to view each panelist's Bookmark placement for each round; measures of central tendency for each round of Bookmark placements; and impact data following each

round. To accompany the Bookmark Processor, EdMetric will provide each panelist with a paper rating form for each round of judgments.

LAPTOPS

UH should request that each panelist provide his/her own laptop for completion of standard setting activities.

NON-DISCLOSURE AGREEMENTS

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

WORKSHOP EVALUATIONS

EdMetric will develop three evaluations to be used during the KĀʻEO Standard Setting Workshop. The first evaluation, to be administered following standard setting training, will determine panelists' understanding of the standard setting process and their preparedness to study the OIB. The second evaluation, to be administered after panelists study the OIB, will determine panelists' readiness to begin the task of placing their Bookmarks. The third evaluation, to be administered upon completion of the workshop, will provide an indication of how satisfied panelists were with the workshop and with the recommended performance standards. Both evaluations will be presented in online format. Panelists will access the evaluations via their "control panel". Results will be entered via Google Forms, allowing workshop facilitators real-time access to results.

SECTION 4. WORKSHOP PARTICIPANTS

WORKSHOP FACILITATORS

Dr. Karla Egan, Principal and Founder of EdMetric LLC, will serve as the lead facilitator for the KĀʻEO Science Standard Setting Workshop. Dr. Egan has presented and published papers in the field of standard setting, and she has designed and led over 40 standard setting workshops, including the KĀʻEO Language Arts and Mathematics standard setting workshop. Dr. Egan is currently serving on the National Academy of Sciences committee that is evaluating the National Assessment of Educational Progress (NAEP) achievement levels in Reading and Mathematics. She has implemented most major standard setting methodologies, including item-mapping procedures, modified Angoff procedures, paper-sorting procedures, and student-centered procedures. Dr. Egan was the lead creator of an innovative framework for achievement level descriptors that was used by the Smarter Balanced Assessment Consortium and by several states in developing their achievement level descriptors. Dr. Egan will provide all training specific to standard setting and the BSSP during the workshop. She will also provide an orientation and training for Table Facilitators.

Dr. Pohai Kukea Shultz, Principal Investigator of the University of Hawaii's Hawaiian Immersion Assessment Project, will co-facilitate the workshop. Dr. Kukea Shultz has served in a key leadership position, overseeing the development of the KĀʻEO assessments since their inception. Dr. Kukea Shultz will provide an overview of the KĀʻEO assessments and the program's history. She will assist Dr. Egan with security, data management, and time management during the workshop. She will also serve as the liaison to the Hawaii Department of Education.

CONTENT SPECIALIST

EdMetric recommends that UH identify a Science content specialist who is familiar with both the content standards and the assessment to serve as a resource throughout the workshop.

PANELISTS

UH should identify a total of 12 panelists to participate in the KĀʻEO Science Standard Setting Workshop. Panelists should be experienced educators who reflect the diverse backgrounds and needs of Hawaiian students. The final panel should represent a sample of expert panelists from a pool of all such qualified experts. These educators should have experience in Hawaiian language immersion programs. They will be responsible for studying the content of the test, participating in group discussions, and making individual decisions about the requirements for each performance level.

TABLE LEADERS

From the selected panelists, UH should identify two Table Leaders. Each Table Leader will be assigned to a table of panelists. Table Leaders will facilitate discussion and keep the process on track within their tables. Table Leaders are

full participants and it is recommended that they be educators of notable status. Table 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. Their primary role will be 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 for participants, or request assistance.

SECTION 5. WORKSHOP IMPLEMENTATION

LOGISTICS

Facilities/Meeting Rooms

UH will be responsible for arranging meeting rooms for the workshop. It is expected that a single meeting room will be used, and that all secure materials will be stored in a locked office when not in use. Room 1 must be large enough to comfortably accommodate two round tables of 6 panelists. The room must be large enough for each table to hold discussions without interrupting the other group. Room 1 should have a projection screen and an LCD projector for training and for sharing results with panelists.

The room should be equipped with wireless internet connectivity for all panelists and workshop facilitators. The room should be able to be secured, allowing participants and workshop facilitators to leave materials onsite for the duration of the workshop.

The facility at which the workshop is conducted should be able to provide a small volume of copying services as needs arise during the workshop.

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

Participant Lodging

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

WORKSHOP DESIGN OVERVIEW

EdMetric proposes a 3-day KĀʻEO Science Standard Setting Workshop during which panelists will finalize range PLDs, recommend three scale score cut points to delineate four performance levels, and draft Reporting PLDs.

Prior to panelists arriving at the workshop, EdMetric will provide access to a recorded training webinar for the two identified Table Leaders. This training can occur anytime during the week preceding the workshop.

Registration and Welcome

The morning of the workshop, panelists will register and receive non-secure workshop materials. All panelists will sign a non-disclosure agreement at the time of registration. When all panelists have arrived and registered, panelists will gather in the meeting room for an Opening Session. During the Opening Session, Dr. Schultz will welcome panelists and provide them with general housekeeping information; a historical overview of the Hawaiian Language Immersion Program and the KĀʻEO assessments; a description of the goals of the standard setting workshop; and

an explanation of the Policy PLDs. The Opening Session should emphasize the important contribution of the standard setting process to the valid interpretation of test scores, discuss the aspects of the KĀʻEO Science assessment that are unique and of cultural significance, and describe the level of rigor that is expected of the recommended cut scores.

Standard Setting Training

Following Dr. Kukea Shultz's welcome and overview, Dr. Egan will train panelists on the standard setting procedure. She will discuss the process of standard setting, explain the organization of the OIB and item map, and clarify the roles and responsibilities of all participants. She will then train participants on how to proceed with a discussion of the Range PLDs and the expectations for the Threshold Student (i.e., the student who just barely achieved Level 3 performance). Following Dr. Egan's training, all panelists will complete a Standard Setting Readiness Survey to ascertain their preparedness to study the OIB. Dr. Egan will provide additional individual training as necessary if requested by panelists.

Finally, she will train participants on how to access the Bookmark Processor via their "control panels" and set their Bookmarks.

Breakout Groups – Range PLDs and Threshold Student Discussions

Following Dr. Egan's training, panelists will move into their table groups and proceed with a discussion of the Policy, Range, and Threshold PLDs. This discussion is expected to last 3 to 4 hours, and it will provide the framework for setting Bookmarks.

The UH staff member who drafts the Range PLDs should plan to attend this session.

Breakout Groups – Round 1

To begin Round 1, panelists will take the operational form of the KĀʻEO Science assessment in their table groups. They will then study the OIB. Table Leaders will facilitate the discussions at each table, leading panelists through the guiding questions regarding what each item measures and why it is more difficult than the preceding items in the booklet. Each table will identify a scribe to make notes on the item map. Table Leaders will then reiterate the process for accessing the online system to set Bookmarks. Following the OIB discussions, each panelist will complete an evaluation to determine their readiness to proceed with the first Bookmark placement. Dr. Egan will provide additional individual training as necessary if requested by panelists. When panelists have indicated that they are prepared to make their first judgments, they will independently set their first Bookmarks, beginning with the Level 3 cut, and proceeding to the Level 2 and Level 4 cuts. Workshop staff will calculate impact data (the percent of students falling within each performance level) for each table and for the group as a whole based on the median recommended cuts, for Round 1; however, impact data will not be shared with panelists at this time.

Breakout Groups – Round 2

During Round 2, each table will be shown their group median Bookmark placements. The Table Leader will facilitate a discussion of items for which there was not consensus according to the 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 panelists at each table. Following discussion, each panelist will independently make recommendations for adjusting their Round 1 Bookmark placements. Workshop staff will again calculate impact data based on the median recommended cuts for each table and for the group as a whole.

Large Group – Round 3

Dr. Egan will bring all panelists together for a Round 3 discussion. At this time, she will present the Round 2 impact data for each table and for the group as a whole. She will facilitate a discussion of any discrepancies that might exist between the two tables, as well as of the overall impact data. Panelists will then have a final opportunity to reset their Bookmarks. Recommended cut scores will be determined by finding the median of the Round 3 results.

Reporting PLDs

Following Round 3, panelists will engage in a process to refine the Range and Threshold PLDs to Reporting PLDs. These descriptors explain the knowledge, skills and abilities of the students in each achievement level. Panelists will use the information gathered from their study of the test, content standards, and understanding of the Threshold Student to add clarity and conciseness to the Reporting PLDs. This may be accomplished by moving panelists back into their table groups with a discussion led by the Table Facilitators, or with a large group discussion facilitated by Dr. Egan.

Final 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.

HIGH-LEVEL AGENDA

Table 2 presents a high-level agenda for the KĀʻEO Science Standard Setting Workshop.

Table 2: High-Level Agenda

	Standard Setting Task
Day 1 Morning	Registration, Orientation, and Training
Day 1 Afternoon	Range PLDs & Threshold Students
Day 2 Morning	Study Ordered Item Booklet Threshold Student Discussions and Round 1 Bookmark Placement

Day 2 Afternoon	Round 2 Discussions and Bookmark Placements; Presentation of Impact Data and Round 3 Bookmark Placements
Day 3 Morning	Reporting PLDs Final Workshop Evaluation

WORKSHOP SECURITY

Because the KĀʻEO Science assessment is a secure test used for accountability purposes, security of materials during the standard setting event is imperative. Security will begin with the acquisition of panelists' signatures on non-disclosure agreements at the time of registration. The non-disclosure agreement should specify that participants will not remove any secure materials from the meeting rooms and will not disclose the content of test items after the workshop.

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 meeting 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.

Secure materials are not permitted outside the meeting room. After each day, Table Leaders will follow an auditing procedure in order to account for all secure materials.

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

SECTION 6. DOCUMENTATION AND FOLLOW-UP

EdMetric will provide all necessary documentation to support the validity of the standard setting process. This design document provides UH with a comprehensive description of the workshop materials and the methods by which they are to be created, as well as a detailed discussion of the recommended methodology and the implementation of the workshop itself. The final design document will be made available to UH in a format suitable for formal documentation and submission for United States Department of Education Peer Review purposes.

Following the standard setting, EdMetric will document the process and results in a comprehensive technical report. This report will be designed to assist UH in evaluating the performance standards recommended by the panelists, and to promote clear understanding of the process by stakeholders. The technical report will contain a narrative description of the standard setting event; detailed information about judgments made by panelists; information about standard errors of measurement and of the cut score; graphical representations of panelists' judgments; detailed summaries of panelists' evaluations; and copies of the handouts and slide decks used during the standard setting workshop.

APPENDIX B. WORKSHOP PRESENTATIONS AND HANDOUTS

TRAINING 1 KĀʻEO OPENING SESSION

Kaiapuni Assessment of Educational Outcomes (KĀʻEO)
Opening Session

ALD Training

EDMETRICS

Overview of Training

Standard Setting | ALDs | Q & A

EDMETRICS

OVERVIEW OF STANDARD SETTING

EDMETRICS

Test Development Cycle

EDMETRICS

What is Standard Setting?

EDMETRICS

How do we do this?

You provide:

- Knowledge of student performance
- Knowledge of content

We provide:

- Content Standards
- Policy level descriptors
- Drafts of ALDs
- Process for setting cut scores

EDMETRICS

KĀʻEO Cut Scores

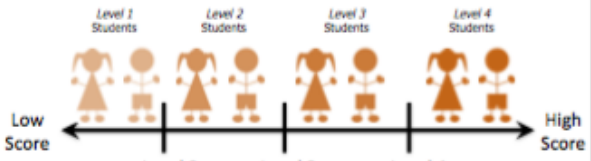
- Three cut scores
  Level 2, Level 3, Level 4
- Four achievement levels
  Level 1, Level 2, Level 3, Level 4

How do we set cut scores?

