



USAID
FROM THE AMERICAN PEOPLE

MEASUREMENT WORKSHOP

LINKING ASSESSMENTS TO A GLOBAL STANDARD
WITH SOCIAL MODERATION – AUGUST 29-31, 2018

SUMMARY REPORT

SEPTEMBER 14, 2018

Revised October 8, 2018

Revised October 17, 2018

This publication was produced for review by the United States Agency for International Development. It was prepared for the Reading and Access Evaluation Project by Management Systems International, A Tetra Tech Company.

MEASUREMENT WORKSHOP

LINKING ASSESSMENTS TO A GLOBAL STANDARD WITH SOCIAL MODERATION – AUGUST 29-31, 2018

SUMMARY REPORT

Contracted under

AID-OAA-M-13-00010
Reading and Access Evaluation

DISCLAIMER

The author's views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

CONTENTS

CONTENTS	3
ACRONYMS	4
ACKNOWLEDGMENTS	5
WORKSHOP OVERVIEW	6
UNDERSTANDING POLICY LINKING	7
FEEDBACK ON POLICY LINKING	8
KEY WORKSHOP TAKEAWAYS	8
MAIN RECOMMENDATIONS	11
ANNEX 1: WORKSHOP AGENDA	14
ANNEX 2: WORKSHOP PARTICIPANTS	17
ANNEX 3: FURTHER READING	20
ANNEX 4: ILLUSTRATIVE POLICY DEFINITIONS AND PERFORMANCE LEVEL DESCRIPTORS	21
POLICY DEFINITIONS	21
PERFORMANCE LEVEL DESCRIPTORS	22

ACRONYMS

COR	Contracting Officer's Representative
E3/ED	Office of Education in the Bureau for Economic Growth, Education and Environment
GRN	Global Reading Network
IBE	International Bureau of Education
MSI	Management Systems International
ORF	Oral Reading Fluency
PLD	Performance Level Descriptor
READ	Reinforcing Education Accountability in Development
SDG	Sustainable Development Goal
SME	Subject Matter Expert
UIS	UNESCO Institute for Statistics
UNESCO	United Nations Educational, Scientific and Cultural Organization
URC	University Research Co., LLC
USAID	United States Agency for International Development
USG	United States Government

ACKNOWLEDGMENTS

This report summarizes the lessons learned during a three-day measurement workshop that was the product of a collaborative effort. The Office of Education in the Bureau for Economic Growth, Education and Environment (E3/ED) of the U.S. Agency for International Development (USAID) has been very supportive of this activity. The team would like to thank Benjamin Sylla for his leadership as the Contracting Officer's Representative (COR) of the Reading and Access Evaluation Project. We also extend a special thanks to Rebecca Rhodes and Melissa Chiappetta of USAID for their support and guidance throughout the entire process.

From the Global Reading Network (GRN), we would like to thank Jennifer Gerst and Deepa Srikantaiah for hosting the workshop at the University Research Co., LLC (URC) in Bethesda, MD.

The following organizations generously contributed participants to the workshop: Abt Associates, American Institutes for Research, Anderson Education Associates, ASER Centre, Bill & Melinda Gates Foundation, CARE, Center for Global Development, Chemonics, Council for School Examinations and Assessment - Bhutan, Creative Associates, DAI Global, Department for International Development – U.K., Education Development Center, EnCompass, Evans and Associates Consulting, FHI360, Global Partnership for Education, Global Reading Network, International Association for the Evaluation of Educational Achievement, International Business & Technical Consultants, Inc., International Rescue Committee, Juarez and Associates, Khulisa, Making Cents International, Management Systems International, Mathematica Policy Research, Ministry of Education – Nepal, National Council of Educational Research and Training – Pakistan, People's Action for Learning Network, Program for Analyzing Education Systems of the Confemen, Room to Read, RTI International, Save the Children International, School-to-School International, Social Impact, U.K. Department for Education, United Nations Educational, Scientific and Cultural Organization, UNESCO/Dakar, UNESCO Institute for Statistics, United Nations Children's Fund, University of Maryland, U.S. Agency for International Development, Uwezo, World Bank, and World Education.

