

Aligning Philippine K to 12 Assessment Policies against International Benchmarks: Implications for Quality Reform

Minie Rose C. Lapinid^{1*}, Voltaire M. Mistades¹, Richard R. Sagcal^{1,4},
Leah E. Gustilo¹, Marilyn U. Balagtas², Richard D.L.C. Gonzales³,
Cesar T. Medula Jr.², and Brando C. Palomar²

¹Brother Andrew Gonzalez College of Education,
De La Salle University, Manila City, NCR 1004 the Philippines

²Philippine Normal University, Manila City, NCR 1004 the Philippines

³College of Humanities and Social Sciences, Graduate School,
Cagayan State University Carig Campus,

Tuguegarao, Cagayan, Region 2 3500 the Philippines

⁴Senior High School Department, Quezon City Science High School,
Quezon City, NCR 1005 the Philippines

This study employs document analysis to examine the alignment between Philippine K-12 assessment policy guidelines and international large-scale assessments (ILSAs). The aim is to benchmark and enhance Philippine assessment standards by identifying ILSA key standard indicators that are well-articulated, require further elaboration, or are missing but could be adopted in the Philippine Department of Education (DepEd) assessment policy guidelines. Four DepEd assessment policy guidelines—classroom assessment, national assessment, system assessment, and interim assessment—alongside ILSA documents were coded and categorized under the Sustainable Development Goals (SDGs)—specifically SDG4 (Quality Education), SDG5 (Gender Equality), and SDG 10 (Reduced Inequality). The coded ILSA key standard indicators were then mapped against those in DepEd documents. The analysis reveals both coherence and gaps between DepEd policies and ILSAs, pointing to the need for more detailed articulation of standards and the adoption of additional ILSA standards in Philippine guidelines. While some alignment in terms of quality, equality, and equitable key assessment standards was noted, areas needing enhancement include [1] defining qualifications for expert consultants, [2] balancing test-item difficulty and classification in assessments, [3] using sample items in ILSAs as exemplars in assessments, and [4] tagging approved items in the item bank with classification details. Recommended ILSA standards for adoption include [1] updating item development frameworks to address current issues and trends, [2] implementing adaptive testing and computer-based test administration, and [3] linking test results with students' backgrounds. These findings underscore the importance of assessments that ensure validity, reliability, fairness, and timeliness while informing policy formulation, reforms, and improved assessment practices.

Keywords: document analysis, international benchmarks, assessment, policy guidelines, international large-scale assessments, quality reforms

*Corresponding author: minie.lapinid@dlsu.edu.ph

INTRODUCTION

Human capital is necessary for economic growth, making quality education paramount and indispensable for its development (Berman *et al.* 2019). Hanushek and Woessmann (2012) found that student achievement significantly predicts a country's economic growth, positioning education as a key development index (Saniel *et al.* 2022). Educational quality refers to the comprehensive standards and effectiveness encompassing various aspects – including efficiency, equity, and holistic learner development influenced by pre-formal education, health, early experiences, home support, and learning environments in shaping a quality learner (Skedsmo and Huber 2021). A comprehensive literature review highlights the multi-faceted nature of educational quality, elucidating key facets “which include physical, psychosocial, and service delivery components” (Skedsmo and Huber 2021).

Assessment is integral to achieving quality education, providing feedback to enhance teaching, identifying learning gaps, and informing instructional decisions (Adarkwah 2021). Formative assessment motivates students (Leenknecht *et al.* 2020) to learn and help build their need for autonomy and competence. This connects instructions and learning processes with learners' needs and intended learning outcomes (Campbell and Levin 2009). National assessment in many countries is used as a mechanism for transparency and accountability of their educational system in compliance with set standards (Berman *et al.* 2019; Skedsmo and Huber 2021). Finally, student performance in standardized tests like the Program for International Student Assessment (PISA) can predict GDP growth (Hanushek and Woessmann 2012), linking education to workforce readiness (Shute *et al.* 2016).

While assessment is indispensable for quality education, it can inadvertently introduce potential partiality that may disadvantage specific cohorts of children (Skedsmo and Huber 2021). Tests can both promote fairness and perpetuate inequities in education (Berman *et al.* 2019). For instance, Getenet and Beswick (2021) emphasize addressing socioeconomic disparities and test validity, particularly for indigenous students. Similarly, Şenel's (2021) research on Turkey's Central Examination highlights biases favoring non-visually impaired students. In mid-western China, improved teacher support and training are needed to institutionalize formative assessment practices (Chen *et al.* 2021).

International large-scale assessments (ILSAs) are standardized tests administered to students globally (Rocher and Hastedt 2020) to measure student achievement in specific disciplines and evaluate the quality of educational systems (Jimenez and Modaffari 2021). As

such, ILSAs are often used as proxy measures to examine different countries' human capital in an increasingly competitive global economy (Berman *et al.* 2019). The large data sets from such assessments can provide insights into pedagogical issues, teacher performance, and classroom practices to educational policymakers, researchers, and stakeholders (Berman *et al.* 2019). ILSAs inform policies and reform (Strietholt and Scherer 2018); foster accountability and transparency for the efficient use of public funds and combating complacency and mediocrity (Berman *et al.* 2019; Skedsmo and Huber 2021; Campbell and Levin 2009); allow countries to benchmark against high-ranking countries (Berman *et al.* 2019; OECD 2019); and help identify support and interventions for underperforming students and schools, as well as offering incentives for high performers (Berman *et al.* 2019; UNICEF and SEAMEO 2019; Jimenez and Modaffari 2021).

ILSAs are widely recognized and participated in by many countries, setting high standards for assessment quality that any country can emulate for benchmarking purposes. For instance, 81 countries involving about 700,000 students took part in PISA 2022 (OECD 2023). These exemplary assessment practices within ILSA frameworks include components such as purpose and goals, assessment design (content, format, and administration procedure), sampling, data collection and analysis, validity and reliability, cross-national comparability, equity and fairness, and interpretation of test results are used to evaluate and compare educational outcomes across different countries. As such, this present study refers to these exemplary assessment practices as key standard indicators.

The ILSAs that are mentioned in the Department of Education (DepEd) system assessment policy guidelines in the Philippines are enumerated for each key stage (DepEd 2017). ILSA's Key Stage 1 for students in Kindergarten to Grade 3 includes the International Development and Early Learning Assessment (IDELA) (Save the Children n/d), Early Grade Reading Assessment (EGRA) (RTI International 2016), and Early Grade Mathematics Assessment (EGMA) (Platas *et al.* 2014). ILSAs for Key Stage 2 for Grades 4–6 students contain the Progress in International Reading Literacy Study (PIRLS) (Mullis and Martin 2019; DepEd 2017), South East Asia Primary Learning Metrics (SEA-PLM) (UNICEF and SEAMEO 2019), and Trends in International Mathematics and Science Study (TIMSS) Grade 4 (Mullis and Martin 2019). ILSA's Key Stage 3 for students in Grades 7–10 comprises TIMSS Grade 8 (Mullis and Martin 2019) and PISA (OECD 2019). Each is tailored to measure specific educational domains and skills at different grade levels, contributing to a global holistic understanding of educational assessment practices.

The K to 12 Basic Education Program has been in place for more than 10 years and with the new *Matatag* curriculum, there has been a clamor prompting to revisit, recalibrate, and update the assessment policy guidelines provided by the DepEd for teachers' reference to better prepare students for international standards. DepEd's assessment policies include various directives encompassing DepEd Order (DO) No. 8, s. 2015 (classroom assessment); DO No. 55, s. 2016 (national assessment); DO No. 29, s. 2017 (system assessment); and DO No. 31, s. 2020 (interim assessment). This scenario serves as the context for this study, which aimed to inform the Philippine basic education system by identifying the standards of ILSAs for policy recommendations to ensure quality, equality, and equity. Specifically, the research study sought to answer the following question: which ILSA standards are well articulated, need more elaboration, and are not expressed in the DepEd provisions under Sustainable Development Goals (SDGs) quality, equality, and equity? To the best of the researchers' knowledge, no study has been conducted analyzing assessment policy guidelines *vis-à-vis* ILSA framework documents for benchmarking purposes. Thus, this research study attempts to fill this literature gap.

MATERIALS AND METHODS

Research Design

This study utilized qualitative content analysis (Kyngäs *et al.* 2020), whereby the documents analyzed served as the main data sets, using systematic coding and interpreting textual material to understand patterns, themes, and meanings within the data.