- Pr[illegible]s
  - Ar[illegible]rary
  - Test s[illegible]ic
  - [illegible]t is not 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

Developing Expertise

- You know students and content
- Activities to enhance knowledge:
  1. Study content standards
  2. Review & revise Range & Threshold ALDs
  3. Study test items
     - Take test
     - Study ordered item booklets using item maps

ACHIEVEMENT LEVEL DESCRIPTORS

REMEMBER

ALD

=

Achievement Level Descriptor

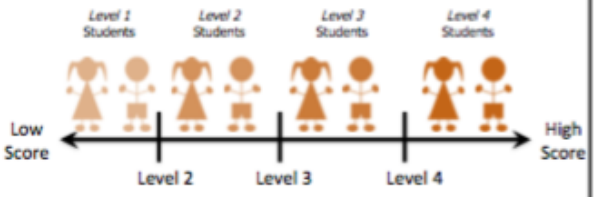
What are Achievement Level Descriptors?

- Define the types of knowledge, skills, and processes **expected of or demonstrated by** students in each level of performance

ALDs and Student Performance

- Test scale represents range of student scores
- Cut scores will separate students into achievement levels
- KÃ'EO : 3 cut score, 4 achievement levels

ALDs Define Student Performance



What are Achievement Level Descriptors?

- Aggregate students into groups based on test performance

  Level 4, Level 3, Level 2, Level 1

- Define the types of knowledge, skills, and processes **expected of or demonstrated by** students in each level of performance
- Connect the assessment and content standards

Who Uses ALDs?

Family of ALDs

Policy ALDs: Broad statements that set the tone and rigor for the other ALDs
- Used by Policymakers

Range ALDs: Expectations for whole of student performance in an achievement level
- Specific to a grade and content area.
- Utilized by item writers and educators

Threshold ALDs: Expectations for student performance at the entry point of an achievement level.
- Specific to a grade and content area.
- Utilized by standard setters.

Reporting ALDs: Students' performance at the entry point of an achievement level.
- Specific to a grade and content area.
- Utilized by stakeholders.

EDMETRICS

POLICY DESCRIPTORS

EDMETRICS

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 accountability purposes.

EDMETRICS

RANGE ALDS

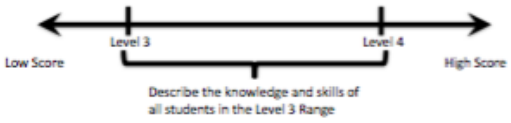
EDMETRICS

Range ALDs

- Define expected knowledge and skills of all students in a particular achievement level
- Depict the expected grade-level progression by content standard
- Guide item development
- Guide curriculum development

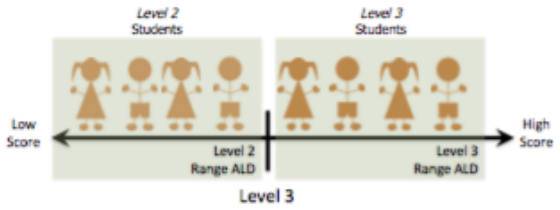
EDMETRICS

Range ALDs for Level 3



EDMETRICS

Range ALDs

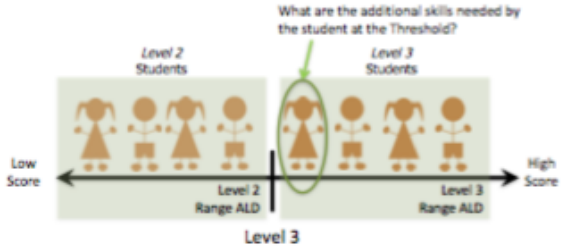


Range ALDs

- Describe *range* of knowledge, skills, and processes for all students in each achievement level
  - Level 1
  - Level 2
  - Level 3
  - Level 4
- During the workshop, the existing Range ALDs will be reviewed and revised by the standard setting panel

THRESHOLD ALDS

Range ALDs to Threshold ALDs



Threshold ALD

Threshold ALDs

- During the workshop, you will work with other panelists to create the Threshold ALDs
- Threshold ALDs will guide your work during standard setting

REPORTING ALDS

EDMETRICS

Threshold ALDs to Reporting ALDs

- Threshold ALDs guide the standard setting
- Reporting ALDs are based on the final cut scores and on test items
- You will create Reporting ALDs at the workshop
- Additional training will be provided immediately prior to creating the Reporting ALDs

EDMETRICS

Reporting ALDs

- Reconciles the threshold ALDs with the final cut scores
- Defines the knowledge and skills of students who have entered the achievement level

EDMETRICS

ALDS AND STANDARD SETTING

EDMETRICS

EDMETRICS

Adapted from Egan, Schneider, & Ferrara (2012)

AGENDA & NEXT STEPS

EDMETRICS

Agenda

Day 1
- **Morning** | Opening Session; Study Content Standards; Review & Revise Range ALDs
- **Afternoon** | Create Threshold ALDs; Take Operational Test

Day 2
- **Morning** | Study OIB
- **Afternoon** | Study OIB, Round 1 Bookmarks

Day 3
- **Morning** | Round 2, Round 3
- **Afternoon** | ALD Writing, Evaluation

EDMETRICS

Next Steps

- Readiness Survey
- Content Standards
- Review Range & Threshold ALDs

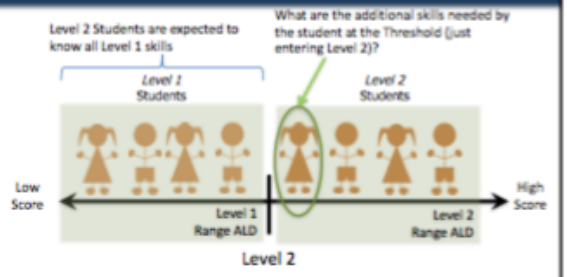
EDMETRICS

Create Threshold ALDs

- Study the Range ALDs
- Revise Threshold ALD for Level 2
  - Study Range ALDs for Levels 1 & 2
  - Consider the Range ALD for Level 1. What additional KSPs are needed in order for a student to enter Level 2?
    - Remember: This is the student who just makes it into Level 2
- Revise Threshold ALD for Level 3
  - Study Range ALDs for Levels 2 & 3
  - Consider the Range ALD for Level 2. What additional KSPs are needed in order for a student to enter Level 3?

EDMETRICS

Range ALDs to Threshold ALDs



EDMETRICS

Questions?

EDMETRICS

Kaiapuni Assessment of Educational Outcomes (KĀ'EO)

Achievement Level Descriptors

ALD Training

EDMETRICS

REMEMBER

ALD

=

Achievement Level Descriptor

EDMETRICS

What are Achievement Level Descriptors?

- Define the types of knowledge, skills, and processes **expected of or demonstrated by** students in each level of performance

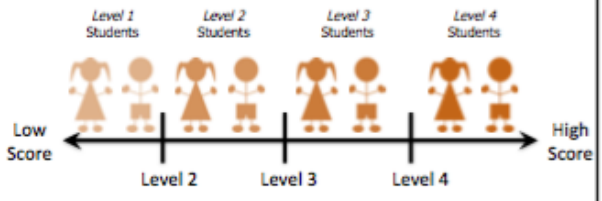
EDMETRICS

ALDs and Student Performance

- Test scale represents range of student scores
- Cut scores will separate students into achievement levels
- KĀ'EO : 3 cut score, 4 achievement levels

EDMETRICS

ALDs Define Student Performance



EDMETRICS

What are Achievement Level Descriptors?

- Aggregate students into groups based on test performance

  Level 4, Level 3, Level 2, Level 1
- Define the types of knowledge, skills, and processes **expected of or demonstrated by** students in each level of performance
- Connect the assessment and content standards

EDMETRICS

Who Uses ALDs?


Policy Makers


Educators


Item Writers


Parents

EDMETRICS

Family of ALDs

Policy ALDs: Broad statements that set the tone and rigor for the other ALDs
- Used by Policymakers

Range ALDs: Expectations for whole of student performance in an achievement level
- Specific to a grade and content area.
- Utilized by item writers and educators

Threshold ALDs: Expectations for student performance at the entry point of an achievement level.
- Specific to a grade and content area.
- Utilized by standard setters.

Reporting ALDs: Students' performance at the entry point of an achievement level.
- Specific to a grade and content area.
- Utilized by stakeholders.

EDMETRICS

POLICY DESCRIPTORS

EDMETRICS

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 accountability purposes.

EDMETRICS

RANGE ALDS

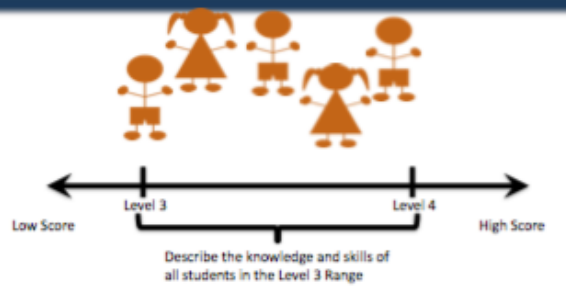
EDMETRICS

Range ALDs

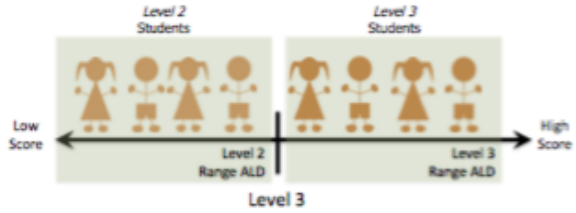
- Define expected knowledge and skills of all students in a particular achievement level
- Depict the expected grade-level progression by content standard
- Guide item development
- Guide curriculum development

EDMETRICS

Range ALDs for Level 3



Range ALDs

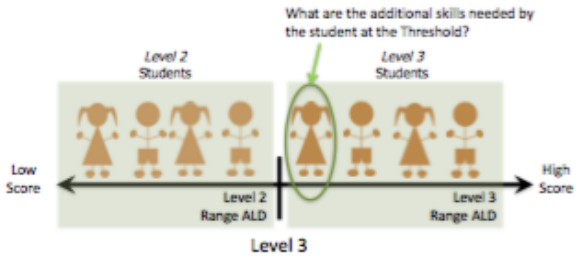


Range ALDs

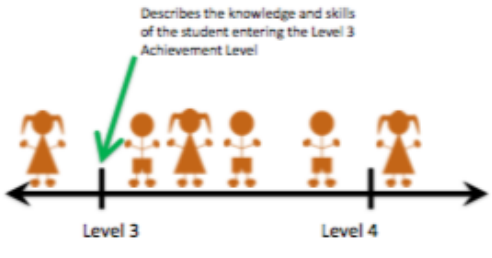
- Describe *range* of knowledge, skills, and processes for all students in each achievement level
  - Level 1
  - Level 2
  - Level 3
  - Level 4
- During the workshop, the existing Range ALDs will be reviewed and revised by the standard setting panel

THRESHOLD ALDS

Range ALDs to Threshold ALDs



Threshold ALD



Threshold ALDs

- During the workshop, you will work with other panelists to create the Threshold ALDs
- Threshold ALDs will guide your work during standard setting

REPORTING ALDS

Threshold ALDs to Reporting ALDs

- Threshold ALDs guide the standard setting
- Reporting ALDs are based on the final cut scores and on test items
- You will create Reporting ALDs at the workshop
- Additional training will be provided immediately prior to creating the Reporting ALDs