This workshop would not have succeeded if not for the effort of presenters, panelists, moderators and facilitators: Abdullah Ferdous, Christine Beggs, Dana Kelly, Deepa Srikantaiah, Hong Jiao, Jeff Davis, Jennifer Gerst, Luis Crouch, Marguerite Clarke, Pooja Reddy Nakamura, Ramya Vivekanandan, Silvia Montoya, and Thomaz Alvares de Azevedo.

Dana Kelly of Management Systems International (MSI) oversaw this activity with support from Idalia Rodriguez Morales, Jonathan Weinstock, and Sean Kelly.

WORKSHOP OVERVIEW

The Global Reading Network (GRN) Measurement Workshop on *Linking Assessments to a Global Standard with Social Moderation* was held at the University Research Co., LLC (URC) from Wednesday, August 29 to Friday, August 31, 2018. Workshop participants represented a diverse group of international education stakeholders from ministries of education, implementing partners, evaluation partners, foundations, regional and international assessment programs, universities, research centers and donor agencies. Some attendees participated remotely. Participants engaged in a variety of presentations, discussions, panels, and working groups over the three days.

The workshop was organized in response to the 2017 Reinforcing Education Accountability in Development (READ) Act, which intersects with the reporting requirements of the 2030 Sustainable Development Goals (SDGs). The READ Act requires the U.S. Government (USG) to:

- 1) Apply rigorous monitoring and evaluation methodologies to determine if programs and activities...accomplish measurable improvements in literacy, numeracy, and other basic skills development that prepare an individual to be an active, productive member of society and the workforce; and
- 2) Include methodological guidance in the implementation plan and support systemic data collection using internationally comparable indicators, norms, and methodologies, to the extent practicable and appropriate.

The workshop was responsive to the second requirement, i.e., to “support the development of internationally comparable indicators, norms, and methodologies” for measurement and reporting purposes. Specifically, we explored social moderation – also called *policy linking* (and this is the terminology that we will use moving forward) – as a method to link different learning assessments by subject area and grade level to global standards of minimum proficiency. Implementing policy linking would allow the USG to benefit from two types of analyses that are not currently possible for results in basic education from different development projects and/or countries on a global basis: 1) comparisons of results across assessments and 2) aggregation of results from different assessments.

The workshop had four objectives for the policy linking method: 1) develop a common understanding of it, 2) get stakeholder feedback on its applicability, 3) raise issues around implementing it, and 4) gain technical input into its operationalization. Issues involving policy linking, including its application for determining the percentage of students achieving the same standards of minimum proficiency across projects and/or countries, were presented and discussed during the workshop.

The sections below summarize the workshop. The first section covers the development of a common understanding of policy linking. The second section gives a brief description of the feedback sessions on policy linking in which stakeholder information was provided, implementation issues were raised, and technical input into operationalization was solicited. The following two sections provide key takeaways and recommendations from the workshop. The final part of this report includes annexes with the workshop agenda ([Annex 1](#)), list of workshop participants ([Annex 2](#)), and a short bibliography for further reading ([Annex 3](#)).

UNDERSTANDING POLICY LINKING

On the first day of the workshop, there were four presentations designed to build a common understanding of policy linking: 1) reporting to a global standard, 2) forms of linking assessments, 3) overview of policy linking, and 4) linking assessments to a global standard with policy linking. The first day concluded with a panel discussion on global reporting using policy linking.

The presentations provided information on two types of assessment linking: *statistical linking* and *policy (non-statistical) linking*. Statistical linking, which includes test equating, is the most widely-used form of linking. It requires building a link between assessments that measure a similar construct using a sample of common persons, i.e., the same students taking different assessments, or a subset of common items, i.e., the same items on different assessments. Once one of these requirements (along with other assumptions about the construct, populations, and testing conditions) is met and the assessment data are collected, a psychometrician analyzes the data set (or data sets) to create a conversion table that shows the equivalence of scores on the different assessments. For instance, a score of 40 points on one (more difficult) assessment might equal a score of 44 points on another (easier) assessment. The conversion table usually includes the relationship between the score points on the different assessments to a common scale, which then facilitates placing additional assessments on the common scale.