Materials and Coding Procedures

The latest version of the aforementioned seven ILSA framework documents and four DepEd assessment policy guidelines from January–November 2023 were analyzed, following a hierarchical structure and well-defined order to accurately extract information from the policy. This was pivotal in effectively managing and interpreting the data for analytical purposes. Three researchers each coded 2–3 ILSA documents, classifying policy provisions according to three specific United Nations (UN) SDGs: SDG 4 for quality education indicators (test construction, teachers, materials, and administration), SDG 5 for gender equality, and SDG 10 for reducing inequalities among student populations (*e.g.* contextual learner factors) (UN n/d). Each researcher then reviewed one other's work to ensure accuracy.

A separate group of four researchers, each working on one DepEd document, followed the same coding procedure. Each researcher independently verified another's coding

by tracing back to the original document text for context verification. Relevant DepEd document references were noted for each code.

Two layers of reviews and validation were conducted by a pair of validators to correct any deviations from coding parameters. Revisions were coded in red beside the original to track changes, and each reviewer added their name for transparency and accountability. In each review phase, researcher coders and validators held an online meeting to discuss comments requiring revision, reconsideration, removal, or addition.

Researcher coders mapped concepts between DepEd policy guidelines and ILSA standards, aligning each ILSA standard with the relevant SDG category in the DepEd provisions. Coding parameters shown in Appendix Table I categorized provisions as well-articulated, needing elaboration, or missing provisions in the DepEd documents. Well-articulated provisions fully capture standards, those needing elaboration partially capture standards, and missing provisions are absent standards recommended for adoption. Each category has specific criteria to determine alignment level and guide necessary improvements or adoption actions.

One validator then reviewed the concept matchups and their categorization based on the SDGs and whether the ILSA provisions were well articulated, needed more elaboration, or were absent in the DepEd documents. Additionally, the validator indicated the actions “retain,” “revise,” and “reject” across each mapping. Of the 158 coded contents mapped (and not mapped for absence), no code was rejected, and there were only nine codes suggested for revision – yielding 94.3% agreement, signifying an almost perfect agreement. Suggestions were given on how to revise the mapping. An online meeting between the coders and the validator was conducted to discuss the latter's suggestions and clarifications. The finalized mapping was agreed upon. Then, the analysis was summarized into tables, and texts lifted from the documents were used as evidence to support each identified standard indicator. Please see Appendix Figure I for the visual presentation of the qualitative data analysis.

RESULTS

The presentation of the results between the DepEd assessment policy guidelines and the ILSAs comprises three sets: [1] concepts well-articulated, [2] those that need more elaboration, and (3) those not present in DepEd assessment policy guidelines. Each of these are discussed in turn. In each set, indicators under SDGs 4, 5, and 10 are delineated. An excerpt of the content in the document analyzed that exemplifies our points and the document

where this concept was lifted up can be found in their corresponding tables in the Appendix section.

Standards Articulated in the Assessment Policy Guidelines

Six ILSA quality standards were found to be well articulated in DepEd documents – including specific schedules and regularity of the assessments, clear purpose statements and definitions of assessments, aligning assessments with essential learning competencies, consideration of students' learning, ensuring standard administration of assessment, and the inclusion of assessment of 21st-century skills.

Equality Standards Articulated in the Assessment Policy Guidelines

ILSAs include gender as a variable in student background questionnaires to provide insights and analyze gender differences. The PISA framework identifies engagement and metacognition as predictors of reading achievement, with gender mediating performance, and examining financial literacy, thus supporting SDG 5 on gender equality (OECD 2019). The National Assessment Policy Guidelines (NAPG) incorporates this in the examinee's profile and allows providing test results by gender upon request by external stakeholders (DepEd 2016).

Equity Standards Articulated in the Assessment Policy Guidelines

Appendix Table III lists three ILSA equity key standards that are well-articulated in DepEd assessment policy guidelines. First, DepEd acknowledges the importance of participating in assessments but does not impose; instead, it promotes understanding through advocacy activities, including orientations before tests and debriefings afterward to explain their purposes and allow students to share experiences (DepEd 2017). ILSAs help inform policy decisions such as determining aspects of the education system that need change, frequency and type of teacher training, safe space provision for reported bullying incidents, and retention policies – all of which help address poor performance. For example, PISA provides data to evaluate grade repetition policies and their impact on education costs (OECD 2019). Second, DepEd gives regard to the test-taker's health, safety, and overall well-being through conducting advocacy activities in the form of orientations and debriefings (DepEd 2017). During the pandemic, the interim assessment policy guidelines guided assessments in remote or hybrid settings, prioritizing the mental health of both teachers and students (DepEd 2020). Lastly, it is commendable that the System Assessment Policy Guidelines (SAPG) stresses the importance of covering diversity and unique contexts

across different subject areas, languages, geographical locations, and practically different learning environments. NAPG enumerated the different assessments provided to different types of learners based on their needs (*e.g.* exit assessments in each key stage, career assessment for determining learners' aptitudes and occupational interests aligned to the Senior High School tracks, and accreditation and equivalency for continuing education and placement) and allows testing personnel and examinees to exercise their religion and consideration for learners under special circumstances and those with special needs (DepEd 2016).

Quality Standards Needing Elaboration in the Assessment Policy Guidelines

Some key indicators in the DepEd assessment policy guidelines require further elaboration for clearer directives. Appendix Table IV lists eight key standards for quality – primarily from the NAPG and Classroom Assessment Policy Guidelines (CAPG), with one from the SAPG – covering aspects of validity and reliability, variation in test formats and response types, item distribution, and proficiency scales.

ILSAs measure learners' application of knowledge and skills in unfamiliar settings. For instance, EGMA emphasizes numeracy skills for problem-solving, highlighting their significance in daily life (Platas *et al.* 2014). SEA-PLM recognizes writing as a literacy skill in real life, which is a new domain in ILSA (UNICEF and SEAMEO 2019). Both CAPG and SAPG stress the importance of applying knowledge in new real-life contexts, which ILSAs elaborately clarify in various contexts (*e.g.* PISA's reading literacy in personal, occupational, and educational contexts). Additionally, the policy guidelines may specify that teachers may use the available ILSA test items as resources to help students acclimate to ILSA tests.

Educational assessment should exemplify the knowledge, processes, strategies, practices, and habits that students and teachers should develop [Fredericksen (1984), as cited by Bennett (2015)]. Ideally, test items should accurately measure complex cognitive abilities such as critical thinking and articulating thoughts, as well as resistance to cheating. However, such items often necessitate extended testing periods, complicating the test administration process.

The CAPG distinguishes how early grade-level pupils in kindergarten should be graded compared to those in Grades 1–12. However, ILSA tests are developmentally appropriate. For example, EGRA, EGMA, and IDELA are administered orally and individually to young children (RTI International 2016; Save the Children n/d), who might otherwise be put at a disadvantage if administered in paper-and-pen, where young beginner reader pupils are left alone to answer by themselves. The

test administrator sits with the child and follows a scripted protocol, allowing them to demonstrate what they have learned including motor skills like hopping on one foot, copying a shape, or engaging in socio-emotional tasks like conflict resolution or empathy in IDELA (Save the Children n/d). The CAPG outlines various testing types and formats but should explicitly include provisions for oral and individual administration for Kindergarten and Grade 1, incorporating age-appropriate visuals and larger typeset in font sizes.

ILSAs maintain strict quality control measures; for instance, EGMA demonstrates predictive, content, and convergent validities, whereas PISA adheres to high psychometric standards for global educational comparisons (Strietholt and Scherer 2018). DepEd's NAPG outlines the test development process – including defining roles within DepEd, formulating a table of specifications (TOS), and consulting experts for framework enhancement. For transparency purposes, the NAPG should mandate Bureaus to report on the contracted services [*e.g.* PISA mentioned contracted services; see OECD (2019)], the specific frameworks used in national assessment, the distribution of test items, the translation process (*e.g.* the inclusion of back translation by language experts), and the qualifications of the consultants to avoid biases (Jimenez and Modaffari 2021).

PISA and SEA-PLM utilize both closed-constructed and open-constructed response items, allowing students to demonstrate their reasoning and problem-solving processes, which provides an in-depth understanding of their thinking (OECD 2019; UNICEF and SEAMEO 2019). While this format offers valuable insights, it requires manual scoring and trained raters to ensure score validity (Kuramoto and Koizumi 2018). The National Achievement Test (NAT) could benefit from including varied response formats beyond multiple-choice to minimize guessing, as multiple-choice tests prioritize efficiency over depth, potentially fostering rote learning (Bennett 2023). Familiarity with ILSA-type tests can alleviate anxiety and enhance metacognitive self-regulatory behavior among learners.

