



United Nations  
Educational, Scientific and  
Cultural Organization



UNESCO  
INSTITUTE  
FOR  
STATISTICS



GLOBAL  
ALLIANCE  
TO MONITOR  
LEARNING

# Constructing UIS proficiency scales and linking to assessments to support SDG Indicator 4.1.1 reporting

GAML4/REF/8

This technical concept paper was prepared to contribute to the discussion and understanding of the important development challenges facing policymakers and practitioners

**Prepared by:**  
**Management Systems International (MSI)**

October 2017



## TABLE OF CONTENTS

<b>ACRONYMS</b> .....	3
<b>INTRODUCTION</b> .....	4
<b>PROCEDURES</b> .....	5
Constructing UIS Proficiency Scales (UIS-PSs) .....	5
Step 1: Define Content Standards .....	5
Step 2: Determine Performance Levels .....	5
2a: Decide Number of Performance Levels and Labels on the UIS Proficiency Scales .....	5
2b: Write Policy Definitions for the Performance Levels of the UIS Proficiency Scales .....	6
Step 3: Develop Full Descriptions for the Performance Levels of the UIS Proficiency Scales .....	7
Linking the UIS Proficiency Scales with NAs and CNAs .....	9
Step 1: Evaluate Alignment of Performance Level Descriptors (PLDs).....	10
Step 2: Set Socially Moderated Performance Standards for NAs and CNAs .....	11
<b>MORE INFORMATION</b> .....	12
Construction of the UIS-PSs .....	12
Linking UIS-PSs with NAs and CNAs.....	12
Tentative Timeline .....	12
<b>REFERENCES</b> .....	14



## Acronyms

---

UIS	UNESCO Institute for Statistics
SDG	Sustainable Development Goals
ACER	Australian Council for Educational Research
GAML	Global Alliance to Monitor Learning
PASEC	Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN
SACMEQ	Southern Africa Consortium for Monitoring Educational Quality
TIMSS	Third International Mathematics and Science Study
PIRLS	Progress in International Reading Literacy Study



## Introduction

UIS' goal, as custodial agency for reporting against the SDGs, is to develop reporting scales that will support national governments to effectively measure and monitor student learning outcomes in mathematics and reading against SDG Indicator 4.1.1 over time, and to utilize the data for making informed policy decisions. It is a further goal to support the use of existing national assessments (NAs) and cross-national assessments (CNAs) to facilitate measurement and reporting of learning outcomes, rather than requiring a single assessment be used by all countries for SDG reporting purposes.

The GAML and its technical partner, the ACER Centre for Global Education Monitoring, are exploring the development of reporting scales—the UIS Reporting Scales, or UIS-RS—that would achieve these goals. Draft learning progressions have been developed and are currently undergoing broad review, and a proposal for validating the scales has been put forward for consideration. There is broad support for continuing the work to explore the development of these reporting scales, but also recognition that this is a longer-term effort.

Another promising effort that is being pursued is the statistical linking of TIMSS and regional assessments in 2019, when TIMSS and several of the regional assessments will be administered. As similar exercise could be done for PIRLS (reading), which will be administered in 2021. This would provide TIMSS/PIRLS projected scores for countries in the participating regional assessments. What is proposed in this paper to construct UIS proficiency scales and establish standards does not preclude those proposed linking activities.

A more immediate need in order for UIS and countries to report against Indicator 4.1.1 is to define “minimum proficiency levels” (and ideally additional levels of performance) for reading and mathematics and produce a reporting metric and a mechanism for linking existing assessments and their performance levels to this metric. This paper has been prepared by MSI to provide UIS and the Global Alliance to Monitor Learning (GAML) with a proposed approach to meeting this immediate need.