Policy (non-statistical) linking is based on the concept that the linkage from one assessment to another is a set of external criteria, as in performance level descriptors (PLDs) that reflect students' knowledge and skills. Policy linking also requires that linked assessments measure a similar construct, but it does not require common persons or common items to build the link. For different assessments, subject matter experts (SMEs) use a common set of PLDs and a standard setting method to make judgments on the difficulty of the assessments. An analyst then produces a concordance table that matches the scores on different assessments to each other. Sometimes, only the cut scores, e.g., between passing and failing or proficient and not proficient, on different assessments are matched to each other. For instance, a cut score of 50 points on one (more difficult) assessment might equal a cut score of 55 points on another (easier) assessment. Policy linking is less precise than statistical linking, but it is a more practical method in situations where the requirements of statistical linking are not met and/or resources are limited. It is increasingly useful in supporting the "growing demand among policy makers for multiple assessment pathways that serve the same, or effectively similar, purposes" (Buckendahl & Foley, 2015).

The explanation of forms of linking assessments was followed by presentations on the steps in implementing policy linking. First, policy linking requires general (or policy) definition(s) of a single or multiple performance standard(s). For a single performance standard of minimum proficiency, a *policy definition* would be developed across subject areas and grade levels. This policy definition, which is developed by policymakers, is generally short, i.e., a few sentences for a single performance standard such as minimum proficiency, and broad so that it can be widely applied. [See Annex 4 for illustrative policy definitions.](#)

Second, it uses the policy definition as the basis for PLDs for each targeted subject area and grade level. The PLDs are developed by SMEs. Different teams would need to develop, for example, four separate PLDs for minimum proficiency in two subject areas (e.g., reading and math) at two grade levels (e.g., end of lower primary and end of upper primary school). [See Annex 4 for illustrative PLDs.](#)

Third, it involves applying the PLDs through a *standard setting* procedure for each subject area and grade level to set cut scores. A cut score is needed for each performance standard, e.g., minimum proficiency by subject area and grade level. Internationally-accepted procedures include Angoff, modified Angoff, Bookmark, and Body-of-Work. The choice of the standard setting procedure depends primarily on the items types in the assessment. For policy linking to a standard of minimum proficiency, the PLDs would be applied using a procedure to set cut scores on any or all assessments that measure a similar construct for a subject area and a grade level. Each assessment will have a different cut score (unless assessments coincidentally have the same cut scores) that aligns with the PLDs.

Fourth, it entails the *validation* of the links. After applying the policy linking method for different assessments and setting the cut scores by subject area and grade level, both quantitative and qualitative evidence should be gathered to examine the strength of the policy links. Quantitative data involve statistics that are calculated on the consistency of the judgments for setting the cut scores, e.g., inter- and intra-rater reliability and location data. Qualitative data involves descriptions of student performance around the cut scores, i.e., information on the consistency of performance in relation to the PLDs.

During the discussions, the possibility of using both statistical and policy linking methods to form a portfolio of options for linking assessments was raised. This was based on the practical value of being able to link some assessments using statistical methods and other assessments using non-statistical methods, with all assessments linked to a global standard for comparison and aggregation purposes.

FEEDBACK ON POLICY LINKING

On the second day of the workshop, the discussion shifted to the second and third objective – the applicability of and issues with implementing policy linking to link assessments to a global standard of minimum proficiency. It began with a case study about setting performance standards, such as minimum proficiency, in reading and math. This was followed by a reflection on the politics, process and logistics of developing a global performance standard. Then there was presentation on developing PLDs, which guided a session in which the participants were divided into different working groups, with each group proposing a policy definition for minimum proficiency.

The final day of the workshop was devoted to the fourth objective of eliciting technical input so that USAID could better evaluate the possible use of policy linking for reporting to a global performance standard of minimum proficiency. Working groups discussed issues around governance, communications, quality assurance, and validation. A final session of the workshop featured comments and recommendations from representatives of the World Bank and UNESCO Institute for Statistics (UIS). Final comments were made by other participants, facilitators, GRN hosts, and USAID representatives.

Key takeaways from the three days of the workshop are presented below, along with main recommendations based on those takeaways.