ILSAs provide clear frameworks and item examples to illustrate item classifications and criteria, as seen in SEA-PLM, which outlines specific skills related to the text (*e.g.* brainstorming, structuring and organizing, managing coherence, selecting diction, and managing syntax and grammar) (Courtney 2021). Sample items can serve as valuable resources for teachers to create similar assessments without merely teaching to the test (Bennett 2023). Japan's practice of publishing national assessment questions emphasizes their educational value and sets clear expectations for students (Kuramoto and Koizumi 2018). Additionally, this practice clearly communicates

expectations to students regarding the tests.

The CAPG specifies cognitive process dimensions based on Anderson and Krathwol's concept development model [2001, cited in a work by DepEd (2015)], whereas the NAPG mandates a TOS for well-balanced item distribution across learning objectives and competencies. However, ILSAs ensure different item classifications provided in the framework. Consequently, items eligible for inclusion in the item bank for organizing the national assessment's final form must have item classification and a distribution table per item classification must be provided to ensure a well-spread of items.

PISA incorporates technology for mathematical problem formulation and EGMA permits counters for problem-solving (OECD 2019; RTI International 2014). While CAPG lists assessment tools, it can explicitly allow specific tools for assessment purposes, as long as these do not violate test validity and reliability.

The PISA database is publicly accessible to anyone (OECD 2019). It is suggested to develop a similar mechanism of data storage, offering a model for national assessments to provide data management and dissemination without compromising test taker privacy. Data can be made available to school administrators and teachers, in addition to providing these results to students using a report card, instead of making the data available upon request.

The national assessment document mentions that the proficiency level should be at least 75%. However, the quality of output at the national level can be improved just like SEA-PLM and PISA where each level achieved based on the correct answers obtained by the student may be used to describe students' achievement in the classroom and national tests for more descriptive and accurate assessment result reporting (OECD 2019; UNICEF and SEAMEO 2019). Secondly, proficiency scales are constantly reviewed and refined in TIMSS and PISA such that scales in the upper and lower levels are extended based on the item response theory of empirical data. This way, stakeholders are assured of assessment output and reporting quality.

Aside from these, reporting can include linking assessment results to various student demographics and trend indicators “between student-level, school-level, and system-level background variables and outcomes” (OECD 2019). Notwithstanding, the government should help build capacity among teachers and school leaders to properly utilize the needed data – when they need it – to guide student learning and help inform decision-making. For instance, what support has already been introduced to address at-risk students? Were these corrective interventions promptly implemented (Yi *et al.* 2021)?

Equity Standards Needing Elaboration in the Assessment Policy Guidelines

Appendix Table V shows the three ILSA equity standard indicators that need more articulation in the NAPG: [1] prioritizing test-taker welfare; [2] improving translation quality for mother tongue accuracy; and [3] utilizing test results for comparative analysis across regions, divisions, districts, and schools for priority interventions.

ILSA developers strive for clear instructions and concise items to avoid misinterpretation and ensure each question assesses only the intended competency. An assessment should not subject the test-takers to unnecessary anxiety, stress, fatigue, and disengagement caused by vague questions leading to multiple interpretations, the absence of correct answers from the choices, the use of convoluted vocabulary, and long duration of test administration. Thus, it is essential to include provisions ensuring that tests follow assessment principles. For younger test-takers such as those at the kindergarten level, stop rules or breaks in between can also be introduced to reduce fatigue.

The Basic Education Exit Assessment (BEEA) assessment uses both English and Filipino languages, making accurate translation essential for fairness. Including this in the NAPG would promote equity, as clear translation prevents misinterpretation of questions and options (DepEd 2016).

The NAPG specifically allows ELLNA data to be categorized by competency skills, learners with special needs, school topology, type of school, and school location (DepEd 2016). However, it could better specify analysis for immediate intervention for budgetary considerations. Including a framework and standardized procedures as comprehensive as the ILSAs, would enable cross-country data comparison (Strietholt and Scherer 2018). A cross-region analysis by DepEd could identify educational strengths and weaknesses for tailored interventions, especially in marginalized and conflict-affected areas.

Quality Standard Gaps in the Assessment Policy Guidelines

Appendix Table VI lists three critical quality standards for inclusion in DepEd's assessment policy guidelines – namely establishing a framework for item development, utilizing matrix sampling, and adaptive testing format and computer-based methods.

PISA's triennial administration, rotating among the three major domains is crucial for trend analysis and benchmarking. But more importantly, PISA frameworks evolve to reflect new developments, ensuring relevance and effectiveness (OECD 2019). The frameworks are revisited, revised, and updated in each cycle to address contemporary theories, practices, issues, and trends. While CAPG mentions collaborative learning to enhance 21st-

century skills, and SAPG uses ILSAs as a proxy indicator to track skill development, the NAPG and CAPG could guide teachers and developers by referencing these ILSA frameworks, which are theory-based, consultative, and comprehensive. PISA's framework for item development includes various dimensions of knowledge and skills and provides illustrative items (OECD 2019). SEA-PLM employs common tasks across grade levels to measure longitudinal progress (UNICEF and SEAMEO 2019). In accordance with this recommendation, the initial report of the Second Congressional Commission on Education (EDCOM II) recommends creating a unified framework across all levels, with regional and division standardizations (EDCOM II 2024).

An additional strategy PISA and PIRLS employ to address testing fatigue is matrix sampling of **test items** to reduce testing time, scoring duration, and assessment costs by rotating test items across multiple booklets, grouping them by units, problem scenarios, difficulty, and alignment with 15-year-old students' developmental levels (Mullis and Martin 2019). Students take different tests as there are available sets of tests by placing clusters in booklets in a rotated design.

Additionally, PISA is moving towards an adaptive testing format tailored to students' cognitive abilities to enhance measurement precision with fewer items to reduce fatigue and stress. This type of assessment provides greater resolution and sensitivity, particularly for students performing at lower levels by offering questions that match student's skill levels (OECD 2019). Likewise, PIRLS now incorporates three levels of passage complexity (challenging, typical, and simple) that are merged into two tiers of booklet difficulty. Every country conducts the full assessment with the ratio of challenging and easy booklets differs based on the students' reading skills in that country (Mullis and Martin 2019).

NAPG can include such provisions or at least similar measures to help test-takers avoid pressure and fatigue while maintaining test results' integrity such as creating parallel test sets and implementing strategic seating to prevent cheating while allowing cross-regional comparisons. For example, the PISA has 66 test forms, yet cross-country results are still comparable. Aside from these strategies, although unconventional, providing examinee choice and learner agency in language, response types, interests, and cultural identity – and can result in slightly higher ability estimates, a slight reduction in test anxiety, promote test engagement, and motivation to perform, leading to a perception that the assessment is fair [Pitkin and Vispoel (2001), as cited by Bennett (2023)].

Some ILSAs, previously paper-and-pen-based tests are transitioning to computer-based assessment due to its adaptive feature, increased accuracy, and rapid result processing (Shute *et al.* 2016; RTI International 2016). EGRA switched to electronic data collection in 2010 to minimize errors, whereas PISA and ePIRLS now offer digital formats, with ePIRLS simulating online reading tasks (RTI International 2016; Mullis and Martin 2019). Computer-based assessments are cost-effective, provide advanced formats like its drag-and-drop feature, and allow for real-time scoring, offering valuable data on response times and processes that inform policy and research (Shute *et al.* 2016; Strietholt and Scherer 2018; OECD 2023).

As digital technology becomes integral in post-pandemic education, computer-based assessments can be considered for inclusion in NAPG, at least in areas with adequate local funding and infrastructure. This provision may require investments in hardware, networking, technology training, and proctoring resources. EDCOM II's (2024) Year One report recommends this to facilitate timely test result release and analysis. Nevertheless, risks associated with computer-based assessments exist such as widening social gaps between students with access to the required infrastructure and those without (Wikström and Eklöf 2019).

Equity Standard Gaps in the K to 12 Assessment Policy Guidelines

A lone key equity gap in national assessments is the integration of learning assessment data with students' backgrounds. PISA links information on student academic performance with data on students' demographics and attitudes, school attributes, and educational systems that influence their learning, whereas PIRLS and SEA-PLM include student attitudes, self-beliefs, and contextual factors to better understand learning needs (Mullis and Martin 2019; UNICEF and SEAMEO 2019). IDELA even assesses health and hygiene factors (Save the Children n/d).

NAPG can include such analysis by aggregating DepEd's school data, the learner information system, and other contextual factors with student achievements to identify tailored interventions. Research supports this approach; for example, Bernardo *et al.* (2022) identified factors affecting Filipino students' performance in PISA math using machine learning, revealing differences between private and public school vulnerabilities. Furthermore, ILSA data sets have been used to inform decisions such as hiring more teachers, raising teacher qualifications, improving resource distribution, and increasing early childhood education investment (Carnoy *et al.* 2016).

