DEVELOPMENT AND VALIDATION OF CONTEXTUALIZED PISA-BASED ASSESSMENT ITEMS IN FUNCTIONS FOR GRADE 11 MATHEMATICS
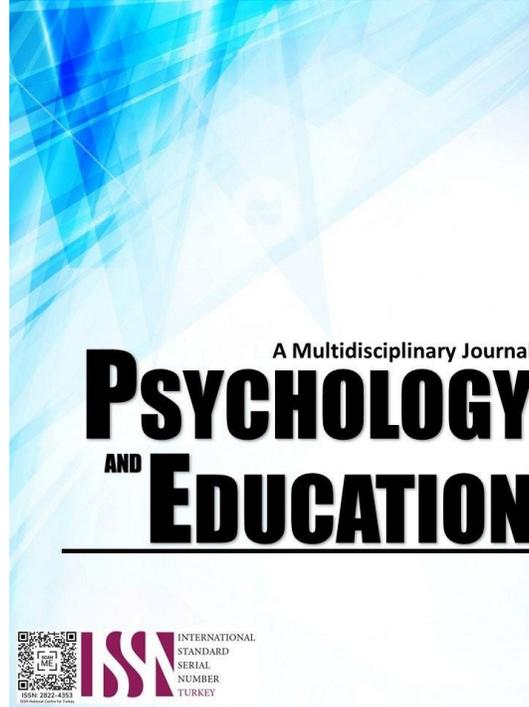
Development and Validation of Contextualized PISA-Based Assessment Items in Functions for Grade 11 Mathematics

Eloiza Jane O. Serdeña*
For affiliations and correspondence, see the last page.

Abstract

This study aimed to develop and validate contextualized Program for International Student Assessment (PISA)-based assessment items in Functions for Grade 11 Mathematics in a secondary school in the Imelda District, Zamboanga Sibugay. Using a mixed-methods design, specifically an explanatory sequential approach, the study first conducted quantitative procedures to determine the psychometric properties of a 70-item test, followed by qualitative procedures to gather student feedback for refinement. Content validation was performed by three mathematics experts and one language validator using the Content Validity Index (CVI). A total of 72 Grade 11 students participated in the pilot testing, which generated data for computing reliability, item-total correlation, discrimination index, and difficulty index. Cronbach's alpha was 0.889, indicating high reliability. Item analysis showed that 30 items were rated excellent and 37 were rated good in terms of content validity, while three items were rated poor. Based on discrimination and difficulty indices, 12 items were identified as most suitable for final use. Pretest results (M = 11.70, MPS = 23.39%) indicated Not Proficient performance, while posttest results (M = 15.72, MPS = 31.03%) indicated Low Proficient performance. A paired-samples t-test revealed a significant difference between pretest and posttest scores, t(71) = − 6.736, p = .001, demonstrating measurable improvement after exposure to contextualized items. Qualitative feedback from 15 students highlighted varied perceptions of difficulty, with themes of ease, moderate challenge, high difficulty, and mixed experiences. Overall, the study concluded that the contextualized PISA-based assessment items are valid, reliable, and contextually appropriate, although continuous refinement and targeted instructional support are needed to improve students' proficiency in Functions.

Keywords: *PISA-based assessment, functions, Grade 11 mathematics, contextualized assessment, student performance*

Introduction

Mathematics education increasingly emphasizes the development of students' ability to apply concepts in real-world contexts, particularly in the area of Functions. Traditional assessments in the Philippine classroom remain largely procedural, limiting students' opportunities to demonstrate reasoning, problem-solving, and mathematical literacy—competencies emphasized in the Program for International Student Assessment (PISA) (OECD, 2019). Research shows that contextualized and authentic tasks improve engagement and comprehension by situating mathematical ideas within students' lived experiences (Mitari & Zulkardi, 2019). However, many learners continue to struggle with interpreting problems, identifying relevant information, and applying conceptual understanding to unfamiliar situations, resulting in persistent performance gaps (Borgonovi et al., 2020; Jaudinez, 2019).