The paper presents the steps involved in constructing a “UIS proficiency scale”, or UIS-PS, (for each domain and education level in Indicator 4.1.1) and linking NAs and CNAs to them, describing the development of the following:

### Proficiency Scale

- 1) *Content standards*: what students are expected to **learn** in reading and mathematics at the three levels of education defined in Indicator 4.1.1 – grades 2/3, end of primary, and end of lower secondary. (For the purposes of this paper, and reflecting discussions at the September 21-22 Hamburg meeting, these levels will be referred to as by the end of grade 3 (lower primary), 6 (end of primary), and 9 (lower secondary).)
- 2) *Performance levels*: number of levels and names to be used.
- 3) *Policy descriptors*: what students expected are to **perform** (in a generic terms without content) at each level.
- 4) *Performance standards*: what students are expected to **perform** in terms of content (with respect to knowledge, skills, and abilities) at each performance level.

### Linking



- 1) *Proficiency scale map(s)*: how proficiency scales (i.e., performance levels) of various NAs and CNAs map to the UIS proficiency scale.
- 2) *Socially moderated performance standards*: what students should obtain (score) on their NAs and CNAs to be classified into the “desired” performance level for SDG reporting.

Outputs 1-4 are relevant for constructing UIS proficiency scales, and outputs 5-6 for linking of the UIS proficiency scales with other NAs and CNAs.

## Procedures

### **Constructing UIS Proficiency Scales (UIS-PSs)**

The following process is proposed to construct UIS proficiency scales.

#### *Step 1: Define Content Standards*

In order to develop standalone reporting scales for grades 3, 6, and 9 reading and mathematics, the first step is to define the content standards for each domain and for each grade span of K-3, 4-6, and 7-9 separately. The common content standards are predefined knowledge and skills that students are expected to **learn** in reading and mathematics by the end of grades 3, 6, and 9 across countries. The UNESCO IBE has already made significant progress in describing these content standards for each domain and grade. It has already reviewed and analyzed over 140 NAs and CNAs to identify the content standards of various grades being assessed (reference forthcoming).

#### *Step 2: Determine Performance Levels*

##### *2a: Decide Number of Performance Levels and Labels on the UIS Proficiency Scales*

In this step, the number of levels to be used and their names on the scales are determined. This could be done by the GAML. Typically, no more than four performance levels are needed (Perie, 2008). Beyond four levels, it becomes difficult to describe meaningful differences across the levels. Three is probably advisable for UIS proficiency scales. After determining the number of levels, the next task is to name the levels. There are no clear-cut guidelines on how to develop names for the levels, however it is recommended that they be thoughtfully chosen to relate to the purpose of reporting and supportable inferences arising from the classifications (Cizek & Bunch, 2007).

Below are some example labels, based on those used in various assessment programs, which UIS could consider adopting for the UIS proficiency scales:

- does not meet minimum proficiency/partially meets minimum proficiency/**meets minimum proficiency**/exceeds minimum proficiency;<sup>1</sup>
- does not meet standards/partially meets standards/**meets standards**/exceeds standards;
- below basic/basic/**proficient**/advanced;
- beginning step/nearing proficient/**proficient**/advanced;
- level 1/level 2/**level 3**/level 4;
- Novice/apprentice/**proficient**/distinguished.

<sup>1</sup> Although we have proposed labels for four performance levels we recommend that UIS consider have *three* levels, given the complexity of defining proficiency levels for a diverse set of countries to report against and in linking NAs and CNAs to the UIS scales. Four levels may suggest a level of precision that is not supported.



The levels in bold are considered the “desired” level of student performance that policymakers expect all students to achieve.

These are just examples; UIS could choose different labels. However, when naming levels it is strongly recommended avoiding the following types of terms (Beck, 2003):

- Nebulous, unclear, or unreasonable terms (e.g., needs improvement, reasonable mastery);
- Normative terms (e.g., average, typical);
- Moving terms (e.g., nearly X, approaching the standards, emerging, progressing);
- Noneducational terms (e.g., normal, inadequate, novice, apprentice);
- Nonparallel terms (e.g., outstanding, pass, warning).