Reporting ALDs

- Reconciles the threshold ALDs with the final cut scores
- Defines the knowledge and skills of students who have entered the achievement level

ALDS AND STANDARD SETTING

Adapted from Egan, Schneider, & Ferrara (2012)


Questions?
EDMETRICS

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

Bookmark Standard Setting Procedure

EDMETRICS

Bookmark Standard Setting Process

EDMETRICS

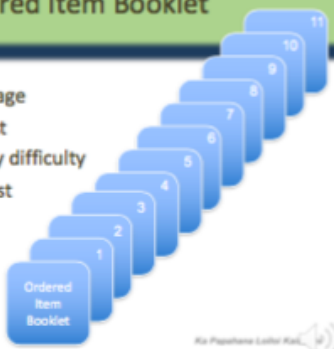
Bookmark Materials

- Ordered Item Booklet
- Item Maps (online)
- Rubrics

EDMETRICS

Ordered Item Booklet

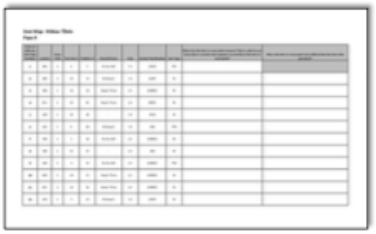
- One item per page
- Easiest item first
- Items ascend by difficulty
- Hardest item last



EDMETRICS

Item Map: Key Questions

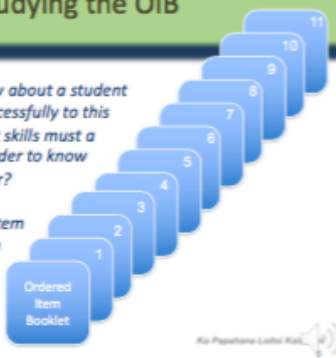
(1) What does this item measure?
(2) Why is this item more difficult than the items that precede it?



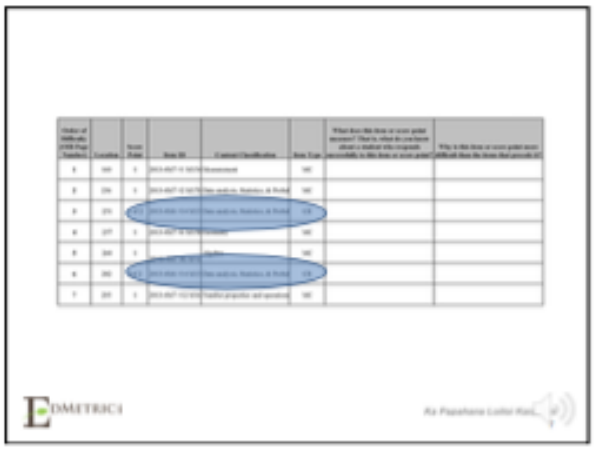
EDMETRICS

Studying 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?*



EDMETRICS



The Bookmark Process

- Study content in a book
- Tell a story about the content students in each level should be able to do
- Set a bookmark

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

Next Steps

- Study the OIB

Questions?

Kaiapuni Assessment of Educational Outcomes (KĀʻEO)
Standard Setting

Bookmark Standard Setting Procedure

EDMETRICS

Bookmark Standard Setting Process

EDMETRICS

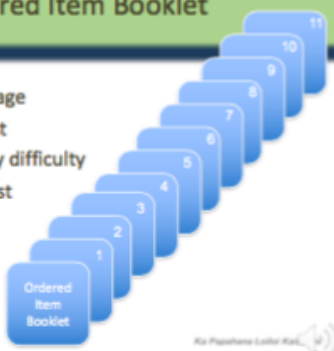
Bookmark Materials

- Ordered Item Booklet
- Item Maps (online)
- Rubrics

EDMETRICS

Ordered Item Booklet

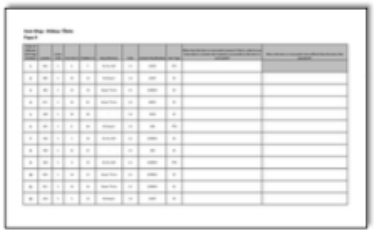
- One item per page
- Easiest item first
- Items ascend by difficulty
- Hardest item last



EDMETRICS

Item Map: Key Questions

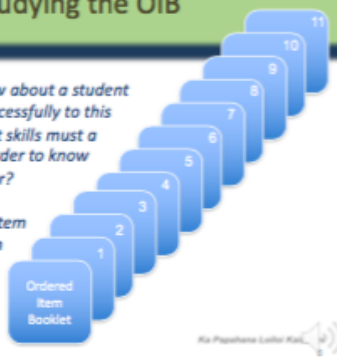
(1) What does this item measure?
(2) Why is this item more difficult than the items that precede it?



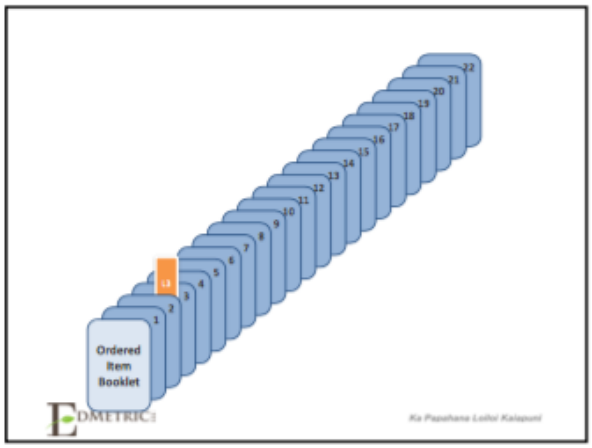
EDMETRICS

Studying 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?*

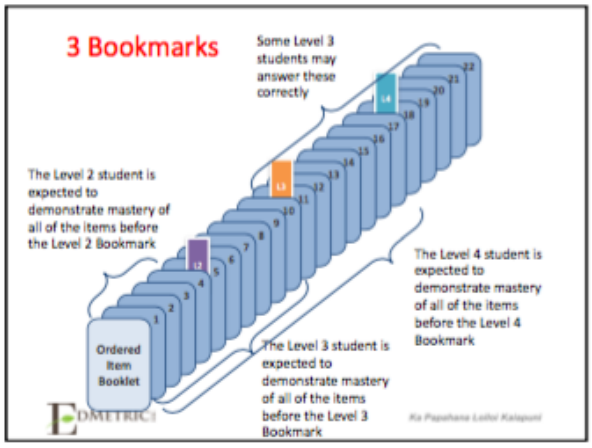


EDMETRICS



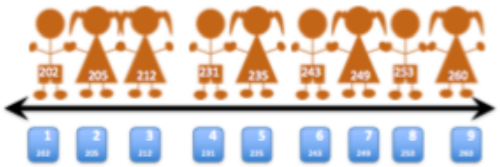
Cut Scores

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



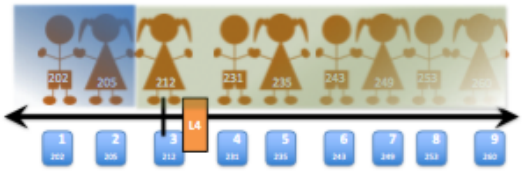
Test Scale

- Items are ordered by difficulty
- Students are ordered by achievement



Bookmarks and Cut Scores

- Bookmark: Separates items
- Cut Scores: Separate students

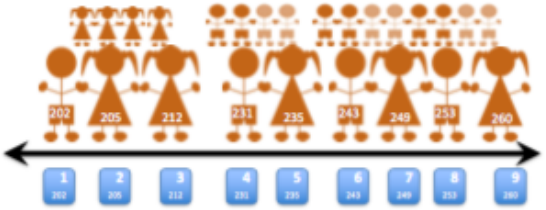


What is meant by Mastery?

- Mastery: the point where a student has a 50/50 chance of answering the item correctly

Mastery & Item Location

- Item Location: Scale score necessary for a student to have 1/2 chance of answering an item correctly

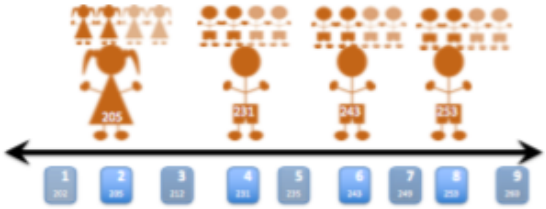


EDMETRIC

Ka Papahana Loiloi Kaiapuni

Mastery & Item Location

- Item Location: Scale score necessary for a student to have 50/50 chance of answering correctly

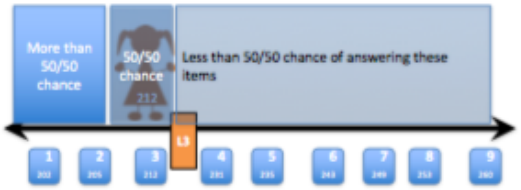


EDMETRIC

Ka Papahana Loiloi Kaiapuni

Mastery and the Target Student

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



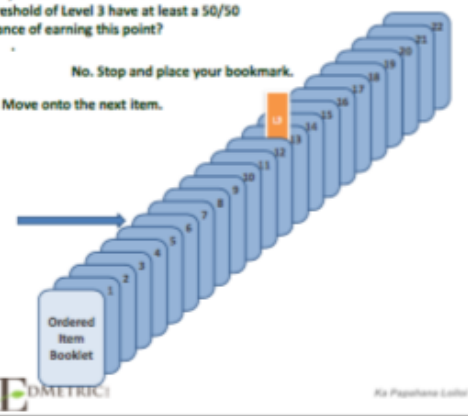
EDMETRIC

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.



EDMETRIC

Ka Papahana Loiloi Kaiapuni

Bookmark Placement

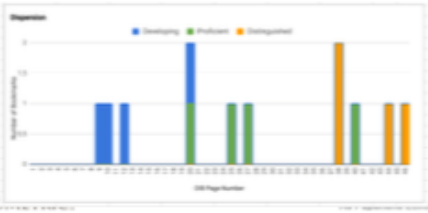
- Individual activity
- Enter results
  - Find your Table Links on Google Drive

EDMETRIC

Ka Papahana Loiloi Kaiapuni

Round 2

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



EDMETRIC

Ka Papahana Loiloi Kaiapuni

Next Steps

- Readiness Survey
- Place your Bookmarks

EDMETRIC

Ka Papahana Loihi Kaiapuni

Questions?

EDMETRIC

Ka Papahana Loihi Kaiapuni

Kaiapuni Assessment of Educational Outcomes (KĀ'ĒO)

Final Session

Reporting ALDs

EDMETRIC

Ka Papahana Loiloi Kaiapuni

Family of ALDs

Policy ALDs: Broad statements that set the tone and rigor for the other ALDs
- Used by Policymakers

Range ALDs: Expectations for whole of student performance in an achievement level
- Specific to a grade and content area.
- Utilized by item writers and educators

Threshold ALDs: Expectations for student performance at the entry point of an achievement level.
- Specific to a grade and content area.
- Utilized by standard setters.

Reporting ALDs: Students' performance at the entry point of an achievement level.
- Specific to a grade and content area.
- Utilized by stakeholders.

EDMETRIC

Ka Papahana Loiloi Kaiapuni

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 accountability purposes.

