2024

International Assessment Benchmarks: Inputs to Enhance the K to 12 Assessment Policies

Minie Rose C. Lapinid
De La Salle University, Manila, minie.lapinid@dlsu.edu.ph

Voltaire Mistades
De La Salle University, Manila

Richard R. Sagcal
De La Salle University, Manila

Leah E. Gustilo
De La Salle University, Manila

Marilyn U. Balagtas
Philippine Normal University, Manila

See next page for additional authors

Authors

Minie Rose C. Lapinid; Voltaire Mistades; Richard R. Sagcal; Leah E. Gustilo; Marilyn U. Balagtas; Richard DLC. Gonzales; Cesar T. Medula, Jr.; Brando C. Palomar; and Ria Lorraine H. Pagliawan

POLICY BRIEF
2024-04-028, APRIL 2024
ISSN # 2094-3342

De La Salle University
AKI
Angelo King Institute
for Economic and Business Studies

International Assessment Benchmarks: Inputs to Enhance the K to 12 Assessment Policies

Written by:

Minie Rose C. Lapinid[1]
Voltaire M. Mistades[1]
Richard R. Sagcal[1]
Leah E. Gustilo[1]
Marilyn U. Balagtas[2]
Richard DLC. Gonzales[3]
Cesar T. Medula, Jr.[2]
Brando C. Palomar[2]
Ria Lorraine H. Pagliawan[2]

[1]*De La Salle University - Manila*
[2]*Philippine Normal University - Manila*
[3]*Cagayan State University*

International Assessment Benchmarks: Inputs to Enhance the K to 12 Assessment Policies

Minie Rose C. Lapinid[1], Voltaire M. Mistades[1], Richard R. Sagcal[1]
Leah E. Gustilo[1], Marilyn U. Balagtas[2], Richard DLC. Gonzales[3]
Cesar T. Medula, Jr.[2], Brando C. Palomar[2], Ria Lorraine H. Pagliawan[2]

[1]De La Salle University - Manila
[2]Philippine Normal University - Manila
[3]Cagayan State University

The study conducted a document analysis of the four Philippine Department of Education (DepEd) K to 12 assessment policy guidelines and seven international large-scale assessment (ILSA) framework documents for benchmarking purposes and policy advice. Each identified ILSA standard indicator is mapped to each DepEd document policy provision to determine the gaps in the assessment policy guidelines. ILSA key standard indicators with no equivalent concept corresponding to DepEd assessment policy guidelines were considered gaps for adoption and policy recommendations. We found five ILSA key indicators not fully captured in the classroom and national assessment policies and three considered gaps for adoption into the classroom, national, and system assessment policies. The classroom assessment policies could elaborate more on the contexts by which students can apply what they know and include ILSA items for classroom use. The national assessment policies can elaborate more on (1) improving its test development process, (2) considering open-constructed response formats of test items than the usual closed-constructed response type, and (3) classifying acceptable items in the item bank for balanced and well-spread test items based on test type. The ILSA standards identified as gaps for adoption for classroom and national assessments are: (1) regularly revisiting and updating assessment frameworks based on research findings and (2) including some innovative assessments of new sets of skills. Adaptive testing and computer-based assessments may be considered in the appropriate assessment policies. Implications in policy change, reform, and future directions are thereafter suggested.

Recommended Citation:

Lapinid, M.R.C., Mistades, V.M., Sagcal, R.R., Gustilo, L.E., Balagtas, M.U., Gonzales, R.DLC., Medula, C.T., Palomar, B.C., & Pagliawan, R.L.H. (2024). International assessment benchmarks Recommendations for K to 12 assessment policy guidelines. *DLSU-AKI Policy Brief, 2024-04-028*.

Large-scale standardized national and international assessment results are being used by many countries to inform policy (Hernandez-Torrano & Courtney, 2021). Policymakers worldwide use ILSA results for cross-national comparisons to establish benchmarks for improving their educational system. Due to the low turnout of Filipino performance in international large-scale assessments (ILSAs) (OECD, 2023), this study deemed it necessary to revisit and review the four Department of Education (DepEd) K to 12 assessment policy guidelines: DepEd Order No. 8, s. 2015 - Classroom Assessment (CA); DepEd Order No. 55, s. 2016 - National Assessment (NA); DepEd Order No. 29, s. 2017 - System Assessment (SA); and DepEd Order No. 31, s. 2020 - Interim Assessment. The most recent of the following ILSA framework documents were also analyzed: International Development and Early Learning Assessment (IDELA), Early Grade Reading Assessment (EGRA), Early Grade Mathematics Assessment (EGMA), Progress in International Reading Literacy Study (PIRLS), SEA-PLM, TIMSS, and PISA. Key standard indicators from ILSA framework documents and the assessment policy provisions from the mentioned DepEd documents were identified. ILSA standard indicators that cannot be mapped to the standard indicators of the DepEd Order assessment policy guidelines are considered gaps for adoption as policy recommendations. Additionally, some DepEd policy provisions must be elaborated to best capture the ILSA indicators.