### *2b: Write Policy Definitions for the Performance Levels of the UIS Proficiency Scales*

The next step is to develop a generic policy definition for each performance level. These definitions are not linked to content but are more general statements that assert policymakers’ position on the desired level of performance. They are particularly useful in the context of reporting multiple assessments. First, they facilitate the articulation of performance levels across grades by ensuring the same level of rigor at each level across each grade. Second, they allow a reader to interpret proficiency in a similar manner regardless of subject assessed. The policy definitions need to be written for each level, not including the lowest level of performance. Figure 1 presents some examples from assessment programs in the United States.

**Figure 1: Illustrative Policy Definitions for Performance Levels**

Assessment	Performance Levels		
National Assessment of Educational Progress (NAEP)	<b>Basic:</b> This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.	<b>Proficient:</b> Solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.	<b>Advanced:</b> This level signifies superior performance beyond proficient.
Pennsylvania Statewide Testing Program	<b>Basic:</b> Marginal academic performance, work approaching, but not yet reaching, satisfactory performance, indicating partial understanding and limited display of the skills included in	<b>Proficient:</b> Satisfactory academic performance indicating a solid understanding and adequate display of the skills included in Pennsylvania’s academic standards.	<b>Advanced:</b> Superior academic performance indicating an in-depth understanding and exemplary display of the skills included in Pennsylvania’s academic standards.

	Pennsylvania's academic standards.		
Arizona Statewide Testing Program	<b>Approaches the Standard:</b> This level denotes understanding of the knowledge and application of the skills that are fundamental for proficiency in the standards.	<b>Meets the Standard:</b> This level denotes demonstration of solid academic performance on challenging subject matter reflected by the content standards. This includes knowledge of subject matter, application of such knowledge to real-world situations, and content-relevant analytical skills. Attainment of at least this level is the expectation for all Arizona students.	<b>Exceeds the Standard:</b> This level denotes demonstration of superior academic performance evidenced by achievement substantially beyond the expected goal of all students.

In writing policy definitions for performance levels, it is strongly recommended that the words used are memorable and distinguish clearly among the levels. The definitions should clearly state the degree of knowledge and skills expected of students at each level. They should be concise, approximately 1–2 sentences, and clear (Perie, 2008). Because it is the backbone of all further writing (i.e., the full descriptions focused on content, as in Step 3, below), UIS should carefully consider the wording and be sure each definition communicates the intended goals and clearly distinguishes one level from the next.

### *Step 3: Develop Full Descriptions for the Performance Levels of the UIS Proficiency Scales*

After the policy definitions have been adopted, content descriptions should be added to develop full descriptions of the performance levels. The full descriptions express the knowledge and skills required to achieve each performance level. They can be used to provide stakeholders with more information on what students at each performance level know and are able to do and what they need to know and be able to do to reach next performance level.

To develop full descriptions, for each domain a performance level descriptor (PLD) writing workshop is conducted with of subject matter experts (SMEs). Five to eight people per subject and grade span will suffice (Perie, 2008). Those with an understanding of the policy context should work alongside those with an understanding of teaching and learning the subject-matter content to write the PLDs. The SMEs will start with the policy definitions (see Step 2b) and expand those definitions in terms of specific knowledge, skills, and abilities at each level for each domain and for each grade. The PLDs should be very detailed and reflect the content standards defined in Step 1.

Here are some examples of Grade 6 English PLDs adapted from an existing U.S. state-wide assessment program:

**Partially Meets Minimum Proficiency:** A student performing at this level demonstrates limited comprehension of literary and informational texts and may use textual evidence to summarize and/or analyze a text. The student inconsistently analyzes how an element of literature or informational text develops and influences the text. The student may



determine a central idea in an informational text. The student may determine how the author uses organization, structure, form, text features, figurative language, and/or word choice to achieve a purpose. The student determines the point of view in a text. The student provides an incomplete comparison between texts in different forms or genres. The student may identify the development of an argument and may evaluate the author's claims and evidence in a text. The student may use context and word structure to determine the meanings of words, may interpret figurative language, and may understand some word meanings. In writing, the student inconsistently uses reasoning and evidence to develop an argumentative/informational essay on a topic for an intended audience. The student organizes a narrative using limited narrative techniques. The student writes a text-dependent analysis essay that responds to a text or texts and demonstrates a weak analysis that may include inadequate evidence to support its intended purpose. The student may use transitions. The student recognizes and demonstrates a partial command of the conventions of standard English grammar, usage, and mechanics.