KEY WORKSHOP TAKEAWAYS

The key takeaways that emerged from the workshop were the following:

1. **Multiple stakeholders are highly interested in linking assessments to global performance standards.** The workshop had excellent participation from a wide variety of key stakeholders. Those from ministries of education, regional and cross-national assessment programs, implementing partners, evaluation partners, foundations, universities, research centers and donor agencies were highly engaged during the sessions, made important contributions, and expressed support for moving forward on linking of assessments to a global description of minimum proficiency.
2. **Policy linking is a viable method for linking different assessments to a global standard of minimum proficiency.** Statistical linking is a stronger technical method but policy linking – as a non-statistical technical method – is an alternative that is recognized by psychometricians and has a record of being used in situations where statistical linking is not practical or technically possible. This method is particularly useful when common item and common person statistical linking designs are either not possible or when they would be too expensive and/or time-consuming.
3. **Policy linking can be conducted ex post, with assessments that have already been administered, since it does not require common persons or items.** Policy linking is more practical in situations where statistical linking has not been planned in advance. In some cases, statistical linking using common persons can be done after the operational administration of the assessments, but it would require additional sampling and data collection, while policy linking does not require additional student-level data to establish the linkages.
4. **A portfolio approach for assessment linking, using a combination of statistical linking and policy linking, is technically feasible.** A portfolio, or combination, of statistical linking and policy linking is technically possible and could be practical. For instance, statistical linking could be used to link different assessments when certain assumptions are met. An example would be if regional assessments (e.g., LLECE, PASEC, SACMEQ, SEA-PLM) and international assessments (e.g., PIRLS, TIMSS) measured similar constructs with similar populations and employed a common person or common item linking design. Such an approach would allow for comparability between those assessments at each score point. Policy linking could be used to link different assessments when the requirements of statistical linking are not met, though it is generally only conducted only for cut scores separating performance categories, i.e., where there are common PLDs. Whether statistical or policy linking is used to link assessments to each other, policy linking would be needed to link any of these assessments to a *global standard* for comparing and reporting, e.g., in relation to a global performance standard of minimum proficiency.
5. **Either the development or selection of a policy definition, with a concentration on minimum proficiency, is possible.** Group exercises during the workshop demonstrated that a diverse set of policy experts could reference existing definitions of minimum proficiency to adapt a global policy definition. Another option would be to adopt one of the existing definitions as the global definition. The provision for either adapting or adopting a policy definition is that it needs to be appropriate for all assessments that will be linked.
6. **Resources for developing detailed PLDs based on global content standards are available.** Following up on the approach used to arrive at a policy definition, participants agreed that a diverse group of SMEs could use existing global content standards and descriptors to adapt or adopt detailed PLDs. The SMEs would include language experts, teachers, inspectors, curriculum developers, teacher trainers, and assessment experts from various countries by subject area –

reading and math – and grade level – lower primary and end of upper primary school. A viable option would be to use, as a starting point, the reading and math topics that have been recently extracted by the International Bureau of Education (IBE-UNESCO) from 140 assessments across the globe. The PLDs would need to be broad enough for a range of assessment content and formats as well as the diverse rates of reading and math acquisition in the global context.