EDMETRIC

Ka Papahana Loiloi Kaiapuni

Threshold ALDs to Reporting ALDs

EDMETRIC

Ka Papahana Loiloi Kaiapuni

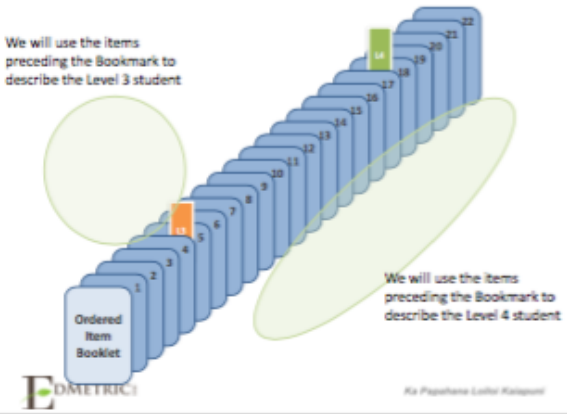
Describe Target Students

Revise item descriptions
- Use action verbs
- Describe concepts found in a group of items without giving away item itself

Goal
- Provide meaningful descriptions for teachers and parents

EDMETRIC

Ka Papahana Loiloi Kaiapuni



EDMETRIC

Ka Papahana Loiloi Kaiapuni

Next Steps

- Readiness Survey
- Place your Bookmarks

EDMETRIC

Ka Papahana Loiloi Kaiapuni

Questions?

EDMETRIC

Ka Papahana Loiloi Kaiapuni

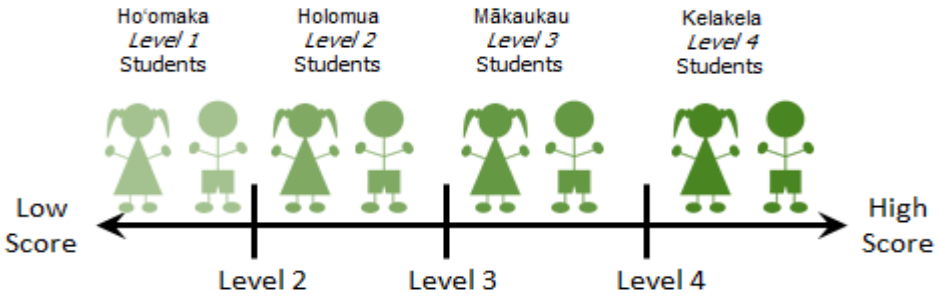
Achievement Level Descriptor Review

What is an Achievement Level Descriptor?

With assessment, students are placed along a continuum of scores (like when you give a test in your classroom. So KĀʻEO has a range of scores between 400-700 approximately. But what does that meet if you get a 558?

Achievement level descriptors or ALDs describe the knowledge and skills of students in each performance category.

But even before standard setting, we start developing ALDS to help inform the development process.



Types of ALDs

It used to be that ALDs were developed at the end of a test cycle but it is becoming more common to start developing them early to guide test construction.

So you will see when you open grade and content level documents that there are different types of ALDs and different levels of detail.

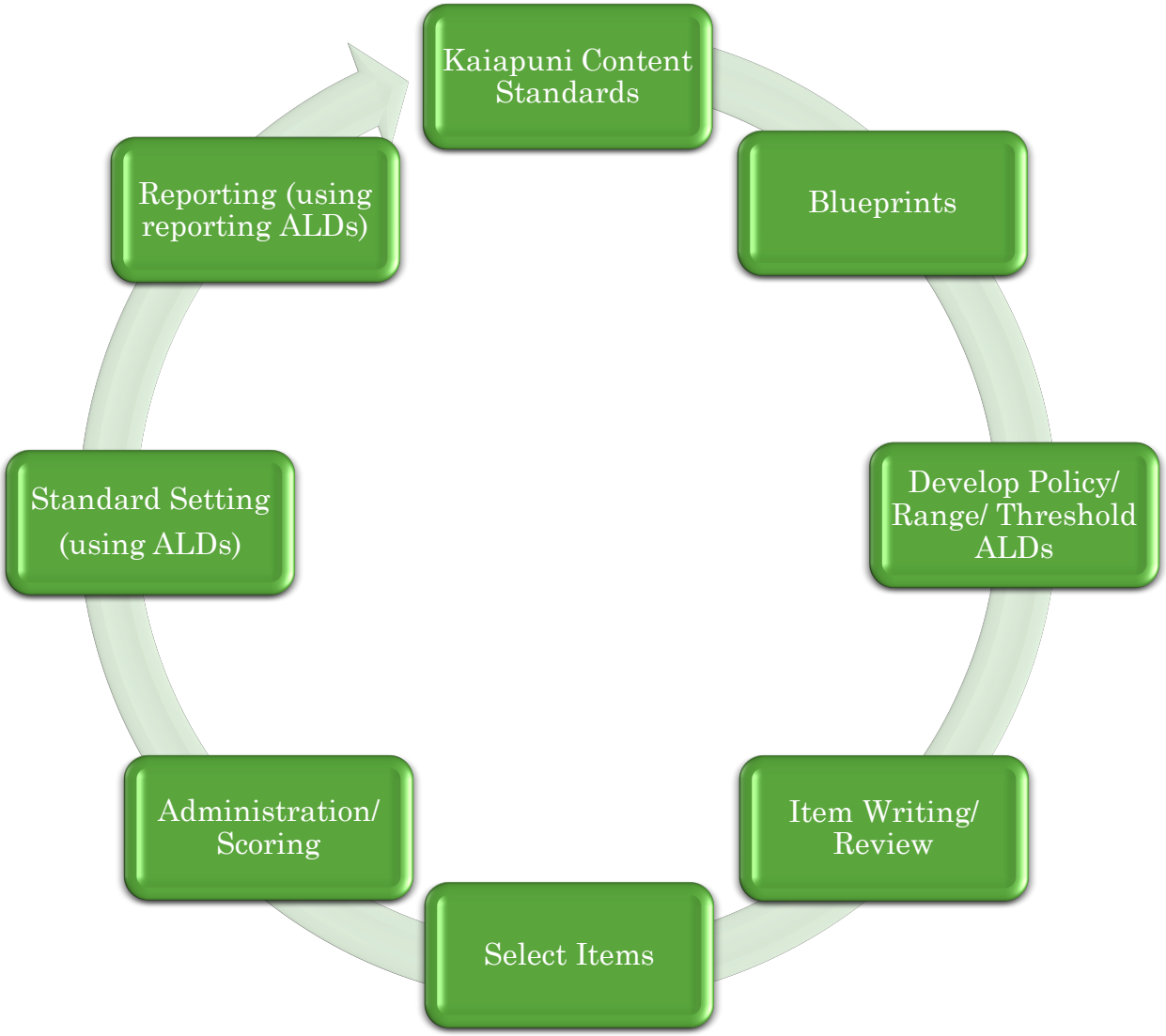
When you look at the range ALDs you will see 'Targets'. These are the statements that are one level up from the 'standards'.

Range ALDs are written for each Target. Targets are the content statements right above the standards in the nested structure.

ALD Type	Purpose	Typical Stakeholder Audience	Notes
Policy ALDs	Written at a higher level that generally state goals and rigor for performance	Policy makers	Typically written as one set for use across the assessment program (not for each grade level)
Range ALDs	Describe the range of knowledge, skills, and cognitive processes required for *within* each of the performance levels	Item writers	Usually written at the beginning of the test cycle. These are cumulative (e.g., if a student is in level 3, they have met the requirements of levels 1 and 2)
Threshold ALDs	State the minimum level of knowledge and skills required to advance to the next performance level	Standard setting participants	As with range ALDs, these represent cumulative skills.
Reporting ALDs	These are written based on the results of standard setting and help stakeholders interpret assessment scores		
General	These are the ALDs that will be on the reports	Parents, policy makers	
Detailed	These expand upon the general reporting ALDs by providing more specific information	Teachers, administrators	

Overview of ALD Development Process

It is helpful to look at ALDs as part of the alignment study because of the integral role they play in test development.



Content Review (discussion by content area)

On Monday, we talked about the backbone of the assessment and that there were certain documents that really helped establish content validity (e.g., standards, items, blueprint). And making sure those documents were all consistent and aligned.

When reviewing the ALDs, note any areas where they don't progress smoothly (e.g., there is a big jump in knowledge from level 1 – 2 but then it levels off from 2 – 3). Then read the next ALD and check for the same progression across and then check the alignment between the levels from one row to the next.

Consider the following questions about the ALDs:

1) Do they capture the important content in the standards (is there content you would suggest be added/deleted in the ALDs?)?

2) Do they represent important progression in skills from one level to the other?

3) Do they adequately represent the skills you saw in the items (is there content you would suggest be added/deleted in the ALDs?)?

4) Do the ALDs communicate important information to stakeholders?

5) Do they progress in a meaningful way from grade 3 to 4?

6) Helpful for the stakeholders (e.g., parents, teachers)?

Ka Papahana Loiloi Kaiapuni

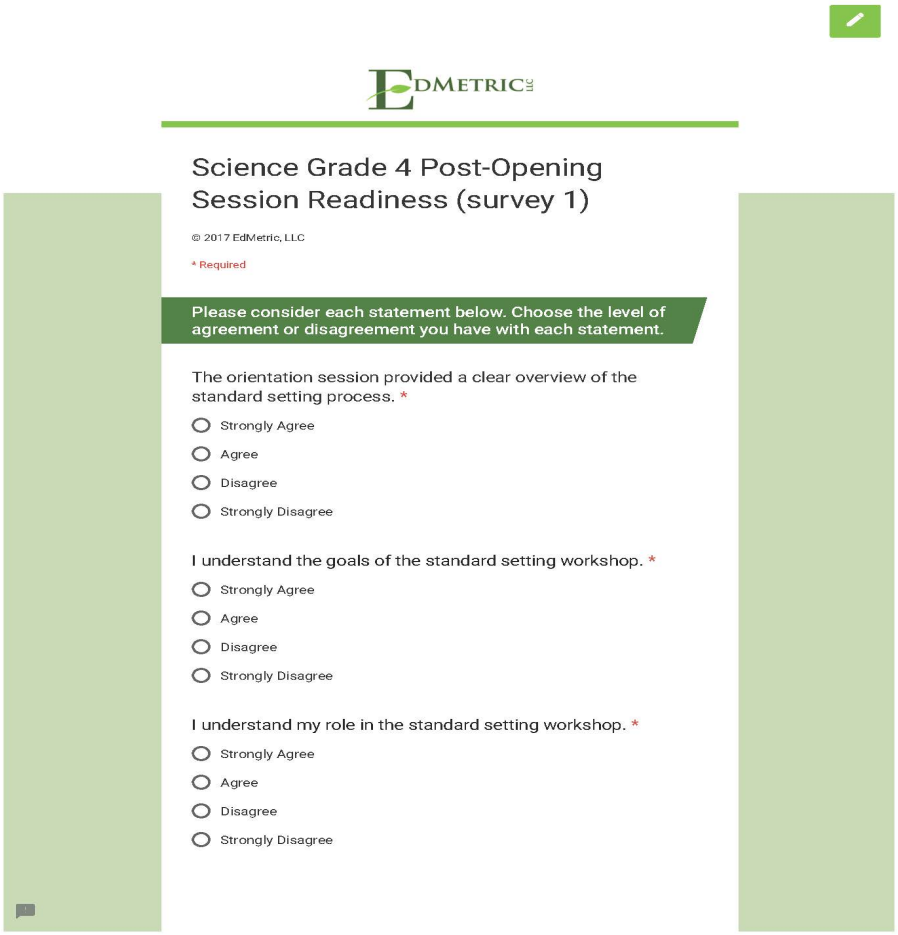
Standard Setting High-Level Agenda[1]

Monday, July 17: Day 1	
Monday	**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:** Discuss and revise Range PLDs **Noon:** Lunch **1:00 PM:** Create Threshold PLDs (Target Student Descriptors) Complete Grade 3 operational form **11:00 AM:** Study ordered item booklet (OIB) **1:00 PM:** Continue study of the OIB **4:00 PM:** Session Close
Tuesday, July 18: Day 2	
Tuesday	**9:00 AM**: Complete Operational Form **10:00 AM:** Study ordered item booklet (OIB) **Noon:** Lunch **1:00 PM:** Continue study of the OIB **3:00 PM:** Bookmark Training **3:45 PM:** Round 1 Ratings **4:00 PM:** Secure materials collection .