DISCUSSION

An effective assessment tool provides valuable information about students' learning, knowledge, and skills. Quality assessments are characterized by validity, reliability, fairness, clarity of purpose, authenticity, feasibility, format variety, timeliness, timely and constructive feedback quality, transparency, security of materials, and integrity of results (Strietholt and Scherer 2018). These are what ILSAs can provide for educational research, revealing not only student performance but also the significance and nature of learning within the local environment (Moss *et al.* 2006). Consequently, assessments should guide teachers in improving their instruction, professional development planning, curricula and textbook adoption, development of more targeted and specific interventions for struggling learners, and engaging stakeholders for support (Moss *et al.* 2006).

Of the four assessment policy guidelines, the NAPG can adopt more changes, followed by the CAPG and the other two DepEd policy guidelines. Countries like the United Kingdom, the United States, South Africa, Canada, and Australia utilized national assessments to track performance against international benchmarks, primarily for monitoring and accountability (Getenet and Beswick 2021). Specifically, identifying learners at risk at an early stage and addressing these promptly through effective interventions that could mitigate learning deficits (Yi *et al.* 2021).

Given the high stakes of national assessments that influence students' career paths and school incentives, quality assessments are essential. However, proposed changes may entail resource allocation coupled with challenges requiring upskilling teachers to design, create, and administer these assessments, especially for large class sizes, and under-resourced schools. Thus, implementation must be carefully considered to achieve desired results (Cerna 2013) as implementing it on a larger scale will take time, careful consideration, revision, iteration, and political skills (Bennett 2023).

CONCLUSION

The four DepEd assessment policy guidelines have articulated quality and equity in education following the ILSAs. Nonetheless, there are still other aspects the Philippine DepEd can adopt from the ILSAs that need more clarifications and additions in the policy guidelines for teachers' reference and better assessment practices at the national level. The DepEd assessment policy guidelines are reference documents for teachers, test developers, and concerned DepEd Bureaus in charge of

assessments. As such, well-articulated guidelines convey the expected standards for assessment practices. Quality assessments identify gaps in learning outcomes regardless of student background, ensuring quality education through immediate feedback and on-point intervention programs. An equitable assessment guarantees the inclusion of marginalized students for a fair assessment, which is necessary for inclusive education and reducing inequalities.

Throughout the research, quality, equality, and equity have been explicitly defined within the SDG framework, recognizing the significance of establishing precise interpretations of these crucial concepts for the study. However, it must be noted that the investigation was limited to this specific context, and other perspectives or contexts may offer different interpretations or variations of these terms. As a result, the findings should be cautiously generalized beyond the scope of the identified referenced SDGs. This limitation was acknowledged as an inherent constraint that requires attention and awareness.

RECOMMENDATIONS

Educational policies in different countries often translate international standards to meet local educational needs, leading to diverse outcomes (Steiner-Khamsi *et al.* 2022). ILSA results have triggered policy reforms such as Norway's, which was driven by their standing in PISA and PIRLS (Tveit 2014; IEA 2011). New Zealand utilized TIMSS data to guide significant education reforms (Korsnakova *et al.* 2021), and both Germany and Denmark established national standards aligned with PISA to support low-performing and disadvantaged students (Fischman *et al.* 2019). The Philippines could adopt a similar strategy by integrating ILSA results into its curriculum to align with PISA and other international standards. Countries like Finland and Estonia have also reformed their education systems based on ILSA results. Even with its high performance, Finland refined its assessment strategies to uphold educational equity and quality (Dobbins and Martens 2012).

At present, the Philippine assessment system could improve by adopting adaptive tests to capture student learning differences, promoting inclusivity and reflecting student development. Examples from various countries highlight the importance of diverse approaches to educational reform, which the Philippines could leverage to enhance its education policy. Future research should focus on global methods and best practices applicable to the Philippine context.

This study identified gaps in the DepEd assessment policy

guidelines by benchmarking against ILSAs, prompting a reevaluation of assessment practices and encouraging dialog to improve Philippine education. However, caution is necessary, as lessons from high-performing countries show potential adverse effects such as increased anxiety and test aversion (Kuramoto and Koizumi 2018). The risks of standardized testing, including a narrowed curriculum and pressure to teach to the test, have been noted in contexts like the United States "No Child Left Behind" Act, which shifted teaching methods away from holistic education (Teltemann and Jude 2019).

Moreover, having a policy does not guarantee effective implementation (Cerna 2013), as evidenced by delays in test result releases following the SAPG's three-month timeline. The EDCOM II Year One report (2024) indicated that 24 out of 27 key stage assessments faced issues, with 13 delayed and 11 canceled. Furthermore, the unavailability of baseline and end-line assessment results hampers progress monitoring and timely interventions. These challenges are scribed to procurement and staffing issues in DepEd's Bureau of Education Assessment, highlighting that successful policy implementation relies on local context, shared values, managerial skills, and the capacity of implementers (Cerna 2013).

Teachers need adequate support with the newly instituted Philippine MATATAG curriculum, which is now being piloted in several schools. Despite the perceived decongested curriculum, teachers expressed challenges in time constraints, lack of teaching and learning resources, and limited facilities (EDCOM II 2024). Particularly in assessment, the study by Garcia *et al.* (2024) showed that teachers struggle to create realistic PISA-like items. Additionally, future studies can investigate teacher assessment literacy and competence in **test-item** construction, scoring, and interpreting results (Herppich *et al.* 2018). Lastly, there is a need to examine how DepEd assessment policy guidelines are being implemented on the ground, at the classroom and school levels, for documenting best practices and strengths worth emulating, and identify areas for appropriate intervention.

ACKNOWLEDGMENTS

This study was made possible through the funding of the Department of Science and Technology–National Research Council of the Philippines, as well as the leadership of the Philippine Normal University. Our sincere gratitude for providing the team the opportunity to conduct the study and the support extended to its fulfillment.

DECLARATION OF AI USE

The Discourse Analyzer kit was used to refine the stylistic features to achieve grammaticality and improve the readability of the text. No AI was used in the generation of text and in the analysis of the data.

REFERENCES

- ADARKWAH MA. 2021. The power of assessment feedback in teaching and learning: a narrative review and synthesis of the literature. *SN Social Sciences* 1(75). doi.org/10.1007/s43545-021-00086-w
- BENNETT RE. 2015. The changing nature of educational assessment. *Review of Research in Education* 39: 370–407. doi.org/10.3102/0091732X14554179
- BENNETT RE. 2023. Toward a theory of socioculturally responsive assessment. *Educational Assessment* 28(2): 83–104. doi.org/10.1080/10627197.2023.2202312
- BERMAN AI, FEUER MJ, PELLEGRINO JW. 2019. What use is educational assessment? *The ANNALS of the American Academy of Political and Social Science* 683(1): 8–20. doi.org/10.1177/0002716219843871
- BERNARDO ABI, CORDEL II MO, LAPINID MRC, TEVES JMM, YAP SA, CHUA UC. 2022. Contrasting profiles of low-performing mathematics students in public and private schools in the Philippines: insights from machine learning. *Journal of Intelligence* 10(61). doi.org/10.3390/jintelligence10030061
- CAMPBELL C, LEVIN B. 2009. Using data to support educational improvement. *Educational Assessment, Evaluation, and Accountability* 21(1): 47–65. doi.org/10.1007/s11092-008-9063-x
- CARNOY M, KHAVENSON T, LOYALKAP, SCHMIDT WH, ZAKHAROV A. 2016. Revisiting the relationship between international assessment outcomes and educational production: evidence from a longitudinal PISA-TIMSS sample. *American Educational Research Journal* 53(4): 1054–1085. doi.org/10.3102/0002831216653180
- CERNA L. 2013. The nature of policy change and implementation: a review of different theoretical approaches. Paris: OECD Publishing. Retrieved on 05 Jan 2023 from <https://web-archival.oecd.org/2014-07-16/310811-the%20nature%20of%20policy%20change%20and%20implementation.pdf>
- CHEN Q, ZHANG J, LI L. 2021. Problematizing formative assessment in an undeveloped region of China: voices from practitioners. *Educational Assessment, Evaluation, and Accountability* 33(4): 649–673. doi.org/10.1007/s11092-021-09369-5
- COURTNEY L. 2021. Real life language: comparing student writing skills in SEA-PLM. British Council, London. Retrieved on 12 Feb 2024 from https://asiapacificml-bucket.s3.us-east-2.amazonaws.com/image/BC-Academic%20papers-10.Louise%20Courtney%20%5BA4%5D_Jul_2.pdf
- [DepEd] Department of Education. 2015. DepEd Order No. 8, s. 2015 on policy guidelines on the classroom assessment in the K to 12 Basic Education Program. Retrieved on 05 Jan 2023 from https://www.deped.gov.ph/wp-content/uploads/2015/04/DO_s2015_08.pdf
- [DepEd] Department of Education. 2016. DepEd Order No. 55, s. 2016 on policy guidelines on the national assessment in the K to 12 Basic Education Program. Retrieved on 05 Jan 2023 from https://www.deped.gov.ph/wp-content/uploads/2016/06/DO_s2016_55-3.pdf
- [DepEd] Department of Education. 2017. DepEd Order No. 29, s. 2017 on policy guidelines on system assessment in the K to 12 Basic Education Program. Retrieved on 05 Jan 2023 from https://www.deped.gov.ph/wp-content/uploads/2017/06/DO_s2017_029.pdf
- [DepEd] Department of Education. 2020. DepEd Order No. 31, s. 2020 on interim guidelines for assessment and grading in light of the basic education learning continuity plan. Retrieved on 05 Jan 2023 from https://www.deped.gov.ph/wp-content/uploads/2020/10/DO_s2020_031.pdf
- DOBBINS M, MARTENS K. 2012. Towards an education approach à la finlandaise? French education policy after PISA. *Journal of Education Policy* 27(1): 23–43. <https://doi.org/10.1080/02680939.2011.622413>
- [EDCOM II] Second Congressional Commission on Education. 2024. Miseducation: the failed system of Philippine education [EDCOM II year one report]. Retrieved on 03 Feb 2024 from <https://edcom2.gov.ph/media/2024/02/EDCOM-II-Year-One-Report-PDF-022924.pdf>
- FISCHMAN GE, TOPPER AM, SILOVA I, GOEBEL J, HOLLOWAY JL. 2019. Examining the influence of international large-scale assessments on national education policies. *Journal of Education Policy* 34(4): 470–499.
- GARCIA S, GAYLO DN, VISTRO-YU CP. 2024. Writing PISA-like mathematics items: the case of tertiary mathematics instructors from a state university in the Philippines. *European Journal of Educational Research* 13(4): 1441–1453. <https://doi.org/10.12973/eu-jer.13.4.1443>
- GETENET S, BESWICK K. 2021. Predictors of children’s achievement: analysis of the Australian national nu-