Key Findings

The results indicate that no ILSA indicator needs to be elaborated on in the DepEd system and interim assessment policy guidelines. Moreover, the DepEd interim assessment policy guidelines document has no gap as regards the ILSA standards.

ILSA Standards Needing Elaboration in the K to 12 Assessment Policy Guidelines

There are five ILSA key standards needing elaboration in the CA and NA policy guidelines.

Classroom Assessment

Although the CA clearly states that assessment entails teachers gathering evidence to understand what learners comprehend and are capable of doing, there is a need to elaborate on what specific unfamiliar contexts and real-life situations students' knowledge and skills can be transferred as exemplified in PISA's major subject domains' frameworks.

Additionally, ILSA items were developed following a certain set of standards; as such, these items can serve as exemplars for teachers' use and to construct similar test items for formative and summative assessments.

National Assessment

Like in CA, ILSA test items can serve as exemplars for developing test items. Additionally, ILSA test items include open-ended response format tests, allowing students to demonstrate a deeper understanding of concepts, apply critical thinking skills, and articulate reasoning (OECD, 2019). This response format enables the test item to assess higher-order thinking skills that mirror real-world problem-solving situations where students can apply their knowledge and skills in meaningful and authentic contexts. It allows distinguishing between a surface from a deep understanding and prevents guessing.

ILSA documents ensure the representation of the different item classifications provided in the framework. For example, PISA has a well-defined framework that allows item profiling and provides tables specifying items' desired distribution based on their difficulty, mathematical processes, content

category, and context category. A well-specified item classification and a table of desired distribution ensure a well-spread of items representing each category type. The DepEd NA policy guidelines articulated the need for a Table of Specifications (TOS) but failed to include its contents.

An assessment should not subject the test-takers to unnecessary stress and fatigue due to vague questions, non-availability of the correct answer in the choices, and use of convoluted vocabulary. ILSA's stimulus material and questions are crafted using clear, simple, succinct language yet effectively conveying the intended meaning (OECD, 2019). This highlights the important role that language plays in assessment. The test development in ILSAs includes a rigorous language translation process so that students are not disadvantaged due to their linguistic background. A clear articulation of the detailed translation process that includes back-translation and expert validation is crucial for maintaining the assessment as valid, high-quality, accurate, and able to mitigate bias (National Research Council, 2002).

ILSA Standards for Gaps for Adoption to the K to 12 Assessment Policy Guidelines

There are three ILSA key standard indicators for adoption in the CA, NA, and SA policy guidelines.

Classroom Assessment

ILSA frameworks are regularly revisited, revised, and updated to keep up with the latest assessment developments from research that represent global assessment standards. Although the CA policy guidelines state that summative assessment is an assessment of learning that informs decisions regarding future learning trajectories and career compatibility, there is a need to include more comprehensive assessment practices (e.g., collaborative problem-solving and global citizenship) that can measure and develop new sets of skills to meet the changing workforce needs driven by technological advancements, automation, and shifts in industry sectors.

Some ILSAs make use of computer-based assessments (CBA) owing to the fact that our learners are digital natives and the advantages CBA brings. CBA at the classroom level offers (1) immediate feedback, allowing students to identify their strengths and weaknesses; (2) automated tasks such as in test administration, scoring, and data analysis, which improve efficiency and unburden educators; (3) multi-media integration; and (4) environment friendly (Thelwall, 2000).

National Assessment

ILSA frameworks continue to evolve and improve in the succeeding cycles of their implementation. ILSAs often contract or consult experts in education, psychometrics, and related fields in the assessment development process. E.g., PISA reviews test results, updates reporting of students' performance, and adds proficiency levels if needed. The national assessment document mentioned the use of levels of progression in reporting national assessment results such as the Basic Education Exit Assessment (BEEA) and that proficiency level should be at least 75%. Notwithstanding, the output quality at the national level may still be enhanced to mimic PISA since a clear, accurate, and comprehensive reporting of assessment results helps stakeholders make informed decisions and interventions through meaningful insights into students' strengths, weaknesses, and progress.

The NA policy guidelines can include how test items in specific subject areas can be at par with international standards. This entails a comprehensive examination of various factors, including the design of the assessments, the alignment with curriculum standards, the rigor in the framework of the development of test items, and the quality of the scoring and reporting processes. The national test items can be compared with established international test items by analyzing their content, cognitive

complexity, and relevance to desired learning outcomes as a way of benchmarking for improving test items.

Some ILSAs utilize adaptive testing that enables precise measurement while using fewer items per student to avoid fatigue or stress and for better assessment engagement. Different sets of questions may be prepared, and the administration of tests at the national level may include strategic seating arrangements to avoid cheating and maintain test integrity.