International and local studies indicate that while PISA-based assessments can enhance mathematical literacy, there is limited availability of validated, context-specific tools explicitly designed for senior high school mathematics in the Philippine setting. Barba et al. (2024) found only partial alignment between the Philippine Mathematics curriculum and PISA's cognitive demands, indicating a need for assessment instruments that reflect the competencies required in global benchmarks. Internal assessments in the Imelda District, Zamboanga Sibugay, further revealed low proficiency among Grade 11 learners, particularly in understanding and solving function-related word problems, suggesting gaps between curriculum intentions and actual student performance.

Despite the increasing advocacy for PISA-aligned assessment practices, few studies in the Philippines have systematically developed and validated contextualized PISA-based assessment items for Grade 11 Functions. This gap underscores the need for tools that are psychometrically sound, culturally relevant, and aligned with higher-order thinking skills. Addressing this need, the present study aims to develop and validate contextualized PISA-based assessment items that measure students' mathematical literacy and real-world problem-solving skills in the domain of Functions.

Research Questions

This study aimed to develop and validate contextualized PISA-based assessment items in Functions for Grade 11 Mathematics for use in a secondary school in the Imelda District, Division of Zamboanga Sibugay, during the School Year 2025–2026. The study examined the content validity, reliability, and item analysis of the developed assessment items and evaluated their suitability for measuring students' mathematical literacy and real-world problem-solving skills. Specifically, the study sought to answer the following research questions:

1. What is the content validity of the contextualized PISA-based assessment items in Functions for Grade 11 Mathematics based on the Content Validity Index (CVI)?

Research Article

2. What is the reliability of the contextualized PISA-based assessment items in Functions for Grade 11 Mathematics based on Cronbach's alpha?
3. What are the item-total correlation, difficulty index, and discrimination index of the contextualized PISA-based assessment items based on the item analysis results?
4. Based on the validity, reliability, and item analysis findings, which contextualized PISA-based assessment items are deemed suitable for final use?
5. What is the level of students' achievement based on the pretest and posttest results of the contextualized PISA-based assessment in Functions?
6. Is there a significant difference in students' achievement between their pretest and posttest results?
7. What feedback do students provide regarding the contextualized PISA-based assessment of items in Functions for Grade 11 Mathematics?

Literature Review

Mathematical literacy has become a central focus of contemporary educational reforms, particularly as schools aim to prepare learners for real-world problem-solving. The Program for International Student Assessment (PISA) defines mathematical literacy as the capacity to formulate, employ, and interpret mathematics in authentic contexts (OECD, 2019). This emphasis shifts mathematics learning from procedural computation to reasoning, communication, and higher-order thinking skills. Studies consistently highlight that traditional assessments often fail to measure these competencies, limiting students' opportunities to apply mathematics meaningfully (Schleicher, 2017; Andrade & Brookhart, 2020).

PISA-Based and Contextualized Assessment Practices

Across international contexts, PISA-based assessments have shown promise in evaluating students' ability to navigate non-routine, real-life problems. Mitari and Zulkardi (2019) demonstrated that contextualizing PISA-like tasks using culturally familiar scenarios enhances students' engagement and comprehension. Similar findings in Indonesia, Malaysia, and other Asian settings reveal that contextualized tasks strengthen problem interpretation and mathematical modelling skills (Kholid & Nissa, 2022; Shafa et al., 2023). These studies collectively indicate that contextualization is not merely an instructional strategy but a valid assessment approach that mirrors real-world cognitive demands.