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

Wednesday, July 19: Day 3	
Wednesday	**9:00 AM:** Discussion of Round 1 results and Round 2 ratings **10:30 AM:** Discussion of Round 2 results and Round 3 ratings **11:30 AM:** Orientation to Reporting PLDs **Noon:** Lunch **1:00 PM:** Create Reporting PLDs **3:00 PM:** Break **3:50 PM:** Secure materials collection. **4:00 PM**: Close

SURVEYS



EDMETRIC LLC

Science Grade 4 Post-Opening Session Readiness (survey 1)

* Required

Please consider each statement below. Choose the level of agreement or disagreement you have with each statement.

The orientation session provided a clear overview of the standard setting process. *

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

I understand the goals of the standard setting workshop. *

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

I understand my role in the standard setting workshop. *

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

I understand the purpose of the Range ALDs. *

- ○ Strongly Agree
- ○ Agree
- ○ Disagree
- ○ Strongly Disagree

I understand the purpose of the Threshold ALDs. *

- ○ Strongly Agree
- ○ Agree
- ○ Disagree
- ○ Strongly Disagree

The training on achievement level descriptors was helpful to me. *

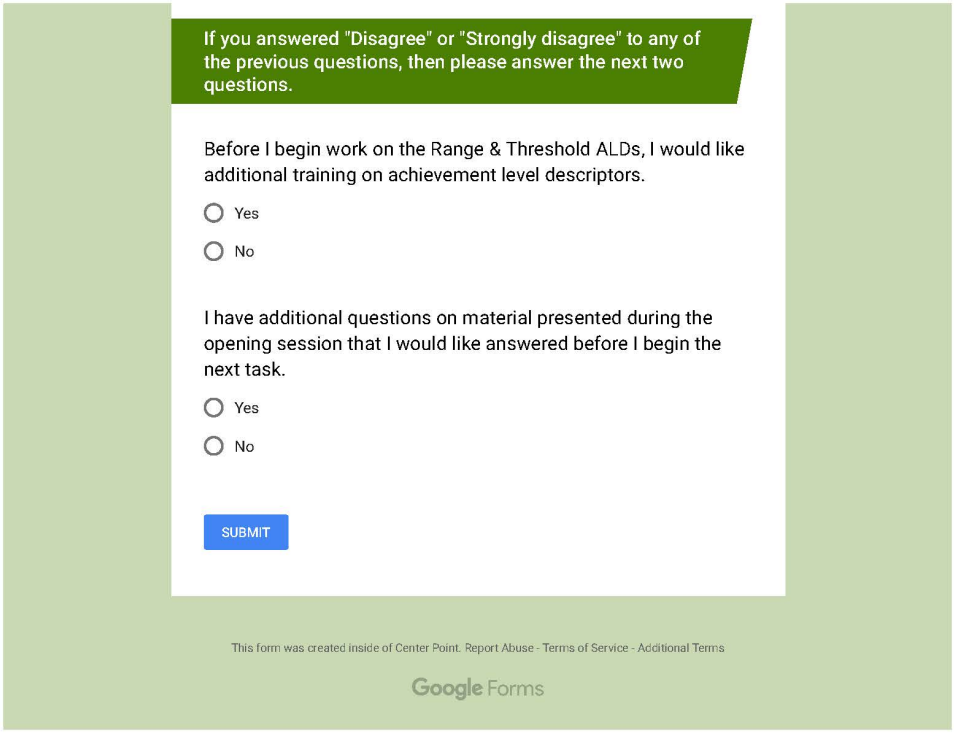
- ○ Strongly Agree
- ○ Agree
- ○ Disagree
- ○ Strongly Disagree

I understand the steps necessary to begin working on Threshold ALDs. *

- ○ Strongly Agree
- ○ Agree
- ○ Disagree
- ○ Strongly Disagree

I understand that I will receive additional training throughout the workshop. *

- ○ Strongly Agree
- ○ Agree
- ○ Disagree
- ○ Strongly Disagree


If you answered "Disagree" or "Strongly disagree" to any of the previous questions, then please answer the next two questions.
Before I begin work on the Range & Threshold ALDs, I would like additional training on achievement level descriptors.
Yes
No
I have additional questions on material presented during the opening session that I would like answered before I begin the next task.
Yes
No
SUBMIT
This form was created inside of Center Point. Report Abuse - Terms of Service - Additional Terms
Google Forms

Science Grade 4 Post-Bookmark Training Readiness (survey 2)

* Required

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

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

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

I understood how to study items in the ordered item booklet. *

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

I understand how to place my bookmarks. *

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

I understand I will have opportunities to change my bookmarks in Round 2. *

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

If you answered "Disagree" or "Strongly disagree" to any of the previous questions, then please answer the next two Yes/No questions.

I would like additional training on placing my bookmarks for Round 1.

- ○ Yes
- ○ No

I have additional questions that I would like to ask before placing my Round 1 bookmarks.

- ○ Yes
- ○ No

SUBMIT

EDMETRIC LLC

Evaluation of the Standard Setting Workshop (survey 3)

* Required

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

I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking. *

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

My group shared a common understanding of the Threshold ALDs. *

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

I used the Threshold ALDs to help me place my cut scores. *

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

During Round 1, I placed my cut scores independently. *

- Strongly Agree
- Agree
- Disgree
- Strongly Disagree

I had enough time to consider the placement of my cut scores. *

- Strongly Agree
- Agree
- Disgree
- Strongly Disagree

I am satisfied with the Threshold ALDs. *

- Strongly Agree
- Agree
- Disgree
- Strongly Disagree

I am satisfied with the Range PLDs. *

- Strongly Agree
- Agree
- Disgree
- Strongly Disagree

The policy definitions were clearly communicated. *

- Strongly Agree
- Agree
- Disgree
- Strongly Disagree

I understood how to place my cut scores. *

- ○ Strongly Agree
- ○ Agree
- ○ Disagree
- ○ Strongly Disagree

I had enough time to consider the placement of my cut scores. *

- ○ Strongly Agree
- ○ Agree
- ○ Disagree
- ○ Strongly Disagree

I am satisfied with our draft Reporting ALDs. *

- ○ Strongly Agree
- ○ Agree
- ○ Disagree
- ○ Strongly Disagree

I understand the Reporting ALDs will be finalized after the workshop. *

- ○ Strongly Agree
- ○ Agree
- ○ Disagree
- ○ Strongly Disagree

I feel the recommended cut scores that resulted from this process are reasonable. *

- ○ Strongly Agree
- ○ Agree
- ○ Disagree
- ○ Strongly Disagree

I would be able to defend the panel's recommended Level 3 cut scores against criticism that they are too high. *

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

I would be able to defend the panel's recommended Level 3 cut scores against criticism that they are too low. *

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

I would be able to defend the panel's recommended Level 4 cut scores against criticism that they are too high. *

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

I would be able to defend the panel's recommended Level 4 cut scores against criticism that they are too low. *

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

I would be able to defend the panel's recommended Level 2 cut scores against criticism that they are too high. *

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

I would be able to defend the panel's recommended Level 2 cut scores against criticism that they are too low. *

- Strongly Agree
- Agree
- Disgree
- Strongly Disagree

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

- Strongly Agree
- Agree
- Disgree
- Strongly Disagree

Overall, I valued the workshop as a professional development experience. *

- Strongly Agree
- Agree
- Disgree
- Strongly Disagree

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

- Strongly Agree
- Agree
- Disgree
- Strongly Disagree

Participating in the workshop increased my understanding of the KAEO assessments. *

- Strongly Agree
- Agree
- Disgree
- Strongly Disagree

The food and service at the facility met my expectations. *

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

The work space had accommodations appropriate to facilitate our work. *

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

The workshop was well organized. *

- ○ Strongly Agree
- ○ Agree
- ○ Disgree
- ○ Strongly Disagree

Which of the following best describes your current position? *

- ○ Classroom Teacher
- ○ School Administrator
- ○ Non-classroom Teacher
- ○ Curricular and/or Instructional Facilitator
- ○ Other: _____

How many years have you been in your current profession? *

Your answer

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

- [ ] Special Education
- [ ] English Language Learner
- [ ] Hawaiian Language Learner
- [ ] Title I Reading
- [ ] Vocational Eduction
- [ ] Adult Education
- [ ] Other: _____

What is your gender? *

- ○ Female
- ○ Male
- ○ Prefer not to say
- ○ Other: _____

Are you of Hispanic origin? *

- ○ Yes
- ○ No

What is your race? *

- ○ Asian/Pacific Islander
- ○ Black/African American
- ○ American Indian
- ○ White
- ○ Multi-racial
- ○ Prefer not to say

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

Your answer

SUBMIT

Page 1 of 1

Never submit passwords through Google Forms.

APPENDIX C. TABLE FACILITATOR TRAINING AND MATERIALS

Training 1 KĀʻEO ALD Table Facilitator Training

Training 1

Kaiapuni Assessment of Educational Outcomes (KĀʻEO)

Achievement Level Descriptors

Table Facilitator Training

Overview of Training

Training 1: ALDs

- Define Achievement Level Descriptors (ALDs)
- Discuss types of ALDs
- Discuss role of ALDs at the standard setting workshop

Training 2: Standard Setting

- Overview standard setting
- Define and demonstrate standard setting methodology

REMEMBER

ALD

=

Achievement Level Descriptor

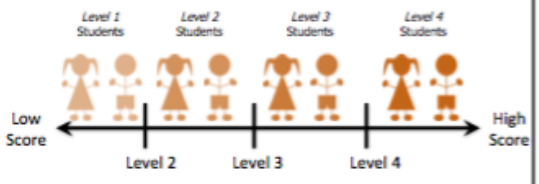
What are Achievement Level Descriptors?

- Define the types of knowledge, skills, and processes **expected of or demonstrated by** students in each level of performance

ALDs and Student Performance

- Test scale represents range of student scores
- Cut scores will separate students into achievement levels
- KĀʻEO : 3 cut score, 4 achievement levels

ALDs Define Student Performance



What are Achievement Level Descriptors?

- Aggregate students into groups based on test performance
  Level 4, Level 3, Level 2, Level 1
- Define the types of knowledge, skills, and processes **expected of or demonstrated by** students in each level of performance
- Connect the assessment and content standards

EDMETRIC

Who Uses ALDs?


Policy Makers


Educators


Item Writers


Parents

EDMETRIC

Family of ALDs

Policy ALDs: Broad statements that set the tone and rigor for the other ALDs
- Used by Policymakers

Range ALDs: Expectations for whole of student performance in an achievement level
- Specific to a grade and content area.
- Utilized by item writers and educators

Threshold ALDs: Expectations for student performance at the entry point of an achievement level.
- Specific to a grade and content area.
- Utilized by standard setters.

Reporting ALDs: Students' performance at the entry point of an achievement level.
- Specific to a grade and content area.
- Utilized by stakeholders.

EDMETRIC

POLICY DESCRIPTORS

EDMETRIC

10

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 accountability purposes.