- meracy assessment program. *Educational Assessment, Evaluation, and Accountability* 33: 591–620. doi.org/10.1007/s11092-021-09364-w
- HANUSHEK EA, WOESSMANN L. 2012. Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth* 17: 267–321. doi.org/10.1007/s10887-012-9081-x
- HERPPICH S, PRAETORIUS AK, FÖRSTER N, FREY IG, KARST K, LEUTNER D, BEHRMANN L, BÖHMER M, UFER S, KLUG J, HETMANEK A, OHLE A, BÖHMER I, KARING C, KAISER J, SÜDKAMP A. 2018. Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education* 76: 181–193. doi.org/10.1016/j.tate.2017.12.001
- [IEA] International Association for the Evaluation of Educational Achievement. 2011. TIMSS and PIRLS – informing educational policy for improved teaching and learning. Accessible at https://timssandpirls.bc.edu/home/pdf/TP_Impact_Statement.pdf
- JIMENEZ L, MODAFFARI J. 2021. Future of testing in education: Effective and equitable assessment systems. Center for American Progress. Retrieved on 05 Jan 2023 from <https://www.americanprogress.org/wp-content/uploads/sites/2/2021/09/EquitableAssessmentSystems1.pdf>
- KORSNAKOVA P, MEINCK S, JAPELJ PB. 2021. Introduction to Dinaric perspectives in TIMSS 2019. In: *Dinaric perspectives on TIMSS 2019*, Vol. 13. Korsnakova P, Meinck S, Pavesic BJ eds. Cham: Springer. p. 1–14.
- KURAMOTO N, KOIZUMI R. 2018. Current issues in large-scale educational assessment in Japan: focus on national assessment of academic ability and university entrance examinations. *Assessment in Education: Principles, Policy, and Practice* 25(4): 415–433. doi.org/10.1080/0969594X.2016.1225667
- KYNGÄS H, MIKKONEN K, KÄÄRIÄINEN M. 2020. The application of content analysis in nursing science research. Springer. <https://sci-hub.ru/10.1007/978-3-030-30199-6>
- LEENKNECHT M, WIJNIA L, KÖHLEN M, FRYER L, RIKERS R, LOYENS S. 2020. Formative assessment as practice: the role of students' motivation. *Assessment and Evaluation in Higher Education* 46(2): 236–255. doi.org/10.1080/02602938.2020.1765228
- MOSS PA, GIRARD BJ, HANIFORD LC. 2006. Chapter 4: validity in educational assessment. *Review of Research in Education* 30: 109–162. doi.org/10.3102/0091732X030001109
- MULLIS IVS, MARTIN MO. 2019. PIRLS 2021 assessment frameworks. TIMSS and PIRLS International Study Center website. Retrieved on 05 Jan 2023 from <https://timssandpirls.bc.edu/pirls2021/frameworks/>
- [OECD] Organization for Economic Co-operation and Development. 2019. PISA 2018 assessment and analytical framework. Paris: OECD Publishing. Retrieved on 05 Jan 2023 from <https://www.oecd-ilibrary.org/deliver/b25efab8-en.pdf?itemId=%2Fcontent%2Fpublication%2Fb25efab8-en&mimeType=pdf>
- [OECD] Organization for Economic Co-operation and Development. 2023. PISA 2022 results: the state of learning and equity in education. Paris: OECD Publishing. Retrieved on 05 Jan 2023 from <https://www.oecd-ilibrary.org/deliver/dfe0bf9c-en.pdf?itemId=%2Fcontent%2Fpublication%2Fdfef0bf9c-en&mimeType=pdf>
- PLATAS LM, KETTERLIN-GELLAR L, BROMBACHER A, SITABKHAN Y. 2014. Early grade mathematics assessment (EGMA) toolkit. Research Triangle Institute. Retrieved from https://ierc-public-files.s3.amazonaws.com/public/resources/EGMA%20Toolkit_March2014.pdf
- ROCHER T, HASTEDT D. 2020. International large-scale assessments in education: a brief guide IEA compass. In: *Briefs in Education* No. 10. Amsterdam, the Netherlands: IEA.
- RTI INTERNATIONAL. 2014. Early grade mathematics assessment toolkit. Retrieved on 05 Jan 2023 from https://ierc-publicfiles.s3.amazonaws.com/public/resources/EGMA%20Toolkit_March2014.pdf
- RTI INTERNATIONAL. 2016. Early grade reading assessment toolkit. Retrieved on 05 Jan 2023 from <https://www.iea.nl/publications/series-journals/iea-compass-briefs-education-series/september-2020-international-large>
- SANIEL DMT, PARCUTILO JO, MARIQUIT TM. 2022. Educational policy analysis: policy formulation and implementation of the five statutory policies in the Philippine basic education. Retrieved from https://www.researchgate.net/publication/371255751_EDUCATIONAL_POLICY_ANALYSIS_POLICY_FORMULATION_AND_IMPLEMENTATION_OF_THE_FIVE_STATUTORY_POLICIES_IN_THE_PHILIPPINE_BASIC_EDUCATION
- SAVE THE CHILDREN. n/d. International development and early learning assessment. Retrieved on 05 Jan 2023 from <https://idela-network.org/about/>