Challenges in Mathematics Performance and Real-World Application

Despite curricular efforts, learners across various educational systems continue to struggle with problem-solving and the application of mathematical concepts. Borgonovi et al. (2020) noted that factors such as weak foundational skills, limited exposure to nonroutine tasks, and misconceptions about mathematics contribute significantly to low performance. In the Philippine context, Jaudinez (2019) reported persistent difficulties in interpreting word problems, while Barba et al. (2024) found only partial alignment between the Philippine Mathematics Curriculum and the PISA framework. These findings collectively reveal a mismatch between curricular goals and actual student outcomes, especially regarding competencies required for real-world application.

Instructional Quality, Equity, and Higher-Order Thinking

Research also emphasizes the role of instructional practices in enhancing or hindering students' development of higher-order thinking skills. Mitani (2021) found that opportunities to engage in inquiry-based and problem-centered learning significantly influence performance in large-scale assessments such as PISA and TIMSS. However, disparities in instructional quality persist, particularly in low-resource schools, where students often receive limited exposure to authentic, cognitively demanding tasks. This reinforces the need for assessments that serve both evaluative and instructional purposes by modeling the kinds of thinking expected in high-quality mathematics instruction.

Development and Validation of Assessment Instruments

Developing high-quality assessment tools requires rigorous validation procedures. Kane (2016) and Polit and Beck (2006) emphasized that valid assessments must reflect the intended constructs and undergo systematic evaluation through content validity, reliability testing, and item analysis. Classical Test Theory remains widely used in educational measurement, ensuring that items discriminate effectively and measure targeted competencies consistently (Reynolds et al., 2021). Studies developing localized assessment tools in the Philippines and abroad highlight the importance of integrating contextual relevance with psychometric soundness to ensure fairness and accuracy (Padilla & Songco, 2016; Mamolo & Sugano, 2021).

Across the literature, there is strong evidence that contextualized, PISA-based assessments improve mathematical literacy and problem-solving. However, most existing studies focus on broad mathematical competencies or junior high school contexts. There is limited research on the systematic development and validation of contextualized PISA-based assessment items specifically for Grade 11 Functions, despite documented learning gaps in this area. Furthermore, few Philippine studies integrate expert validation, item analysis, reliability testing, and student feedback within a single framework.

The present study addresses this gap by developing and validating contextualized PISA-based assessment items tailored to Grade 11

Eloiza Jane O. Serdeña

Functions, ensuring both contextual relevance and psychometric rigor. This contributes to improving assessment practices aligned with real-world mathematical competencies in the Philippine senior high school context.

Methodology

Research Design

This study employed an explanatory sequential mixed-methods design, consisting of a quantitative phase followed by a qualitative phase to explain and elaborate on initial statistical results. According to Creswell and Creswell (2018), this design is appropriate when researchers seek first to measure variables objectively and then use qualitative data to provide deeper insights into the quantitative findings. The design aligned with the study's objective of developing and validating contextualized PISA-based assessment items and understanding learners' perceptions of these items. The quantitative phase focused on establishing content validity, reliability, and item analysis of the instrument, while the qualitative phase gathered student feedback to support refinement.

Respondents

The study was conducted in a public secondary school in the Imelda District, Division of Zamboanga Sibugay, during the School Year 2023–2024. The quantitative respondents consisted of 72 Grade 11 learners enrolled in General Mathematics. These students were selected through cluster sampling based on their availability and participation in the pilot testing. For the qualitative phase, 15 students were purposively selected to represent varying levels of performance based on their test results, ensuring a range of perspectives. Expert validators included three mathematics specialists and one language expert who assessed the content, clarity, and contextual appropriateness of the items.

Instrument

The primary instrument of the study was a researcher-developed, contextually tailored PISA-based assessment in Functions, initially comprising 70 items. The instrument was designed to measure Grade 11 students' mathematical literacy, particularly their ability to formulate, utilize, and interpret mathematical concepts in real-world contexts, consistent with the PISA framework (OECD, 2019). The assessment included multiple-choice and problem-solving items structured around real-life scenarios contextualized within the learners' local environment.