EDMETRIC

RANGE ALDS

EDMETRIC

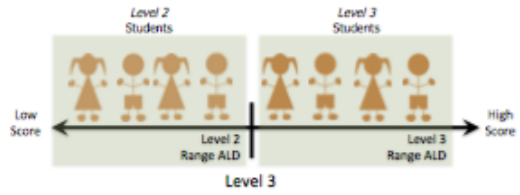
12

Range ALDs

- Define expected knowledge and skills of all students in a particular achievement level
- Depict the expected grade-level progression by content standard
- Guide item development
- Guide curriculum development

Range ALDs for Level 3

Range ALDs

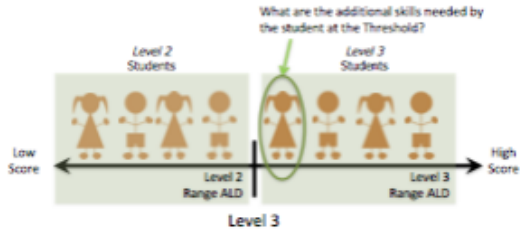


Range ALDs

- Describe *range* of knowledge, skills, and processes for all students in each achievement level
  - Level 1
  - Level 2
  - Level 3
  - Level 4
- During the workshop, the existing Range ALDs will be reviewed and revised by the standard setting panel

THRESHOLD ALDS

Range ALDs to Threshold ALDs



Threshold ALD

Threshold ALDs

- During the workshop, you will work with other panelists to create the Threshold ALDs
- Threshold ALDs will guide your work during standard setting

REPORTING ALDS

Threshold ALDs to Reporting ALDs

- Threshold ALDs guide the standard setting
- Reporting ALDs are based on the final cut scores and on test items
- You will create Reporting ALDs at the workshop
- Additional training will be provided immediately prior to creating the Reporting ALDs

Reporting ALDs

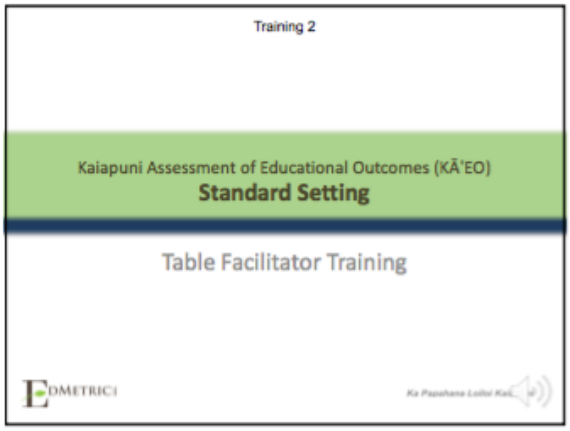
- Reconciles the threshold ALDs with the final cut scores
- Defines the knowledge and skills of students who have entered the achievement level

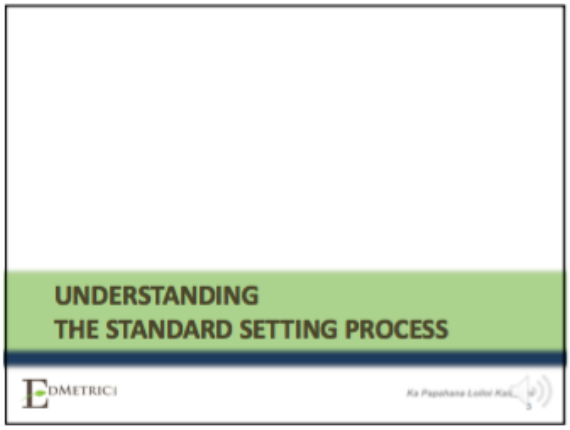
ALDS AND STANDARD SETTING

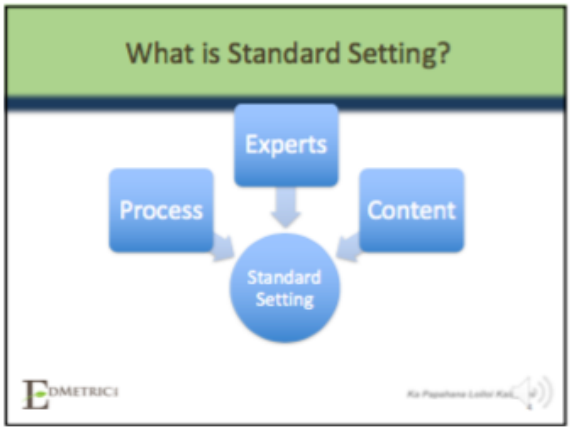

Standard Setting
After Standard Setting
EDMETRIC
Adapted from Egan, Schneider, & Ferrara (2012)

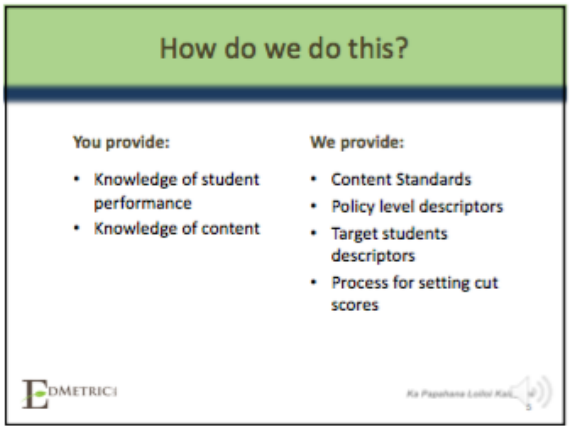

Questions?
1. Email:
karla_egan@edmetricllc.com
2. Q & A
EDMETRIC

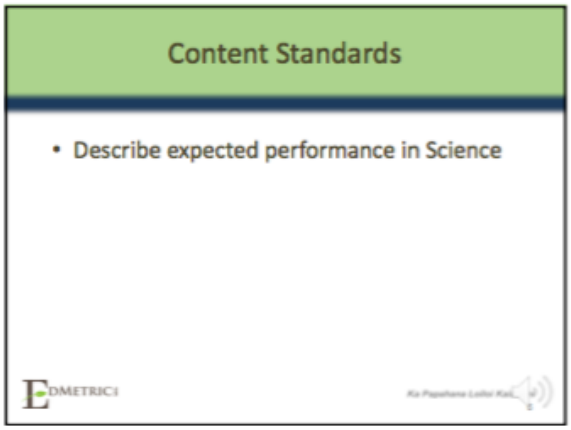
Training 2 KĀʻEO BSSP Table Facilitator Training











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 accountability purposes.

EDMETRICS

KĀʻEO Cut Scores

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

EDMETRICS

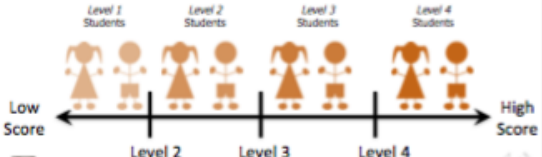
How do we set cut scores?

- Pr[illegible]s
  - Ar[illegible]rary
  - Test s[illegible]ic
  - [illegible] is not considered
- Content
  - Use pre-established content standards
  - Considers educational objectives
- Bookmark Standard Setting Procedure

EDMETRICS

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



EDMETRICS

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

EDMETRICS

Developing Expertise

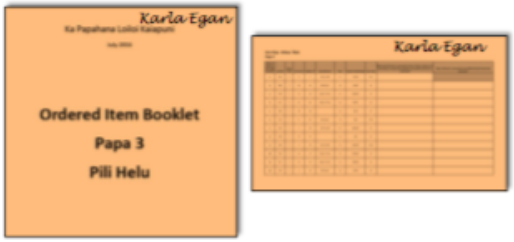
- You know students and content
- Activities to enhance knowledge:
  1. Study content standards
  2. Create Threshold ALDs
  3. Study test items
     - Take test
     - Study ordered item booklets using item maps

EDMETRICS

Bookmark Standard Setting Process

EDMETRICS

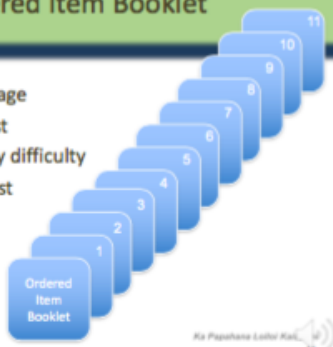
Bookmark Materials



EDMETRICS

Ordered Item Booklet

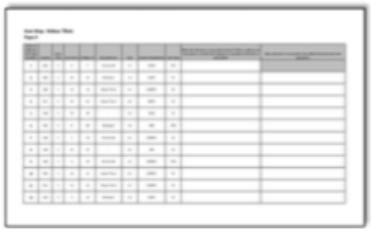
- One item per page
- Easiest item first
- Items ascend by difficulty
- Hardest item last



EDMETRICS

Item Map: Key Questions

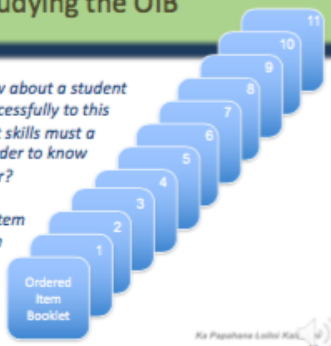
(1) What does this item measure?
(2) Why is this item more difficult than the items that precede it?



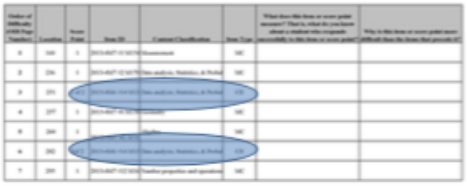
EDMETRICS

Studying 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?*



EDMETRICS



EDMETRICS

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

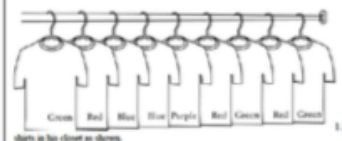
1. How much do these apples weigh?

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

EDMETRICS

Ka Papahana Loiloi Kaiapuni 19

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

EDMETRICS

Ka Papahana Loiloi Kaiapuni 20

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

EDMETRICS

Ka Papahana Loiloi Kaiapuni 21

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.

EDMETRICS

Ka Papahana Loiloi Kaiapuni 22

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.