7. **Proven methodologies for setting performance-based standards are available and applicable.** Participants agreed that existing, research-based standard setting methods meet the requirements for policy linking, i.e., the methods accommodate the variety of assessment formats. These include Angoff, modified Angoff, Bookmark, and Body-of-Work. Depending on the item types used in an assessment, any of these methods could be used in the standard setting process needed to establish the cut scores for minimum proficiency in the different subject areas and grade levels. These methods could be used to set cut scores for additional performance categories below and above minimum proficiency, such as beginning or advanced, if desired.
8. **Early grade math subtasks for use in policy linking are not as well developed as those for reading.** During the workshop, there were scant discussions of the math content that could be used in policy linking for minimum proficiency. Essential components of math knowledge and skills – numbers, operations, measurement, geometry, problem-solving – are broader than their counterparts of oral reading fluency (ORF) and comprehension in reading. Curriculum-based tests, particularly in upper primary, can cover these components but further study is needed for the identification of appropriate math subtasks on early grade assessments such as EGMA, ASER, and Uwezo in examining minimum proficiency.
9. **Each country would determine the assessments to use as a basis for setting minimum proficiency.** Countries could choose from sub-national, national, or cross-national assessments. Several examples were mentioned throughout the workshop by the participants. These included cross-national assessments such as LLECE, PASEC, PIRLS, SACMEQ, SEA-PLM, and TIMSS, as well as sub-national and national assessments in Egypt, Ghana, Guinea, Honduras, Lebanon, Namibia, Pakistan, and Zambia. The only requirements are that the assessments should meet standards of quality (e.g., see the AERA/APA/NCME Standards for Educational and Psychological Testing) and be documented sufficiently for others to judge their technical characteristics. Participants also noted that many countries do not have ongoing assessments that provides sub-nationally or nationally representative data on reading and math performance at lower primary and the end of upper primary. Parallel efforts to further build the capacity of countries to develop and implement national assessments are necessary (and already underway). Again, the policy linking approach allows us to move forward with any reasonably acceptable assessments that exist in each country.
10. **A policy paper that provides a technical justification for policy linking could be prepared.** As a complement to the brief policy linking steps presented in the workshop, a relatively brief description of a step-by-step procedure for implementing policy linking in national and international contexts does not exist. Participants mentioned that an accessible, non-technical explanation of policy linking and an overview of how policy linking would be implemented would be useful for a broad audience, including high-level policymakers.
11. **Opportunities exist for operationalizing policy linking, including field piloting and scaling up.** Participants discussed the need for piloting the operational aspects of policy linking. This would include two parts: 1) developing draft global policy definitions and PLDs, along with

identification of appropriate standard setting methods and 2) piloting the policy linking method with a national assessment and a cross-national assessment in the field to test its feasibility and refine the methodology.

12. **A toolkit that includes sections on each of the activities needed to implement policy linking could be developed.** A draft toolkit would need to be developed prior to field piloting. In addition to providing guidance for the policy definition, PLDs, and standard setting, the toolkit would be piloted in two or three countries and then revised so that it becomes the means for ensuring the quality of the implementation of policy linking in different contexts.
13. **Guidance for validation would lead to information on the technical properties of the cut scores.** Guidance for validation, as a final step in the policy linking process and as a part of the toolkit, would be provided in two areas: 1) information on the ability of workshop participants to follow procedural details that document the policy linking method, including a list of workshop participants, their level of understanding of technical information, their ability to follow the process correctly, and their confidence in the process and results; 2) technical details that document and validate the outcomes, such as calculating standard errors, location statistics (i.e., estimated cut scores of the individual SMEs), magnitude of change among different rounds of ratings, and inter-rater and intra-rater reliability measures. Validation would provide internationally-acceptable assurance that the policy linking method is meeting the need for a valid and reliable method for setting comparable minimum proficiency cut scores on sub-national, national, and cross-national assessments.

MAIN RECOMMENDATIONS

The main recommendations for next steps that emerged from the workshop were the following:

1. **Piloting the steps in policy linking, based on guidance provided in a toolkit would be a next activity for operationalizing the approach.** To build on the interest generated during the workshop, moving forward with activities to pilot the policy linking approach – with existing assessments – is recommended. This would involve creating a toolkit, along with a separate technical paper, if necessary, to provide detailed justification and guidance for the next steps in applying policy linking to the operationalization of a global standard of minimum proficiency. The two subject areas – reading and math – and the two grade levels – lower primary and end of upper primary – covered in the workshop should be targeted. Discussions on the toolkit and technical paper should take place between key stakeholders such as USAID, other donor agencies, and implementing partners.
2. **Taking a decision on whether to adopt or adapt a definition of minimum proficiency will need further consideration.** As the first step in operationalizing policy linking, an existing policy definition of minimum proficiency could be adopted or adapted from an existing definition. This process will need further consideration, since some participants thought that it would be easier to use an existing definition while other participants thought that it would be more broadly acceptable to either adapt or create a new definition.
3. **Developing detailed PLDs based on global content standards can proceed based on existing content standards.** A diverse group of SMEs, including language experts, teachers, inspectors, curriculum developers, teacher trainers, and assessment experts for reading and math in lower and upper primary school could create detailed PLDs from global content standards, such as

those extracted from assessments by IBE-UNESCO. If not adequately covered in the IBE-UNESCO study, content for assessments such as EGRA, EGMA, ASER, and Uwezo is readily available and can be studied in more depth for the early grades PLDs. The resulting global PLDs in reading and math at the two grade levels need to be broad enough to cover a range of assessment content and formats as well as diverse rates of reading and math acquisition.