Content validity was established through expert evaluation using the Content Validity Index (CVI) according to the procedures described by Polit and Beck (2006). Reliability was assessed through a pilot test administered to 72 learners, with internal consistency measured using Cronbach's alpha, consistent with Classical Test Theory (Field, 2018; Pallant, 2020). Item analysis included determining item difficulty, discrimination index, and item-total correlation to refine the final test version. For the qualitative phase, a semi-structured interview guide was used to gather student feedback regarding item clarity, relevance, and difficulty.

Procedure

Data collection proceeded in sequential stages. First, the initial set of assessment items was drafted based on PISA frameworks, curriculum guides, and contextualized situational analyses. These items were then submitted to expert validators, who provided quantitative ratings and qualitative comments. After revisions, the instrument was pilot tested with 72 Grade 11 students. The administration was conducted in a controlled classroom setting, and completed test papers were retrieved immediately to ensure data integrity. The results of the pilot test were encoded and subjected to reliability testing and item analysis. Based on statistical results, items were retained, revised, or discarded. Afterward, 15 students were invited to participate in individual interviews, which were audio-recorded and transcribed for analysis. All procedures adhered to the ethical standards required by educational institutions.

Data Analysis

Quantitative data was analyzed using IBM SPSS. Content validity indices were computed to determine the relevance and clarity of assessment items. Cronbach's alpha was used to measure reliability. Item difficulty and discrimination indices were computed following Classical Test Theory guidelines (Reynolds et al., 2021). A paired-samples t-test determined whether there was a significant difference between students' pretest and posttest scores.

Qualitative data were analyzed using thematic coding to identify patterns in student perceptions of the assessment items. The coding process followed the systematic approach outlined by Saldaña (2021), including initial coding, pattern coding, and theme development to support and explain the quantitative findings.

Ethical Considerations

The study complied with ethical standards for educational research. Approval was obtained from the school head and district supervisor before data collection. Informed consent was secured from all participants and their guardians. Students were assured of confidentiality, anonymity, and voluntary participation, and their responses were used exclusively for research purposes. Audio recordings and documents were stored securely and were destroyed after analysis.

Results and Discussion

Table 1 shows that most of the 70 PISA-based assessment items in Functions for Grade 11 Mathematics were rated highly by the experts. Thirty items were rated excellent, indicating full agreement on their relevance, while 37 items were rated good, showing strong expert consensus. Only 3 items (50, 57, and 62) were rated poor, suggesting they need minor revision or clarification to better align with the PISA framework. Overall, the results indicate that the assessment items are well-constructed and validated, supporting the reliability and practical use of the instrument. Linking these ratings to Bloom's Taxonomy shows that the items effectively target different cognitive levels—from remembering to creating—ensuring that most items promote higher-order thinking and provide appropriate challenges for learners (Mohammed & Omar, 2020).

Table 1. *Tabulation of the results on the Content Validity of the Contextualized PISA-based Assessment items in Functions for Grade 11 Mathematics*