System Assessment

Aside from the CA, using CBAs at the national and system levels can facilitate adopting adaptive testing and speed up data analysis since the administration, scoring, and data generation are automated.

Key Recommendations for Policy and Practice

On the basis of the above findings, the following are the key policy recommendations:

1. Specify what assessing students' knowledge and skill transferability in unfamiliar contexts would entail. Contexts and situations students can relate to may include, but are not limited to their personal, occupational, and educational use.
2. Regularly revisit, revise, and update the CA and NA policies to keep up with the latest trends and developments in international assessments to address contemporary concerns, issues, and research.
3. Teachers and NA test developers can adopt ILSA test items for improved assessment practices.
4. Consider using different response formats besides the multiple-choice or the selected-response type of items in NA.
5. Ensure a well-balanced spread of test items through a well-defined test item classification for profiling and a table of desired distribution of items.
6. The item bank should have a test item profile for acceptable items for storage.
7. Elaborate on the qualifications of the consultants and specialists in item development and validation and articulate the detailed language translation process.
8. Consider using an adaptive testing format where the difficulty level of an item is based on students' cognitive abilities for better assessment engagement, avoiding fatigue among test-takers, and ensuring the integrity of the assessment.
9. Improve the quality of assessment output reports at the national assessment.
10. Provide regular and on-time trend analysis on national and international assessment results for prompt intervention.
11. Consider CBAs in CA, NA, and SA—upgrade testing centers' infrastructures to accommodate computer-based assessments.

Conclusion and Recommendations

Student assessment as reflected in the K to 12 classroom, national, and system assessment policies needs enhancement to incorporate the standards in international large-scale assessments.

Concerned and respective government agencies and bureaus are enjoined to collaborate and engage in further discussions to consider these policy recommendations in their guidelines in close consultation with division superintendents, school principals, and teachers.

About the Authors

Minie Rose C. Lapinid, Ph.D., is an Associate Professor of the Department of Science Education at De La Salle University – Manila. Her research interests include mathematics education, educational technology, and educational assessment.

Voltaire M. Mistades, Ph.D., is an associate professor and the current Chairperson of the Department of Science Education at De La Salle University – Manila. His current research interests include science and physics education.

Richard R. Sagcal is a Department of Science Education graduate student at De La Salle University – Manila. His current research interests include science, chemistry, and physics education.

Leah E. Gustilo, Ph.D., is a full professor and research fellow of the Department of English and Applied Linguistics and the founding editor-in-chief of the Journal of English and Applied Linguistics. Her current research interests include language teaching, linguistics, and discourse analysis.

Marilyn U. Balagtas, Ph.D., is currently the Vice-President for Academics and University Professor of the Philippine Normal University System. Her current research interests include educational assessment, teacher education, educational research, and mathematics education.

Richard DLC. Gonzales, Ph.D., is a part-time faculty member at the Philippine Normal University, Cagayan State University, and University of Santo Tomas Graduate School. He is also an international development consultant for education and human resources development. His current research interests include educational psychology, educational assessment, and language learning.

Cesar T. Medula, Jr., Ph.D., is a professor at the Philippine Normal University. His current research interests include science education, learning and technology.

Brando C. Palomar, Ph.D., is an associate professor of science education and currently a senior program manager of the Research Institute for Teacher Quality of the Philippine Normal University. His current research interests include science education and educational assessment.

Ria Lorraine Pagliawan is currently a research officer at the Philippine Normal University. Her current research interests include mental health, social psychology, parasocial behaviors, and educational assessment.

Acknowledgments

The authors gratefully acknowledge the funding support provided by the National Research Council of the Philippines (NRCP). They also thank the officials of the Philippine Normal University (PNU), Cagayan State University (CSU), and Dela Salle University (DLSU) for the support and additional resources throughout the duration of this study.

References

Hernández-Torrano, D., Courtney, M.G.R. (2021). Modern international large-scale assessment in education: an integrative review and mapping of the literature. *Large-scale Assess Educ* 9, 17. https://doi.org/10.1186/s40536-021-00109-1

Organization for Economic Cooperation and Development. (2023). Programme for International Student Assessment (PISA) 2022 Results: Philippines - Country Note. https://www.oecd.org/publication/pisa-2022-results/country-notes/philippines-a0882a2d/

Organization for Economic Cooperation and Development. (2019). PISA 2018 assessment and analytical framework. PISA, OECD Publishing, Paris. https://doi.org/10.1787/b25efab8-en.

National Research Council. (2002). *Methodological Advances in Cross-National Surveys of Educational Achievement*. Washington, DC: The National Academies Press. https://doi.org/10.17226/10322.

Thelwall, M. (2000). Computer-based assessment: A versatile educational tool. *Computers & Education, 34*(1), 37-49. https://www.sciencedirect.com/science/article/abs/pii/S0360131599000378?via%3Dihub