4. **Proven methodologies for setting performance-based standards should be used for policy linking.** Assessment specialists and psychometricians could recommend internationally-accepted and validated standard setting methods that would be applicable to different assessment formats. Participants agreed that existing, research-based standard setting methods can meet the requirements of different assessment formats. At least on a pilot basis, trained psychometricians and SMEs would apply the appropriate standard setting method to set minimum proficiency cut scores on an initial set of assessments (including testlets, or the relevant parts of assessments, such as a subtasks), based on the PLDs.
5. **Validation exercises should be carried out on the pilot policy linking activities to gather evidence on validity and reliability of the cut scores.** As a final step in the policy linking process, validation evidence should be gathered to provide information on the process and outcomes of pilot activities. This information should be examined to establish confidence in policy linking as a viable method for comparing and aggregating results from sub-national, national, and cross-national assessments. Validation evidence can be used to improve the process after piloting as well as checking on the process and results when/if the method is scaled up to a broad range of assessments.
6. **A portfolio of linking methods should be examined for inclusion in assessment linking.** While very few national and cross-national assessments are currently linked, the possibility of combining statistical linking with policy linking should be retained. The requirements of statistical linking, i.e., common persons or common items, should be fully recognized as a main reason for the current lack of linked assessments. Policy linking provides a less rigorous, while still acceptable, option for linking either existing or new assessments to a global standard of minimum proficiency. The two methods could be combined when the requirements for statistical linking are fulfilled.
7. **Reading comprehension is feasible for use in policy linking, though improvements to validity and reliability of measurements of comprehension should be made.** While it is important and practical to proceed with existing assessments for policy linking, the global education measurement community recognizes that improvements are needed in reading comprehension subtasks. This is especially the case since reading experts in the workshop recommended comprehension as the most appropriate reading domain for establishing minimum proficiency. Such improvements will allow for more rigorous and accurate measures of reading comprehension moving forward. Participants also mentioned that ORF remains a key skill area in early grades, though they raised issues involving the influence of the language of assessment on definitions of minimum proficiency, measurement of ORF, and the resulting comparisons and aggregation of the test results.
8. **Additional studies of math content for linking assessments to a global performance standard should be conducted.** Given that math is a key subject area for inclusion in international basic education interventions, assessments, and reporting, math content should benefit

from further studies. This is essential in identifying the math content needed for measuring and reporting on minimum proficiency in relation to a global performance standard.

- 9. Further consideration should be given to building stakeholder support for reporting to a global minimum proficiency standard.** The three-day workshop, along with the preparation and information sharing, was vital to initial efforts in reporting to a global standard. However, along with discussions on the toolkit and piloting, stakeholders should engage in additional discussions on the longer-term benefits of policy linking for comparing and reporting on global indicators of basic education using student assessment results. This is seen as a critical part of building stakeholder support and creating a viable and sustainable approach, both technically and practically, for improving student learning through assessment, dissemination, and advocacy.

ANNEX I: WORKSHOP AGENDA

Day I – Wednesday, August 29, 2018

Time	Topic
8:30-9:30	Registration
9:30-9:45	Opening remarks
9:45-10:15	Introductions
10:15-10:30	Reporting to a global standard <ul style="list-style-type: none">- Issues and challenges
10:30-10:45	Break
10:45-12:30	Forms of linking assessments: <ul style="list-style-type: none">- Statistical methods (equating, calibration, projection and statistical moderation)- Judgmental method (social moderation/policy linking)
12:30-1:45	Lunch
1:45-2:45	Social moderation/policy linking <ul style="list-style-type: none">- Overview of method- Illustrative examples
2:45-3:30	Linking assessments to a global standard using social moderation/policy linking <ul style="list-style-type: none">- Global linking- Reporting to a global standard
3:30-3:45	Break
3:45-5:00	Panel discussion on linking to a global standard with social moderation
5:00	Adjourn