Item Placement	No. of Exp. Giving 3 or 4	I-CVI	Pc	k*	Remarks	Item Placement	No. of Exp. Giving 3 or 4	I-CVI	Pc	k*	Remarks
Item_1	2	0.67	0.125	0.62	Good	Item_36	2	0.67	0.125	0.62	Good
Item_2	2	0.67	0.125	0.62	Good	Item_37	2	0.67	0.125	0.62	Good
Item_3	2	0.67	0.125	0.62	Good	Item_38	2	0.67	0.125	0.62	Good
Item_4	2	0.67	0.125	0.62	Good	Item_39	2	0.67	0.125	0.62	Good
Item_5	2	0.67	0.125	0.62	Good	Item_40	3	1.00	0.125	1.00	Excellent
Item_6	2	0.67	0.125	0.62	Good	Item_41	2	0.67	0.125	0.62	Good
Item_7	3	1.00	0.125	1.00	Excellent	Item_42	2	0.67	0.125	0.62	Good
Item_8	3	1.00	0.125	1.00	Excellent	Item_43	2	0.67	0.125	0.62	Good
Item_9	3	1.00	0.125	1.00	Excellent	Item_44	2	0.67	0.125	0.62	Good
Item_10	3	1.00	0.125	1.00	Excellent	Item_45	2	0.67	0.125	0.62	Good
Item_11	3	1.00	0.125	1.00	Excellent	Item_46	2	0.67	0.125	0.62	Good
Item_12	3	1.00	0.125	1.00	Excellent	Item_47	2	0.67	0.125	0.62	Good
Item_13	3	1.00	0.125	1.00	Excellent	Item_48	2	0.67	0.125	0.62	Good
Item_14	3	1.00	0.125	1.00	Excellent	Item_49	2	0.67	0.125	0.62	Good
Item_15	3	1.00	0.125	1.00	Excellent	Item_50	1	0.67	0.125	0.24	Poor
Item_16	3	1.00	0.125	1.00	Excellent	Item_51	2	0.67	0.125	0.62	Good
Item_17	3	1.00	0.125	1.00	Excellent	Item_52	2	0.67	0.125	0.62	Good
Item_18	3	1.00	0.125	1.00	Excellent	Item_53	2	0.67	0.125	0.62	Good
Item_19	3	1.00	0.125	1.00	Excellent	Item_54	2	0.67	0.125	0.62	Good
Item_20	3	1.00	0.125	1.00	Excellent	Item_55	2	0.67	0.125	0.62	Good
Item_21	3	1.00	0.125	1.00	Excellent	Item_56	2	0.67	0.125	0.62	Good
Item_22	2	0.67	0.125	0.62	Good	Item_57	1	0.67	0.125	0.24	Poor
Item_23	2	0.67	0.125	0.62	Good	Item_58	2	0.67	0.125	0.62	Good
Item_24	2	0.67	0.125	0.62	Good	Item_59	2	0.67	0.125	0.62	Good
Item_25	3	1.00	0.125	1.00	Excellent	Item_60	2	0.67	0.125	0.62	Good
Item_26	2	0.67	0.125	0.62	Good	Item_61	2	0.67	0.125	0.62	Good
Item_27	2	0.67	0.125	0.62	Good	Item_62	1	0.67	0.125	0.24	Poor
Item_28	3	1.00	0.125	1.00	Excellent	Item_63	2	0.67	0.125	0.62	Good
Item_29	3	1.00	0.125	1.00	Excellent	Item_64	2	0.67	0.125	0.62	Good
Item_30	3	1.00	0.125	1.00	Excellent	Item_65	3	1.00	0.125	1.00	Excellent
Item_31	3	1.00	0.125	1.00	Excellent	Item_66	3	1.00	0.125	1.00	Excellent
Item_32	3	1.00	0.125	1.00	Excellent	Item_67	3	1.00	0.125	1.00	Excellent
Item_33	3	1.00	0.125	1.00	Excellent	Item_68	3	1.00	0.125	1.00	Excellent
Item_34	3	1.00	0.125	1.00	Excellent	Item_69	2	0.67	0.125	0.62	Good
Item_35	3	1.00	0.125	1.00	Excellent	Item_70	3	1.00	0.125	1.00	Excellent

Kappa Evaluation. 0.75 and above (Excellent), 0.60 – 0.74(Good), 0.40-0.59(Fair), Below 0.40 (Poor)

Table 2. *List of Contextualized PISA-based Assessment Items in Grade 11 Mathematics for Pilot Testing*