- ŞENEL S. 2021. Assessing measurement invariance of Turkish “Central Examination for Secondary Education Institutions” for visually impaired students. *Educational Assessment, Evaluation, and Accountability* 33(4): 621–648. doi.org/10.1007/s11092-020-09345-5
- SHUTE VJ, LEIGHTON JP, JANG EE, CHU MW. 2016. Advances in the science of assessment. *Educational Assessment* 21(1): 34–59. https://doi.org/10.1080/10627197.2015.1127752
- SKEDSMO G, HUBER SG. 2021. What does educational quality mean? *Educational Assessment, Evaluation, and Accountability* 33(4): 587–589. doi.org/10.1007/s11092-021-09378-4
- STEINER-KHAMSI G, KARSETH B, SIVESIND K. 2022. Evidence and expertise in Nordic education policy: a comparative network analysis. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-91959-7
- STRIETHOLT R, SCHERER R. 2018. The contribution of international large-scale assessments to educational research: combining individual and institutional data sources. *Scandinavian Journal of Educational Research* 62(3): 368–385. doi.org/10.1080/00313831.2016.1258729
- TELTEMANN J, JUDE N. 2019. Assessments and accountability in secondary education: international trends. *Research in Comparative and International Education* 14(2): 249–271. https://doi.org/10.1177/1745499919846174
- TVEIT S. 2014. Educational assessment in Norway. *Assessment in Education: Principles, Policy, and Practice* 21(2): 221–237. doi.org/10.1080/0969594x.2013.830079
- [UN] United Nations. n/d. Sustainable development goals. Retrieved on 05 Jan 2023 from https://www.un.org/sustainabledevelopment/
- [UNICEF] United Nations Children’s Fund, [SEAMEO] Southeast Asian Ministers of Education Organization. 2019. SEA-PLM 2019 assessment framework (1st ed.). SEA-PLM Secretariat, Bangkok, Thailand. Retrieved on 10 Jan 2023 from https://www.seaplm.org/PUBLICATIONS/frameworks/sea-plm%202019%20assessment%20framework.pdf
- WIKSTRÖM C, EKLÖF H. 2019. Educational assessments in digital environments: insights from different assessment contexts. *Education Inquiry* 10(1): 1–3. doi.org/10.1080/20004508.2019.1568779
- YI HS, NA W, LEE C. 2021. Predicting secondary school student's academic achievement from their elementary school performance and learning behaviours: a longitudinal study based on South Korea's national assessment of educational achievement. *Asia Pacific Journal of Education* 43(4): 1031–1048. doi.org/10.1080/02188791.2021.1963673

APPENDIX

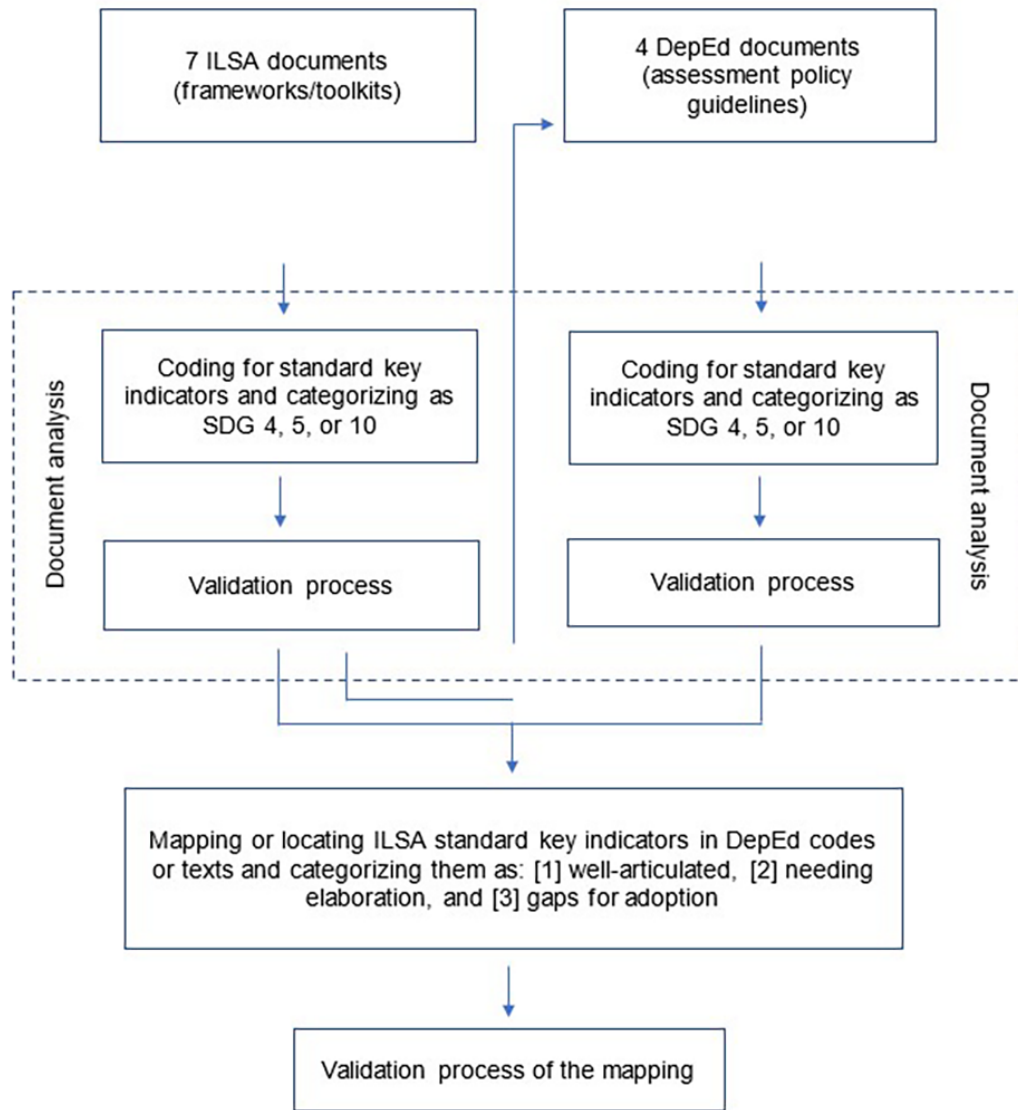


Figure I. Qualitative data analysis procedure.

Table I. Definitions and criteria of provision categories.

Categories	Definitions	Criteria
Well-articulated provisions	These provisions in DepEd assessment policy guidelines fully capture the essence and meaning of the ILSA standards	The provisions in the DepEd assessment policy guidelines align completely with the ILSA standard; no additional details or modifications are necessary
Provisions needing more elaboration	These provisions in DepEd assessment policy guidelines partially capture the concept of the ILSA standard but are deemed lacking or needing more explicit articulation	The provision in the DepEd assessment policy guidelines aligns with the ILSA standard to some extent; additional details or modifications are needed to fully represent the ILSA standard
Missing standards	These are gaps where key standard indicators in the ILSAs do not match any of those in the DepEd policy guidelines	There is no alignment between ILSA key standard indicators and the DepEd Policy guidelines; these standards are absent from the DepEd policy guidelines; they are recommended for adoption in the DepEd policy guidelines

Table II. ILSAs' quality key standards that are well-articulated in DepEd assessment policy guidelines.

ILSA key standards	DepEd assessment policy guidelines
<p>Assessments are longitudinal and regularly administered at intervals for monitoring purposes</p> <p>PISA is a triennial assessment launched in 1997 in reading, mathematics, and science literacies as subject content areas, with each cycle focusing on one subject area as its major domain (OECD 2019, p. 3).</p> <p>TIMSS is being conducted every four years since its first assessments in 1995 (Mullis and Martin 2019).</p>	<p>DepEd conducts the following assessments annually beginning School Year 2016–2017: exit assessments, career assessment to be administered annually, accreditation and equivalency assessment, and grade level placement assessment – with the latter two assessments even accommodating walk-in takers (DepEd 2016)</p> <p>"Early Language, Literacy, and Numeracy Assessment (ELLNA) and the National Achievement Test (NAT) will be administered to the universal population every three years starting in 2018" (DepEd 2017, p. 4)</p>
<p>Clear purpose statements</p> <p>TIMSS is a valuable resource for monitoring the educational system's effectiveness in science, technology, engineering, and mathematics (STEM), which are key areas in the basic education curriculum (Mullis and Martin 2019).</p> <p>PISA assesses and monitors trends on what 15-yr-old students around the world, nearing their end of compulsory education, "can extrapolate from what they have learned and apply their knowledge in new situations" (OECD 2019, p. 12).</p>	<p>The national assessment aims to: "a. monitor the Philippine education system and schools for public accountability; [c] provide information that will guide decisions on instructional practices; [e] measure students' aptitude and occupational interest for career guidance; and assess prior learning for placement, accreditation, and equivalency." (DepEd 2016, p. i)</p> <p>The system assessment policy guidelines are issued for the following purposes: "a. establish baselines for the basic education system and the implementation of the K to 12 curriculum in schools in terms of teaching and learning; [b] monitor the implementation of the K to 12 curriculum in schools in terms of teaching and learning; [c] measure effectiveness of instructional reforms that are part of the K to 12 basic education program ..." (DepEd 2017, p. i)</p>
<p>Essential learning competencies: rationale for assessing particular skills justified, provided, or explained</p> <p>TIMSS utilizes the curriculum model, with the three aspects – namely intended, implemented, and attained curriculum – as the main concept in view of student educational opportunities and how they use these opportunities (Mullis and Martin 2019).</p> <p>EGMA and EGRA spelled out the essential learning competencies for reading literacy and numeracy skills (RTI International 2016)</p>	<p>The CAPG specifies that a formative assessment can be conducted during the teaching and learning process to check whether learners are progressing and achieving the learning competencies; the summative assessment could be conducted towards the end of a learning period to determine the standard attained by the learner.</p> <p>The national assessment aims to "assess the effectiveness and efficiency of the delivery of education services using learning outcomes as indicators; [d] determine if learners are meeting the learning standards of the curriculum" (DepEd 2016, p. 1)</p> <p>"The alternative learning system accreditation and equivalency (ALS A&E) assessment and certification tests are nationally administered to measure the competencies" (DepEd 2016, p. 17).</p> <p>The interim assessment explicitly states that "assessment should be holistic and authentic in capturing the attainment of the most essential learning competencies (MELCs)" (DepEd 2020, p. 1)</p>
<p>Considers the role of students' prior knowledge and understanding in the formal schooling and assessment in Key Stage 1 regarding concept development</p> <p>The core EGMA assumes that children possess informal mathematical skills prior to formal schooling (Platas <i>et al.</i> 2014) based on their daily activity experience and observations</p>	<p>The classroom assessment policy guidelines specify a variety of appropriate assessments are to be given to diverse learners of different contexts, "cultural backgrounds, and life experiences" (DepEd 2015, p. 1).</p>
<p>Standard implementation and administration of assessment</p> <p>A five-day adaptation workshop and training were held to ensure standard implementation and administration; a detailed description of activities in the workshop and procedure of the administration of EGMA is provided (RTI International 2014)</p>	<p>"Specific test administration procedures – including different forms to be accomplished before, during, and after the test – are indicated in the examiner's handbook that will be distributed and discussed during the National Conference/ Consultative Workshop of field-testing personnel to be conducted by BEA" (DepEd 2016, p. 7).</p>
<p>Assesses 21st-century skills</p> <p>21st-century skills are being assessed in PISA – namely collaborative problem solving, financial literacy, critical thinking skills, information literacy, technology literacy, and creative thinking skills, among others (OECD 2019).</p>	<p>"Collaborative formative assessment allows learners to support one another's learning" (DepEd 2015, p. 4)</p> <p>Exit assessment for Grades 6, 10, and 12 "shall cover 21st-century skills (information, media, and technology skills, learning and innovation skills, communication skills, and life and career skills)" (DepEd 2016, p. 9)</p>