**Meets Minimum Proficiency:** A student performing at this level demonstrates comprehension of literary and informational texts by using textual evidence to summarize and/or analyze a text. The student analyzes how an element of literature or informational text develops and influences the text. The student determines a central idea in an informational text. The student determines how the author uses organization, structure, form, text features, figurative language, and/or word choice to achieve a purpose. The student determines the effectiveness of point of view in a text. The student compares and contrasts texts in different forms or genres. The student traces the development of an argument and evaluates the author's claims and evidence in a text. The student uses context and word structure to determine the meanings of words, interprets figurative language, and understands nuances in word meanings. In writing, the student uses logical reasoning and relevant evidence to develop an organized argumentative/informational essay on a topic in a formal style for an intended audience. The student organizes a narrative with a controlling point, using precise words, phrases, and narrative techniques. The student writes a text-dependent analysis essay that responds to a text or texts and demonstrates an organized analysis that cites textual evidence to support its intended purpose. The student uses a variety of appropriate transitional words, phrases, and clauses. The student recognizes and demonstrates a command of the conventions of standard English grammar, usage, and mechanics to convey ideas precisely and for effect.

**Exceeds Minimum Proficiency:** A student performing at this level demonstrates thorough comprehension of literary and informational texts by using key textual evidence to effectively summarize and/or analyze a text. The student thoroughly analyzes how an element of literature or informational text develops and influences the text. The student determines a central idea in an informational text. The student determines how the author uses organization, structure, form, text features, figurative language, and/or word choice to achieve a purpose. The student determines the effectiveness of point of view in a text. The student thoroughly compares and contrasts texts in different forms or genres. The student traces the development of an argument and thoroughly evaluates the author's claims and evidence in a text. The student uses context and word structure to determine the meanings of words, interprets figurative language, and understands nuances in word meanings. In writing, the student uses logical reasoning and substantive evidence to develop a cohesive argumentative/informational essay on a topic in a formal style for an intended audience. The student thoroughly organizes a narrative with a controlling point, using precise words, phrases, and narrative techniques. The student writes a text-





dependent analysis essay that responds to a text or texts and demonstrates an organized and thorough analysis that cites substantial and relevant evidence to support its intended purpose. The student uses a variety of appropriate transitional words, phrases, and clauses. The student recognizes and demonstrates a thorough command of the conventions of Standard English grammar, usage, and mechanics to convey ideas precisely and for effect.

Since the PLDs of UIS proficiency scales will be the basis for linking with NAs and CNAs, it is essential that they are fully elaborated and include details related to each content standard identified in Step 1.