EDMETRICS

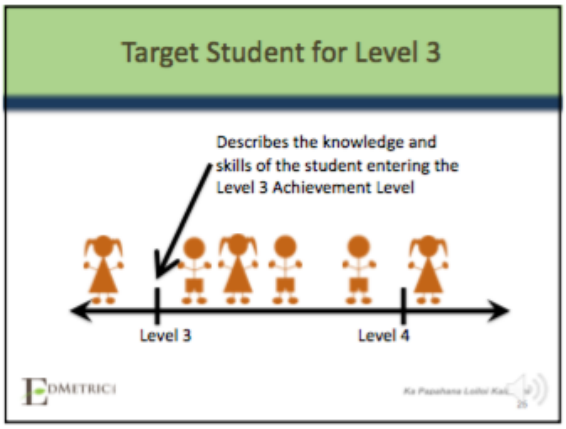
Ka Papahana Loiloi Kaiapuni 23

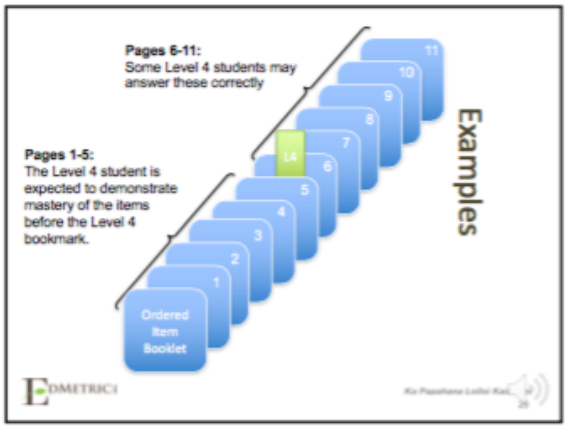
The Bookmark Process

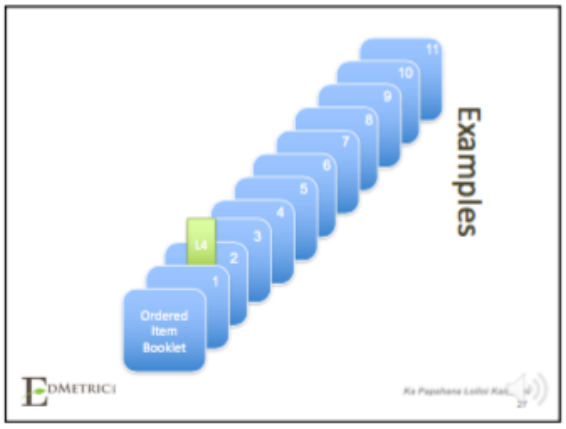
- Study content in a book
- Tell a story about the content students in each level should be able to do
- Set a bookmark

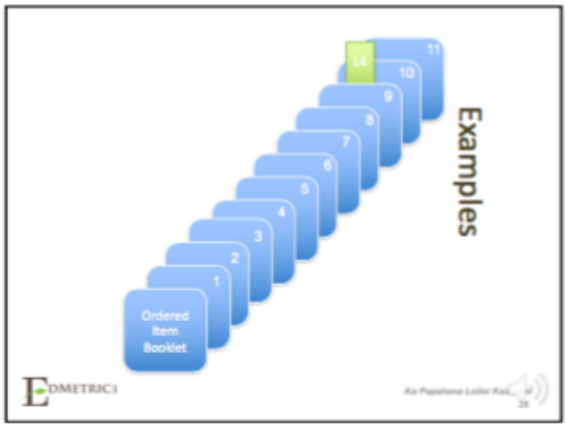
EDMETRICS

Ka Papahana Loiloi Kaiapuni 24

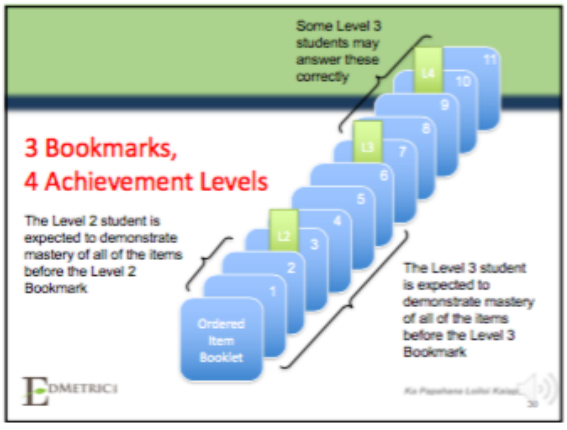

Target Student for Level 3
Describes the knowledge and skills of the student entering the Level 3 Achievement Level
Level 3
Level 4


Pages 6-11:
Some Level 4 students may answer these correctly
Pages 1-5:
The Level 4 student is expected to demonstrate mastery of the items before the Level 4 bookmark.
Examples
Ordered Item Booklet


Examples
Ordered Item Booklet


Examples
Ordered Item Booklet


3 Bookmarks,
4 Achievement Levels
Ordered Item Booklet


Some Level 3 students may answer these correctly
3 Bookmarks,
4 Achievement Levels
The Level 2 student is expected to demonstrate mastery of all of the items before the Level 2 Bookmark
The Level 3 student is expected to demonstrate mastery of all of the items before the Level 3 Bookmark
Ordered Item Booklet

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

Understanding Your Role as Table Facilitator

Table Facilitator Responsibilities

- Facilitate discussion at the table level
  - Studying the OIB
  - Leading table level discussions (Rounds 2 and 3)
- Bring important questions to the attention of the room facilitator
- Check understanding at table
- Keep an eye on time
- Monitor and collect secure materials at your table

Table Facilitator Responsibilities

- Facilitate discussion at the table level
  - Studying the OIB
    - Facilitate conversation around the two questions
      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?"*
    - Remind panelists to put thoughts into comments
    - Monitor time for this task
    - Discourage sidebar conversation
    - Try to get all panelists to participate in discussion

Table Facilitator Responsibilities

- Facilitate discussion at the table level
  - Round 2
  - Round 3

Table Facilitator Responsibilities

- Monitor and collect secure materials at your table
  - At the end of each day we will collect materials.
  - You will have a list of secure materials. Have panelists stack materials according to the list.
  - This allows us to have an efficient and orderly collection of materials to support security.

Agenda

Day 1
- **Morning** | Opening Session; Study Content Standards; Review & Revise Range ALDs
- **Afternoon** | Create Threshold ALDs; Take Operational Test

Day 2
- **Morning** | Study OIB
- **Afternoon** | Study OIB, Round 1 Bookmarks

Day 3
- **Morning** | Round 2, Round 3
- **Afternoon** | ALD Writing, Evaluation

EDMETRICS

Pre-Reading

- Overview of standard setting
- Annotated agenda

EDMETRICS

Questions?

1. Email:

   karla_egan@edmetricllc.com

2. Q & A

EDMETRICS

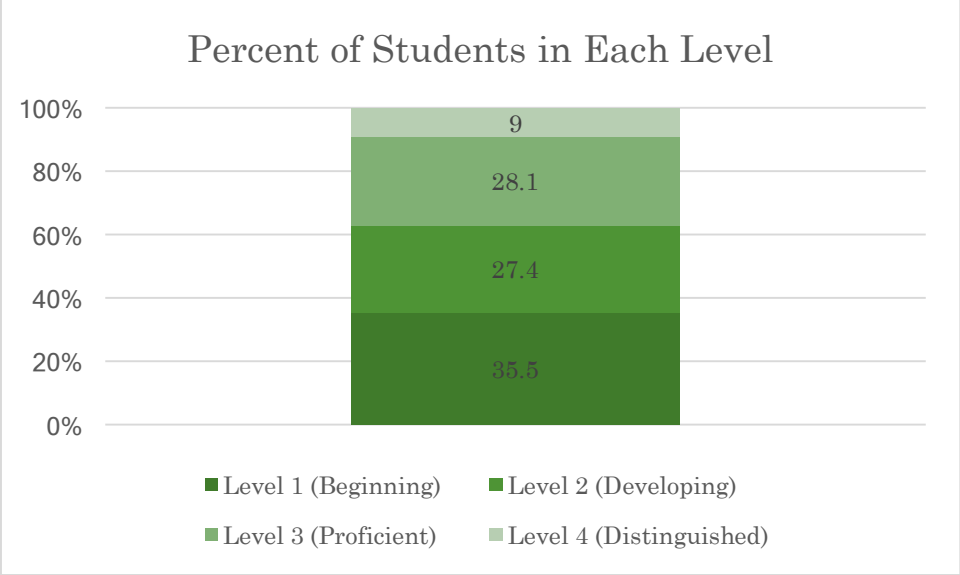
APPENDIX D. DETAILED STANDARD SETTING RESULTS

FINAL CUT SCORES

Achievement Level	Recommended Cut Score
Developing	480
Proficient	519
Distinguished	562

	Beginning	Developing	Proficient	Distinguished
KAEO Science	LOSS-479	480-518	519-561	562-HOSS

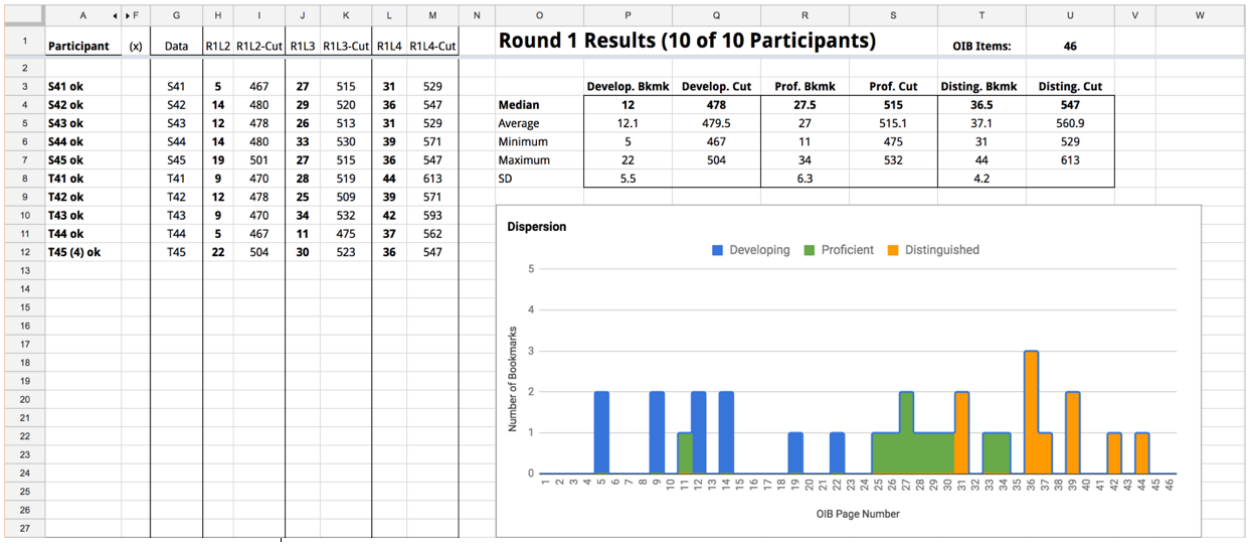
FINAL IMPACT DATA



ROUND BY ROUND RESULTS

Round 1 Bookmarks

KAEO Science Round 1



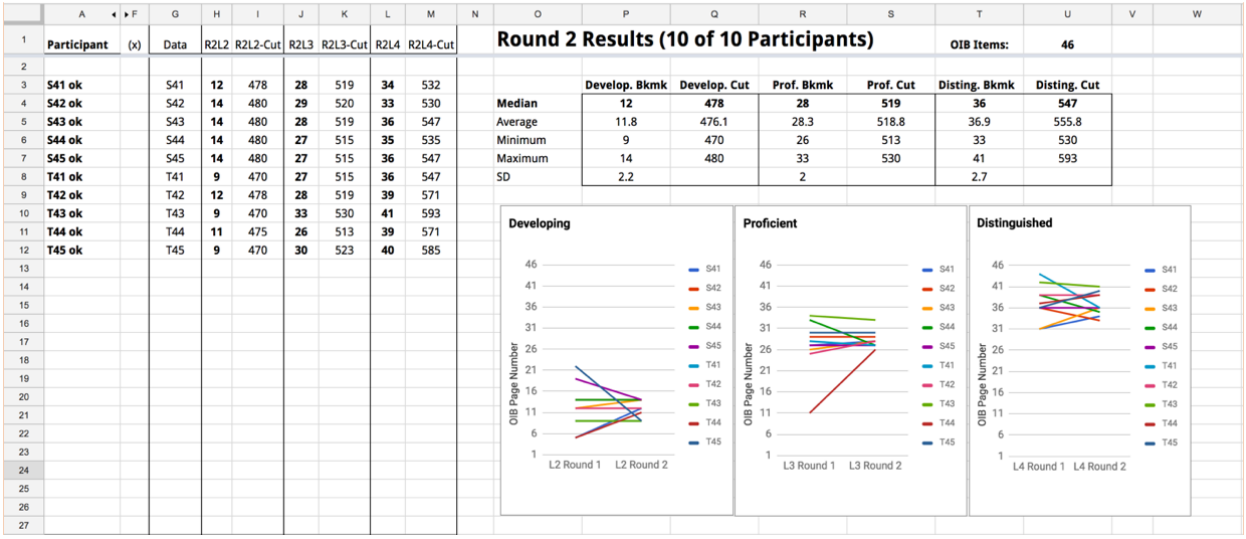
Participant	(x)	Data	R1L2	R1L2-Cut	R1L3	R1L3-Cut	R1L4	R1L4-Cut
S41 ok		S41	**5**	467	**27**	515	**31**	529
S42 ok		S42	**14**	480	**29**	520	**36**	547
S43 ok		S43	**12**	478	**26**	513	**31**	529
S44 ok		S44	**14**	480	**33**	530	**39**	571
S45 ok		S45	**19**	501	**27**	515	**36**	547
T41 ok		T41	**9**	470	**28**	519	**44**	613
T42 ok		T42	**12**	478	**25**	509	**39**	571
T43 ok		T43	**9**	470	**34**	532	**42**	593
T44 ok		T44	**5**	467	**11**	475	**37**	562
T45 (4) ok		T45	**22**	504	**30**	523	**36**	547