Table III. ILSAs' equity key standards that are well-articulated in DepEd assessment policy guidelines.

ILSA key standards	DepEd assessment policy guidelines
<p>Implications of the assessment results for poor-performing students: the need for support and intervention</p> <p>“The results of these analyses can give Ministries of Education the ability to target interventions such as teacher training on specific topics or sub-regions” (RTI International 2016, p. 3)</p>	<p>“It is important to note that to be able to benefit all learners, the results of system performance assessment must connect with program delivery; this will help (schools, divisions, or regional level field offices) to provide relevant technical assistance to schools in the preparation of programs that address learning needs: school improvement plans that directly address teaching and learning improvement concerns in the school should always be based on insights gained from studying data” (DepEd 2017, p. 3)</p>
<p>Considers learner participants and other assessment stakeholders in the assessment process</p> <p>On justifying the use of stop rules: “Students in testing environments are already stressed, and the added stress of repeated struggles to provide answers to questions they do not know creates even more undesirable testing environment; stop rules enable an assessor to stop the administration of a subtest when the student has incorrectly answered a specific number of items in a row (generally four)” (RTI International 2016, p. 9).</p>	<p>“Recognizes the role of learners as co-participants in the assessment process” (DepEd 2015, p.1)</p> <p>Test data are disseminated and reported to stakeholders (DepEd 2016).</p> <p>The National Conference Consultative Workshop of field-testing personnel is being conducted with the examiner’s handbook distributed and discussed during the conference (DepEd 2016).</p> <p>“It is necessary to utilize alternative tools and strategies for assessing and supporting learning while avoiding creating undue pressure on teachers, students, and their families” (DepEd 2020, p. 1).</p>
<p>The assessment reflects and covers diversity and unique contexts; the assessment is sensitive to gender, racial, ethnic, and religious affiliations of examinees and stakeholders</p> <p>“A definition of reading literacy must reflect contemporary developments in school and societal literacy demands, namely the increasing amount of text information available in print and digital forms and the increasing diversity and complexity of situations involving text and reading” (OECD 2019, p. 32)</p> <p>On justifying assessing global competence, “By appreciating the differences in the communities to which they belong – the nation, the region, the city, the neighborhood, the school – young people can learn to live together as global citizens” (p. 166) and “to support the sustainable development goals” (p. 167) on education to support sustainable development and lifestyles while fostering a culture of peace and valuing cultural diversity (OECD 2019).</p> <p>“To ensure fairness given the inter-country and gender differences observed and to ensure a valid assessment of the reflecting and evaluating process, both multiple choice and open constructed-response items continue to be used in PISA” (OECD 2019).</p> <p>Text topics in PIRLS are themes appropriate to the grade level; ethnic, racial, gender, religious contemplation and sensitivity; density of information; and linguistic nature and features (Mullis and Martin 2019).</p>	<p>There is a national assessment for each key stage – ELLNA at the end of Grade 3; Exit Assessment in Grades 6, 10, and 12; and Career Assessment in Grade 9 (DepEd 2016).</p> <p>Administration of “accreditation and equivalency assessment for out-of-school youth and adults” and “grade level placement assessment for learners in special circumstances to determine their appropriate grade level in the formal system” (DepEd 2016, p. ii)</p> <p>Testing personnel and examinees are given free exercise of their religion (DepEd 2016)</p> <p>Annex 1 of SAPG shows the different national and international assessments covering key stages and their details (DepEd 2017)</p>

Table IV. ILSAs’ Quality Key Standards needing elaboration in DepEd assessment policy guidelines

ILSA key standards	DepEd assessment policy guidelines
<p>There is an emphasis on contextual learning and real-life applications to diverse learners in contexts/situations:</p> <p>PISA specified and described the situations and contexts – personal use, occupational use, educational, and the like – and provide each with examples (OECD 2019).</p> <p>"Mathematics enables us to investigate, organize, understand, and describe the world around us" (RTI International 2016 p. 7)</p>	<p>Classroom performance standards include answering the questions: "How well do learners use their learning or understanding in different situations?" and "How do learners apply their learning or understanding in real-life contexts?" (DepEd 2015, p. 3)</p> <p>"Summative assessment should measure student learning in applying their learning in different contexts" (DepEd 2015, p. 2).</p> <p>"Performance tasks give students opportunities to demonstrate and integrate their knowledge, understanding, and skills about topics or lessons learned in a specific real-life situation by performing and/or producing evidence of their learning" (DepEd 2015, p. 8)</p> <p>"Teachers should not limit the assessment methods they use to the examples provided in the table on page 6" (DepEd 2015, p. 5)</p>
<p>Assessment is age-appropriate, <i>e.g.</i> Kindergarten or Grade 1 children, can be assessed orally and individually for their numeracy and reading skills.</p> <p>"EGMA, an orally administered assessment of the core mathematical competencies taught in primary grades" (RTI International 2014, p. 1)</p> <p>"Why assess orally? An oral assessment [EGRA], therefore, can give us more information about what students actually do know and where they are in the reading acquisition process early on" (RTI International 2016, p. 4–5)</p>	<p>Kindergarten learners will be issued checklists and anecdotal records instead of numerical grades for recording of learner progress, whereas learners from Grades 1–12 are graded numerically (DepEd 2015).</p> <p>Classroom assessment can be in the form of oral or written works (DepEd 2015).</p>
<p>Stringent quality assurance mechanisms (validity and reliability measures, field trial, IRT use, rigorous sampling, translation, <i>etc.</i>)</p> <p>"PISA is the most comprehensive and rigorous international programme to assess student performance and to collect data on the student, family, and institutional factors that can help explain differences in performance. Decisions about the scope and nature of the assessments and the background information to be collected are made by leading experts in participating countries and are steered jointly by governments on the basis of shared, policy-driven interests; substantial efforts and resources are devoted to achieving cultural and linguistic breadth and balance in the assessment materials; stringent quality-assurance mechanisms are applied in translation, sampling and data collection; as a consequence, results from PISA have a high degree of validity and reliability" (OECD 2019 p. 13)</p> <p>"The core EGMA is designed to predict later achievement (predictive validity) and align with curricula that support the development of foundational mathematical skills (content validity), ... to identify the linkages among those skills (convergent validity), to discriminate among types of mathematical knowledge, and to differentiate among skill levels" (RTI International 2014, p. 8)</p> <p>Several pilot tests of EGMA were carried out from which validity and reliability of instrument were drawn out (RTI International 2014).</p>	<p>Section 7 of the NAPG listed the test development process and provided a flowchart for its visual representation (see DepEd 2016, p. 24)</p>

Table IV. Cont.