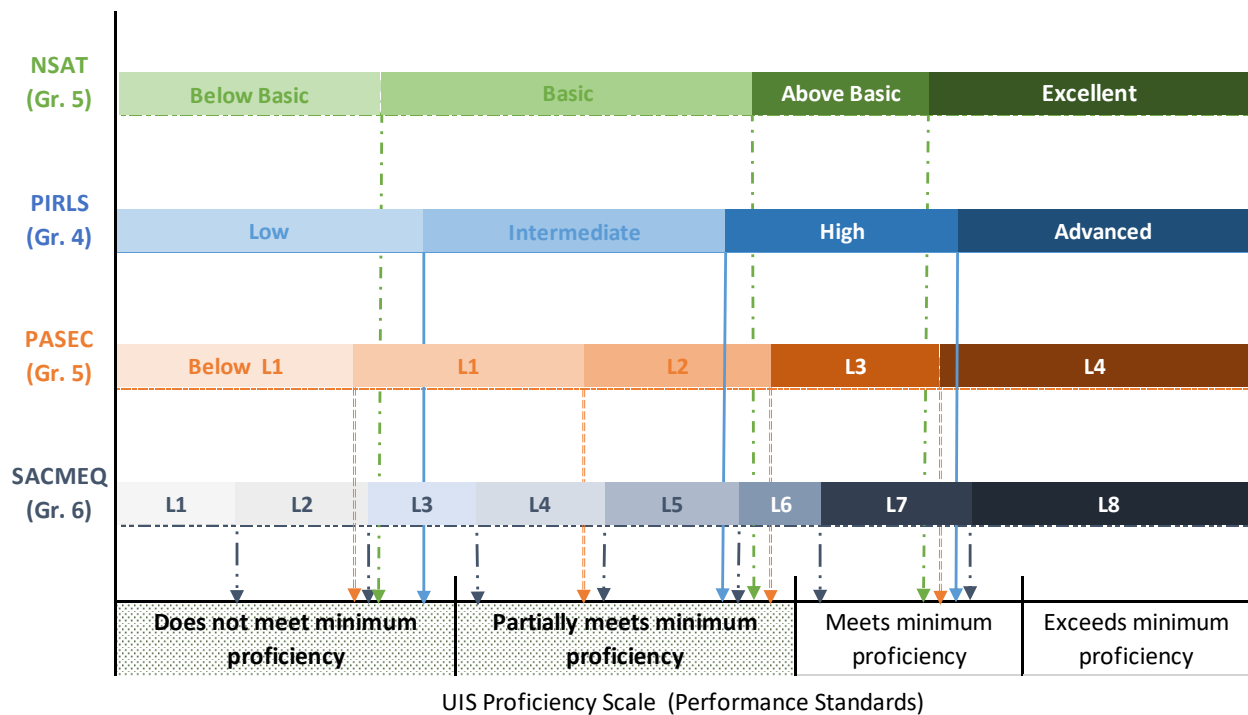
## **Linking the UIS Proficiency Scales with NAs and CNAs**

After performance levels of UIS proficiency scales for each grade and domain are defined, the next step is to link the scales with various NAs and CNAs for the purpose of SDG 4.1.1 reporting. As the UIS does not intend to develop or administer a common assessment for SDG 4.1.1 reporting, it is not possible to statistically link the UIS proficiency scales with NAs or CNAs using test- or item-based linking methods (i.e., equating, calibration, projection, or statistical moderation). However, they can be linked through content-based performance level expectations. The process of linking through the performance level expectations is called *social moderation* or *policy linking* (Buckendahl & Foley, 2015; Reckase, 2000).

In order to explain the linking process and its relevant conditions effectively, let us assume that the UIS proficiency scale for Grade 6 (end of lower secondary) reading has four performance levels (e.g., Does not meet minimum proficiency; Partially meets minimum proficiency; Meets minimum proficiency; and Exceeds minimum proficiency), as depicted in Figure 1. The levels (without the lowest level) have been defined for both policymakers (i.e., policy descriptions) and other stakeholders (i.e., full descriptions). The performance levels and descriptions for Grade 6 reading have been defined based on the Steps 1-3 discussed in the previous section.

It is further assumed that the UIS has decided to count only the percentage of students, who have mastered the required knowledge, skills, and abilities defined in Partially Meets Minimum Proficiency and below of the UIS proficiency scales for SDG 4.1.1 reporting--in other words, the percentage of students classified into Meets Minimum Proficiency and Exceeds Minimum Proficiency levels of the UIS proficiency scale would be reported for Indicator 4.1.1.

**Figure 1. Linking UIS Proficiency Scale with National and Cross-National Assessments: An Example**



*Social moderation* or *policy linking* involves two steps.

### *Step 1: Evaluate Alignment of Performance Level Descriptors (PLDs)*

In this step, one examines whether there is any alignment between Partially Meets level description of the UIS proficiency scale and performance level descriptors of the NA and the CNAs. It is proposed that a three-point holistic scale (*no or limited match, mostly matched, and fully matched*) is utilized to determine the degree of alignment. The rating rubrics include criteria related to strand- (i.e., specific content and competencies articulated in the PLD) and cognitive complexity-level (i.e., low, medium, or high) match (Webb, 1997). For example,

- **No or limited match:** No or little match at strand level.
- **Mostly matched:** Matched fully at strand level, but somewhat matched at cognitive complexity level.
- **Fully matched:** Matched fully at strand and cognitive levels.