Round 1 Results (10 of 10 Participants) — **OIB Items: 46**

	Develop. Bkmk	Develop. Cut	Prof. Bkmk	Prof. Cut	Disting. Bkmk	Disting. Cut
Median	**12**	**478**	**27.5**	**515**	**36.5**	**547**
Average	12.1	479.5	27	515.1	37.1	560.9
Minimum	5	467	11	475	31	529
Maximum	22	504	34	532	44	613
SD	5.5		6.3		4.2	

Bookmark Report – Round 2

KAEO Science Round 2



Participant	(x)	Data	R2L2	R2L2-Cut	R2L3	R2L3-Cut	R2L4	R2L4-Cut
S41 ok		S41	**12**	478	**28**	519	**34**	532
S42 ok		S42	**14**	480	**29**	520	**33**	530
S43 ok		S43	**14**	480	**28**	519	**36**	547
S44 ok		S44	**14**	480	**27**	515	**35**	535
S45 ok		S45	**14**	480	**27**	515	**36**	547
T41 ok		T41	**9**	470	**27**	515	**36**	547
T42 ok		T42	**12**	478	**28**	519	**39**	571
T43 ok		T43	**9**	470	**33**	530	**41**	593
T44 ok		T44	**11**	475	**26**	513	**39**	571
T45 ok		T45	**9**	470	**30**	523	**40**	585

Round 2 Results (10 of 10 Participants) — **OIB Items: 46**

	Develop. Bkmk	Develop. Cut	Prof. Bkmk	Prof. Cut	Disting. Bkmk	Disting. Cut
Median	**12**	**478**	**28**	**519**	**36**	**547**
Average	11.8	476.1	28.3	518.8	36.9	555.8
Minimum	9	470	26	513	33	530
Maximum	14	480	33	530	41	593
SD	2.2		2		2.7	

Bookmark Report – Round 3

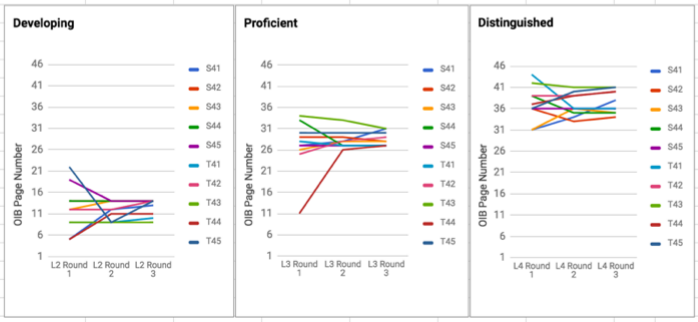
KAEO Science Round 3 (Without Impact Data)

Participant	(x)	Data	R3L2	R3L2-Cut	R3L3	R3L3-Cut	R3L4	R3L4-Cut
S41 ok		S41	**13**	479	**31**	529	**38**	567
S42 ok		S42	**14**	480	**28**	519	**34**	532
S43 ok		S43	**14**	480	**28**	519	**35**	535
S44 ok		S44	**14**	480	**27**	515	**35**	535
S45 ok		S45	**14**	480	**27**	515	**36**	547
T41 ok		T41	**10**	474	**27**	515	**36**	547
T42 ok		T42	**14**	480	**29**	520	**40**	585
T43 ok		T43	**9**	470	**31**	529	**41**	593
T44 ok		T44	**11**	475	**27**	515	**40**	585
T45 ok		T45	**14**	480	**30**	523	**41**	593

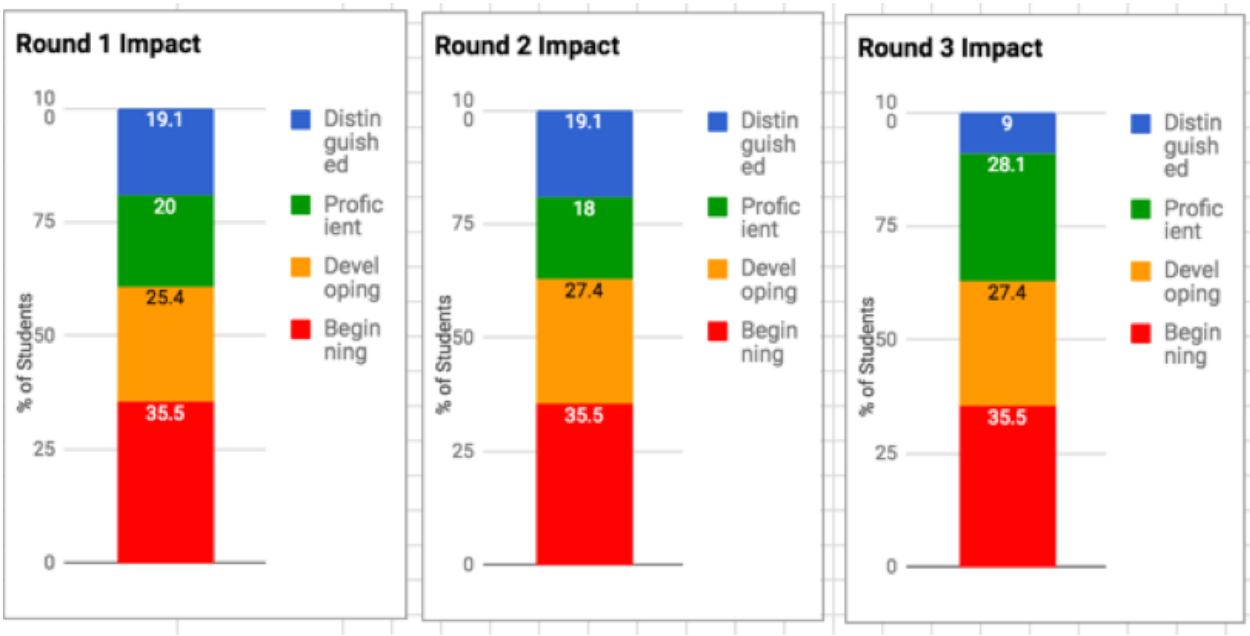
Round 3 Results (10 of 10 Participants)

OIB Items: 46

	Develop. Bkmk	Develop. Cut	Prof. Bkmk	Prof. Cut	Disting. Bkmk	Disting. Cut
Median	**14**	**480**	**28**	**519**	**37**	**562**
Average	12.7	477.8	28.5	519.9	37.6	561.9
Minimum	9	470	27	515	34	532
Maximum	14	480	31	529	41	593
SD	1.9		1.6		2.7	



Impact Data by Round



APPENDIX E. DETAILED RESULTS OF THE PANELIST EVALUATION

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

Statement	Strongly Agree	Agree
I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking.	90	10
My group shared a common understanding of the threshold ALDs.	80	20
I used the Threshold ALDs to help me place my cut scores.	100	0
During Round 1, I placed my cut scores independently.	100	0
I had enough time to consider the placement of my cut scores.	100	0
I am satisfied with the Threshold ALDs.	90	10
I am satisfied with the Range PLDs.	80	20
The policy definitions were clearly communicated.	90	10
I understood how to place my cut scores.	80	20
I had enough time to consider the placement of my cut scores.	90	10
I am satisfied with our draft Reporting ALDs.	70	30
I understand the Reporting ALDs will be finalized after the workshop.	100	0
I feel the recommended cut scores that resulted from this process are reasonable.	80	20
I would be able to defend the panel's recommended Level 3 cut scores against criticism that they are too high.	50	50
I would be able to defend the panel's recommended Level 3 cut scores against criticism that they are too low.	50	50
I would be able to defend the panel's recommended Level 4 cut scores against criticism that they are too high.	60	40
I would be able to defend the panel's recommended Level 4 cut scores against criticism that they are too low.	50	50

Statement	Strongly Agree	Agree
I would be able to defend the panel's recommended Level 2 cut scores against criticism that they are too high.	50	50
I would be able to defend the panel's recommended Level 2 cut scores against criticism that they are too low.	50	50
Overall, I believe that my opinions were considered and valued by my group.	90	10
Overall, I valued the workshop as a professional development experience.	100	0
This experience will help me target instruction for the students in my classroom.	100	0
Participating in the workshop increased my understanding of the KAEO assessments.	100	0
The food and service at the facility met my expectations.	100	0
The work space had accommodations appropriate to facilitate our work.	100	0
The workshop was well organized.	100	0

Table E.2. Number of Panelists disaggregated by Educator role

Educator Role	Frequency
Classroom Teacher	9
Curricular And/Or Instructional Facilitator	1

Table E.3. Number of Panelists disaggregated by Gender

Gender	Frequency
Male	1
Female	9

Table E.4. Mean, Minimum, and Maximum Number of Years Panelists were in current position

N	Mean	Std. Deviation	Minimum	Maximum
10	11.1	5.76	3	21

Table E.5. Panelists' areas of expertise

Area of Expertise	Frequency*
Hawaiian Language Learner	10
ELL	4
Special Education	1
Title I Reading	1
Adult Education	1

** Based on responses of 10 panelists. Panelists could indicate more than one area of expertise.*

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

Island and School	Total
Hawai'i	
Ke kula ʻo ʻEhunuikaimalino	1
Kauai	
Kawaikini New Century Public Charter School	1
Maui	
Ke Kula Kaiapuni ʻo Pāʻia	1
Kula Kaiapuni ʻo Nāhiʻenaʻena	1
Molokai	
Kula Kaiapuni o Kualapuʻu	1
Oahu	
Ke Kula Kaiapuni ʻo Waiau	1

Kula Kaiapuni ʻo Ānuenue	2
Ke Kula Kaiapuni o Pūʻōhala	1
Ke Kula Kaiapuni o Hauʻula	1
Total	**10**

Table E.7. Panelists' Comments

Panelists' Comments
Mahalo to Karla, Pōhai, Pono and all of the other limahana for all that you folks did to help make this workshop successful!
Mahalo for helping us through this process!
Mahalo
Mahalo piha i kā ʻoukou hana nui e hoʻomākaukau i nā mea e pono ai!
Mahalo nui!
The work is huge - but my understanding of the process is becoming clearer. I appreciate the time and effort put into having the information available to all in attendance. When the ALDs are finalized, how can we access a copy?
This was very valuable :)
Please ask me to come back. This was fun!
Mahalo nui no ka ho'onui 'ike.....a me ka mea'ai 'ono loa kekahi!