Use different response formats aside from the usual multiple-choice type of response in large-scale assessments	"Written work may include long quizzes, essays, written reports, and unit or long tests" (DepEd 2015, p. 7)
Vary the assessments to assess students' critical thinking	"Summative tests may be in the form of objective tests, performance tasks, or a combination thereof" (DepEd 2015, p. 8)
"The response format varies depending on the kinds of evidence that is being collected and also according to the pragmatic constraints of large-scale assessment" (OECD 2019, p. 48).	"Assessment facilitates the development of learners' higher-order thinking ... skills" (DepEd 2015, p. 1)
"A further consideration ... is that open constructed response items are particularly important to assess the reflecting and evaluating process, where the intent is often to assess the quality of a student's thinking rather than the student's final response itself" (OECD 2019 p. 49).	
SEA-PLM also has both formats for learners to describe their understanding and thinking processes (UNICEF and SEAMEO 2019).	
There is a good balance and spread of items regarding difficulty; representation of the different item classifications is provided in the framework and tables of desired distribution per item category	"The adapted cognitive process dimensions may be used as a guide not only in lesson development but also in the formulation of assessment tasks and activities" (DepEd 2015, p. 4)
PISA's item difficulty is a function of task and source dimensions (see OECD 2019, p. 44).	Preparation of TOS by subject area in grade level placement assessment is assigned to BEA in consultation with DepEd's Bureau of Curriculum Development (see DepEd 2016, p. 21)
"Desired distribution of items by mathematical process" (OECD 2019, p. 80)	"Acceptable items shall be stored in the item bank for the organization of the final form" (DepEd 2016, p. 24)
"Desired distribution of items by content category" (OECD 2019, p. 86)	
"Desired distribution of items by context category" (OECD 2019, p. 89)	
Tools and materials may be provided to learners for use during the test administration as long as this does not affect the test validity and reliability in terms of attaining the assessment's purpose	Appendix A lists the different assessment tools (see DepEd 2015, p. 27).
The process of formulating situations mathematically in PISA involves "using technology (such as spreadsheet or the list facility on a graphing calculator) to portray a mathematical relationship inherent in a contextualized problem)" (OECD 2019, p. 78)	"Instructional materials and aids posted on the classroom walls should be covered" (DepEd 2016, p. 25)
"Assessors keep track of whether the student used one of three problem-solving strategies: finger/tick marks, paper and pencil calculation, or solved problem in his or her head; students are also provided with counters that can be used to solve the problem" (RTI International 2014, p. 14)	"Each testing room should have the following materials: [a] test materials ..., [b] table and chair for the room examiner, [c] enough seats for the examinees, [d] name grid, [e] board work, [f] pencils, [g] extra sheet of paper for computation" (DepEd 2016, p. 25)
Database of international large-scale assessments are made available to stakeholders and researchers	Advocacy activities are organized and conducted for stakeholders to understand, accept, and appreciate the value of national and international large-scale assessment; learners are briefed, debriefed, and given the results when released (DepEd 2017)
PISA data are readily available to anyone and can be downloadable (OECD 2019)	
Proficiency scales were constantly reviewed based on empirical data and were modified to capture students' capabilities best	CAPG shows the percentages of the different assessment in computing for student grades (DepEd 2015)
"PISA reports students' results through proficiency scales that can be interpreted in educational policy terms" (OECD 2019, p. 53)	"Test scores shall be reported as percentage scores; proficiency level for each cluster of early language, literacy, and numeracy skills is at least 75% (DepEd 2016, p. 7)
PISA provides both quantitative and qualitative results for proficiency levels; the qualitative results are descriptions of what students can do (see OECD 2019, p. 55)	

Table V. ILSAs' equity key standards needing elaboration in DepEd assessment policy guidelines.

Key standards	DepEd assessment policy guidelines
<p>Considers test-takers' welfare in the assessment process</p> <p>"PISA's stimulus material and questions use language that is as clear, simple, brief, and syntactically simplified as possible while still conveying the appropriate meaning; the number of concepts introduced per paragraph is limited; questions within the domain of science that assess reading or mathematical literacy are avoided" (OECD 2019, p. 113).</p> <p>Adaptation workshop, assessor training, pilot study, and instrument finalization are conducted for EGMA prior to its actual implementation (RTI International 2016)</p> <p>"Fatigue can affect outcomes on items towards the end of assessment due to inattention, boredom, anxiety, and cumulative cognitive load" (RTI International 2014, p. 9)</p>	<p>Test takers are briefed before the test and debriefed after the test, and they are given time to go over their test results upon receiving them (DepEd 2016)</p>
<p>Stringent quality assurance mechanisms (IRT use, translation, and the like)</p> <p>"An important aspect of validity is the elimination of possible confounds ... include the language in which the assessment is conducted (low scores might reflect a lack of language-of-assessment skills rather than mathematical skills)" (RTI International 2014, p. 8)</p>	<p>NAPG mentions ELLNA's language can be in English, Filipino, and in 19 mother tongue languages, but there was no mention of ensuring the validity of the translation process (DepEd 2016)</p> <p>SAPG mentioned that the language of assessment in SEA-PLM is "in English except for the global citizenship wherein the Philippines has the option to translate it into local language (Filipino)" (DepEd 2017, p. 13)</p>
<p>Assessment results are to be used to pinpoint regions needing immediate attention and additional support</p> <p>"EGRA can be used by low-income countries to pinpoint regions (or if the sample permits, schools) that merit additional support, including teacher training or other intervention" (RTI International 2016, p. 18)</p>	<p>Section 2 on NAPG for ELLNA data requirements states, "The data should be presented according to the performance profile: [a] competency skills, [b] learners with special needs, [c] school typology, and [d] school location" (DepEd 2016, p. 4), whereas Section 3 for the exit assessment states that "The performance profile shall be presented according to: [a] type of school, [b] competency skills, [c] learners with special needs, and [d] school location" (DepEd 2016, p. 11).</p>

Table VI. ILSAs' quality key standards that are missing in DepEd assessment policy guidelines.

Key standards	Recommended for inclusion in DepEd assessment policy guidelines
Frameworks that serve as the basis for item development must be regularly reviewed and updated, comprehensive, and consultative addressing contemporary concerns, issues, and research.	NCSI
Use of matrix sampling	N
Adaptive testing and/or Computer-based administration	NSI

Legend: DepEd order policy guidelines on assessment documents: [C] classroom, [N] national, [S] system, and [I] interim

GLOSSARY OF TERMS

Formative assessment refers to "assessment for learning so teachers can make adjustments in their instruction; it is also assessment as learning wherein students reflect on their own progress" (DepEd 2015, p. 2).

Large-scale assessments measure "students' learning outcomes in particular learning areas or domains; these are wider and bigger in scope, and they are used to measure what learners know and can do based on a standard criteria and/or expectations; these are uniformly administered to a large number of students" [McDonnell *et al.* (1997), as cited by Saldaña and Shepard (2023, p. 163)].

Proxy indicator/measure is "an indirect measure or sign that approximates or represents a phenomenon in the absence of a direct measure or sign; it is also known as an indirect indicator" (DepEd 2016, p. 3).

System assessment refers to "how effectively and efficiently the education system delivered basic education *vis-à-vis* the set of indicators and *versus* its articulated targets on the Philippine Development Plan" (DepEd 2015, p. 2).

Interim assessment in this study refers to the assessment policy guidelines issued by DepEd for teachers and parents during the COVID-19 pandemic to supplement DepEd Order No. 12, s. 2020 “Adoption of the Basic Education Learning Continuity Plan (BE-LCP) for School Year 2020–2021” (DepEd 2020).

Public accountability refers to “the obligation of DepEd to carry out responsibilities that affect the public; it means that decision makers across all levels will explain publicly, fully, and fairly what education outcomes they intend to bring about, for whom, and why” (DepEd 2016, p. 3).

Summative assessment refers to “assessment of learning, which occurs at the end of a particular unit in order to describe the standard reached by the learner” (DepEd 2016, p. 3).

Adaptive testing refers to a test that “adapts or tailors exam questions in real time to the ability of each test taker; this eventually results in a different set of test questions for each person; the test adapts based on how well the test taker answers earlier questions” (Foster 2021).

Matrix sampling of items refers to “the division of a set of items into different versions of a test form” (Childs and Jaciw 2002).