A group of subject matter experts (SMEs) convene and make consensus ratings on the match between performance levels of the UIS proficiency scale, and the levels used in the NA and CNA. The performance levels of the NA and CNA that have holistic ratings of “mostly” or “fully matched” with the Partially Meets Minimal Proficiency level of the UIS proficiency scale will be used for reporting Indicator 4.1.1. For example, we see in Figure 1 that Level 6 of SACMEQ, Level 2 of PASEC, and Basic level of NSAT (Namibia’s Census-based National Assessment) will most likely be rated as “mostly matched” with the Partially Meets Minimal Proficiency level of the UIS reporting scale. Therefore, the countries that participate in SACMEQ would report the percentage of students at Level 7 and Level 8, and countries that participate in PASEC would report Level 3 and Level 4 for Indicator 4.1.1. Namibia would report the percentage Above Basic and Excellent.



The performance levels of the NAs and CNAs that have holistic ratings of “no or limited match” with the Partially Meets Minimal Proficiency performance level of the UIS reporting scale (e.g., Intermediate level of PIRLS in Figure 1) should go through a social moderation or policy linking procedure described in step 2, below.

### *Step 2: Set Socially Moderated Performance Standards for NAs and CNAs*

In this step, socially moderated performance standards for the NAs and CNAs will be set using a standard setting method in order to link these NAs and CNAs with the UIS proficiency scale. A yes-no variation of Angoff (Plake, Ferdous, & Buckendahl, 2005), Bookmark (Lewis, Green, Mitzel, & Patz, 1999), or Body of Work (Kahl, Crockett, DePascale, & Rindfleisch, 1995) method could be used for estimating three cut scores for Does Not Meet Minimum Proficiency/Partially Meets Minimum Proficiency, Partially Meets Minimum Proficiency/Meets Minimum Proficiency, and Meets Minimum Proficiency/Exceeds Minimum Proficiency on the NAs and CNAs for which there was not a “fully” or “mostly matched” rating (in our example, this would be PIRLS). However, the selection of the standard setting method will be based on item and test formats of the NAs or CNA. If the test contains only multiple-choice items then the yes-no variation of Angoff method would be used. If the test contains both multiple and open-ended items then the Bookmark or the Body of Work method would be used.

For each NA and CNA, a group of 8-10 SMEs per domain convenes for a socially moderated standard setting workshop. During the workshop, the SMEs are provided a thorough orientation to the standard setting method and the UIS-PS PLDs. Then, SMEs are asked to provide individual and independent judgements about each item on the test to set their initial (also called Round 1) cut scores based on their understanding of the PLDs and experience with the student populations. For example, when using the Angoff Yes/No method, the experts would say whether they thought a student meeting minimum proficiency would answer the item correctly. Other methods may require a different type of judgement to be made.

After collecting each SME’s independent and initial judgements, the judgments will be aggregated to estimate the panel-recommended cut scores. The panel is then provided feedback on their judgments and information about the implications of the proposed cut scores (e.g., item difficulty, student classification). The feedback is provided so the SMEs have an opportunity to reconsider their initial judgements and to identify errors or any misconceptions about the process of setting the cut scores and the use of the cut score to classify student scores by category.

The cut scores estimated for Partially Meets Minimum Proficiency/Meets Minimum Proficiency and Meets Minimum Proficiency/Exceeds Minimum Proficiency and percentages of students classified into Meets Minimum Proficiency and Exceeds Minimum Proficiency levels (due to the Partially Meets Minimum Proficiency/Meets Minimum Proficiency and Meets Minimum Proficiency/Exceeds Minimum Proficiency cut scores on the NAs and the CNAs) would be used for Indicator 4.1.1 reporting. In other words, the students classified into the Meets Minimum Proficiency and Exceeds Minimum Proficiency levels of the UIS proficiency scale would demonstrate required knowledge and skills (as defined in the Partially Meets Minimum Proficiency level of the UIS proficiency scale) assessed on the NAs and the CNAs.