Day 2 – Thursday, August 30, 2018

Time	Topic
9:00-9:15	Recap Day 1
9:15-10:45	Setting performance standards <ul style="list-style-type: none">- Best practices and steps- Example from Lebanon
10:45-11:00	Break
11:00-11:30	Setting performance standards (continued)
11:30-12:30	Performance-level descriptors (PLDs) <ul style="list-style-type: none">- Policy definitions and detailed descriptions- Characteristics of good PLDs- Vertical alignment
12:30-1:30	Lunch
1:30-3:15	Group consultation: Policy definitions of minimum proficiency <ul style="list-style-type: none">- Examples of policy definitions- Develop global policy definitions of minimum proficiency
3:15-3:30	Break
3:30-5:00	Group consultation: Operationalizing social moderation <ul style="list-style-type: none">- Applying PLDs- Setting cut scores- Selecting panelists- Defining roles of panelists- Using impact data
5:00	Adjourn

Day 3 – Friday, August 31, 2018

Time	Topic
9:00-9:15	Recap Day 2
9:15-10:00	Group consultation: Operationalizing social moderation (continued) - PLDs, cut scores, panelists and impact data
10:00-11:00	Group discussion and feedback
11:00-11:15	Break
11:15-12:30	Group #1 Consultation: Governance and communication Group #2 Consultation: Quality assurance and validation
12:30-1:30	Lunch
1:30-3:00	Group reporting, discussion and feedback
3:00-3:30	Closing remarks and adjourn

ANNEX 2: WORKSHOP PARTICIPANTS

Organization	First Name	Last Name
Abt Associates	Cris	Price
American Institutes for Research	Todd	Drummond
American Institutes for Research	Pooja	Nakamura
Anderson Education Associates	Kate	Anderson
ASER Centre	Shafer	Banu Vagh
ASER Centre	Ketan	Verma
Bill & Melinda Gates Foundation	Luke	Justice
CARE	Lotte	Renault
Centre for Global Development	Barbara	Bruns
Chemonics	Jordene	Hale
Chemonics	Laura	Harrington
Creative Associates	Corrie	Blankenbeckler
Creative Associates	Karen	Tietjen
DAI	Andrea	Lozano
Department for International Development (DFID)	Emily	Woolf
EDC	Brittany	Hebert
EDC	Daniel	Lavan
EnCompass	Zachariah	Falconer-Stout
EnCompass	Amy	Mulcahy-Dunn
Evans and Associates Consulting	Norma	Evans
FHI 360	Emily	Miksic
George Washington University	Colette	Chabbott
Global Partnership for Education (GPE)	Medjy	Pierre-Louis
Global Partnership for Education (GPE)	Ramya	Vivekanandan
Global Reading Network	Jennifer	Gerst
Global Reading Network	Deepa	Srikantaiah
IBTCI	Steven	Hansch

Organization	First Name	Last Name
IBTCI	Susan	Morawetz
IEA's TIMSS & PIRLS International Study Center, Boston College	Michael	Martin
International Rescue Committee	Jeongmin	Lee
Juarez and Associates, Inc	Rebecca	Westbrook
Making Cents International	Christy	Olenik
Mathematica Policy Research	Emilie	Bagby
Ministry of Education - Nepal	Lekha Nath	Poudel
MSI	Thomaz	Almeida
MSI	Jeff	Davis
MSI	Abdullah	Ferdous
MSI	Sean	Kelly
MSI	Dana	Kelly
MSI	Idalia	Rodriguez Morales
PAL	Winnie	Cherotich
PASEC/Confemen	Hilaire	Hounkpodoté
Room to Read	Christine	Beggs
RTI International	Luis	Crouch
RTI International	Margaret (Peggy)	Dubeck
Save the Children International	Amy Jo	Dowd
Save the Children International	Jonathan	Seiden
School-to-School International	Peter	Cooper
School-to-School International	Hetal	Thukral
Social Impact	Miguel	Albornoz
Social Impact	Mike	Duthie
UK Department for Education	Colin	Watson
UNESCO-Dakar	Davide	Ruscelli
UNESCO-UIS	Silvia	Montoya
UNESCO-UIS	Brenda	Tay-Lim