## More Information

Below is additional information about the activities and workshops that are recommended to carry out the four steps associated with the construction of the UIS proficiency scales and two steps associated with linking of the UIS proficiency scales and the NAs and CNAs. Estimates do not include additional staff days that would be required to prepare for and facilitate the workshops or UIS internal resources for oversight and gathering input from stakeholders.

### *Construction of the UIS-PSs*

**Step 1:** Define common content standards

- This task has already been undertaken by the UNESCO IBE. It will be necessary to review the work to date and determine if additional information on synthesis is required.

**Step 2:** Define number of performance levels, determine labels, and write policy descriptions for the levels of the UIS-PS

- These tasks could be carried out by the GAML, perhaps during the next GAML meeting in November. It would require 1 full day with the GAML.

**Step 3:** Develop full descriptions for the performance levels of the UIS-PS

- One 3-day workshop per domain to define the detailed performance level descriptors. For each domain, a total of 15 subject matter experts (5 lower primary, 5 end of primary, and 5 lower secondary) from different countries attend the workshop.

### *Linking UIS-PSs with NAs and CNAs*

Step 1 is mandatory for examining the alignment between the performance level descriptors (PLDs) of the UIS proficiency scales and NAs and CNAs. However, step 2 will be planned based on the findings of step 1.

**Step 1:** Evaluation of PLDs

- A one-day workshop for each NA or CNA to examine the alignment (and conduct ratings) between PLDs of the UIS-PS. Five experts per domain/grade would participate.

**Step 2:** Set socially moderated standards

- One two-day standard setting workshop for each NA and CNA. A total of 8-10 subject matter experts for each domain/grade would attend.

### *Tentative Timeline*

The following schedule would yield UIS proficiency scales by the end of February 2018 and linked CNAs (and potentially NAs) by July 2018. The 10-month timeline is based on the following assumptions: IBE's analysis of assessments and content standards is sufficient; vetting of policy descriptors and full descriptions can be accomplished between the development of the PLDs in November and workshops in January-February; and a maximum of 1-2 CNAs would require social moderation to be linked to UIS-PMs.

	2017			2018						
	O	N	D	J	F	M	A	M	J	J
<b>Construction of UIS Performance Scales (PSs)</b>										
Determine Common Content Standards (review IBE work)										
UIS-PS Performance Levels: Write Labels and PLDs										
Write Full Descriptions of UIS-PS Levels (workshops)										
<b>Linking UIS-PS with CNAs (and NAs)</b>										
Rate Alignment of PLDs (workshops)										
Set Socially Moderated Standards (workshops)										



## References

---

Beck, M. (2003). *Standard setting: If it is science, it's sociology and linguistics, not psychometrics*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Buckendahl, C. W., & Foley, B. P. (2015). *Policy linking as cut score moderation: Considerations for practice*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.

Kahl, S. R., Crockett, T. J., DePascale, C. A., & Rindfleisch, S. L. (1995). *Setting standards for performance levels using the student-based constructed response method*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Lewis, D. M., Green, D. R., Mitzel, H. C., & Patz, R. J. (1999). *The bookmark standard setting procedure*. Monterey, CA: McGraw-Hill.

Perie, M. (2008). A guide to understanding and developing performance-level descriptors. *Educational Measurement: Issues and Practices*, 27(4), 15-29.

Plake, B. S., Ferdous, A. A., & Buckendahl, C. W. (2007). *Setting Multiple Performance Standards Using the Yes/No Method: An Alternative Item Mapping Method*. Paper presented to the meeting of the National Council on Measurement in Education, Montreal, Canada.

Reckase, M. D. (2000). *The evaluation of the NEAP achievement levels setting process: A summary of the research and development efforts conducted by ACT*. Iowa City, IA: ACT, Inc.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6). Madison: University of Wisconsin, Wisconsin Center for Education Research.