Organization	First Name	Last Name
UNICEF	Camilla	Woeldike
University of Maryland	Hong	Jiao
USAID	Koli	Banik
USAID	Robert	Burch
USAID	Melissa	Chiappetta
USAID	Patrick	Collins
USAID	Brooke	Estes
USAID	Eirini	Gouleta
USAID	Nate	Haight
USAID	Linda	Hiebert
USAID	Katherine	Johnson-Davis
USAID	Michael	Lisman
USAID	Rebeca	Martinez
USAID	Rebecca	Rhodes
USAID	Paola	Rodriguez
USAID	Anjuli	Shivshanker
USAID	Ben	Sylla
USAID	Elena	Walls
World Bank	Marguerite	Clarke
World Bank	Shwetlena	Sabarwal
World Education	Haiyan	Hua

ANNEX 3: FURTHER READING

Buckendahl, C.W., & Foley, B.P. *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, 2015.

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

Plake, B.S., Ferdous, A.A., & Buckendahl, C.W. *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, 2007.

ANNEX 4: ILLUSTRATIVE POLICY DEFINITIONS AND PERFORMANCE LEVEL DESCRIPTORS

POLICY DEFINITIONS

Below are illustrative policy definitions from several large-scale assessments:

Assessment	Performance Levels		
National Assessment of Educational Progress (NAEP)	<p>Basic: This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.</p>	<p>Proficient: 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.</p>	<p>Advanced: This level signifies superior performance beyond proficient.</p>
Pennsylvania Statewide Testing Program	<p>Basic: Marginal academic performance, work approaching, but not yet reaching, satisfactory performance, indicating partial understanding and limited display of the skills included in Pennsylvania’s academic standards.</p>	<p>Proficient: Satisfactory academic performance indicating a solid understanding and adequate display of the skills included in Pennsylvania’s academic standards.</p>	<p>Advanced: Superior academic performance indicating an in-depth understanding and exemplary display of the skills included in Pennsylvania’s academic standards.</p>
Arizona Statewide Testing Program	<p>Approaches the Standard: This level denotes understanding of the knowledge and application of the skills that are fundamental for proficiency in the standards.</p>	<p>Meets the Standard: 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.</p>	<p>Exceeds the Standard: This level denotes demonstration of superior academic performance evidenced by achievement substantially beyond the expected goal of all students.</p>

PERFORMANCE LEVEL DESCRIPTORS

Below are example, abbreviated, Grade 4 mathematics PLDs from the U.S. National Assessment of Educational Progress (NAEP). More detailed descriptors for each level are available.

Basic (214)

Fourth-grade students performing at the *Basic* level should show some evidence of understanding the mathematical concepts and procedures in the five NAEP content areas.

Fourth-graders performing at the *Basic* level should be able to estimate and use basic facts to perform simple computations with whole numbers, show some understanding of fractions and decimals, and solve some simple real-world problems in all NAEP content areas. Students at this level should be able to use—though not always accurately—four-function calculators, rulers, and geometric shapes. Their written responses will often be minimal and presented without supporting information.

Proficient (249)

Fourth-grade students performing at the *Proficient* level should consistently apply integrated procedural knowledge and conceptual understanding to problem solving in the five NAEP content areas.

Fourth-graders performing at the *Proficient* level should be able to use whole numbers to estimate, compute, and determine whether results are reasonable. They should have a conceptual understanding of fractions and decimals; be able to solve real-world problems in all NAEP content areas; and use four-function calculators, rulers, and geometric shapes appropriately. Students performing at the *Proficient* level should employ problem-solving strategies such as identifying and using appropriate information. Their written solutions should be organized and presented both with supporting information and explanations of how they were achieved.

Advanced (282)

Fourth-grade students performing at the *Advanced* level should apply integrated procedural knowledge and conceptual understanding to complex and nonroutine real-world problem solving in the five NAEP content areas.

Fourth-graders performing at the *Advanced* level should be able to solve complex and nonroutine real-world problems in all NAEP content areas. They should display mastery in the use of four-function calculators, rulers, and geometric shapes. The students are expected to draw logical conclusions and justify answers and solution processes by explaining why, as well as how, they were achieved. They should go beyond the obvious in their interpretations and be able to communicate their thoughts clearly and concisely.

Below are example 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.

U.S. Agency for International Development
1300 Pennsylvania Avenue, NW
Washington, D.C. 20004