Original Item Placement	New Item Placement	Remarks	Original Item Placement	New Item Placement	Remarks
Item_1		Revised	Item_36	Item_32	Revised
Item_2		Revised	Item_37	Item_33	Revised
Item_3		Revised	Item_38		Revised
Item_4	Item_1	Revised	Item_39		Revised
Item_5	Item_2	Revised	Item_40	Item_34	Accepted
Item_6	Item_3	Revised	Item_41	Item_35	Revised
Item_7	Item_4	Accepted	Item_42	Item_36	Revised
Item_8	Item_5	Accepted	Item_43	Item_37	Revised
Item_9	Item_6	Accepted	Item_44	Item_38	Revised
Item_10	Item_7	Accepted	Item_45	Item_39	Revised
Item_11	Item_8	Accepted	Item_46	Item_40	Revised

Item_12	Item_9	Accepted	Item_47	Item_41	Revised
Item_13	Item_10	Accepted	Item_48		Revised
Item_14	Item_11	Accepted	Item_49		Revised
Item_15	Item_12	Accepted	Item_50		Rejected
Item_16	Item_13	Accepted	Item_51	Item_42	Revised
Item_17	Item_14	Accepted	Item_52	Item_43	Revised
Item_18	Item_15	Accepted	Item_53		Revised
Item_19	Item_16	Accepted	Item_54		Revised
Item_20	Item_17	Accepted	Item_55		Revised
Item_21	Item_18	Accepted	Item_56		Revised
Item_22	Item_19	Revised	Item_57		Rejected
Item_23	Item_20	Revised	Item_58		Revised
Item_24	Item_21	Revised	Item_59		Revised
Item_25	Item_22	Accepted	Item_60		Revised
Item_26	Item_23	Revised	Item_61		Revised
Item_27		Revised	Item_62		Rejected
Item_28	Item_24	Accepted	Item_63	Item_44	Revised
Item_29	Item_25	Accepted	Item_64	Item_45	Revised
Item_30	Item_26	Accepted	Item_65	Item_46	Accepted
Item_31	Item_27	Accepted	Item_66	Item_47	Accepted
Item_32	Item_28	Accepted	Item_67	Item_48	Accepted
Item_33	Item_29	Accepted	Item_68	Item_49	Accepted
Item_34	Item_30	Accepted	Item_69		Revised
Item_35	Item_31	Accepted	Item_70	Item_50	Accepted

Table 2 shows the results of the pilot testing of 70 contextualized PISA-based assessment items in Grade 11 Mathematics. Items 1–6 were revised for clarity and fairness, while many others were retained but repositioned to improve test flow. Items 50, 57, and 62 were rejected for not meeting the required standards. Expert evaluations, using the Item Content Validity Index (I-CVI) and modified Kappa, rated 30 items as excellent, 37 as good, and 3 as poor. Revisions focused on clarity, fairness, and alignment with Bloom's Taxonomy and the PISA framework, ensuring the test is valid, relevant, and appropriately challenging. Overall, pilot testing and revisions enhanced the assessment's ability to measure students' mathematical thinking and real-world application.

Table 3. *Reliability of the Contextualized PISA-based Assessment Items Using Cronbach's Alpha*

No. of items	$\sum_{i=1}^{k} \sigma_y^2$	σ_x^2	Cronbach's Alpha	Remarks
50	7.443	57.840	0.889	High Reliability

Interpretation Scale. 1.0 (Perfect Reliability), 0.70-0.99 (High Reliability), 0.40 – 0.69(Moderate Reliability), 0.10 – 0.39(Low Reliability), 0 (No Reliability)

Table 3 shows that the 50-item contextualized PISA-based assessment instrument achieved a Cronbach's Alpha of 0.889, indicating high internal consistency and reliability. The sum of item variances (7.443) and total score variance (57.840) confirms that each item consistently contributes to the overall score. These results suggest that the assessment items are well-aligned with real-life contexts and the PISA framework, making the instrument reliable for measuring students' mathematical competencies. Guided by Constructivist Learning Theory, the items engage learners in meaningful, real-world scenarios, promoting active knowledge construction (McLeod, 2025). Classical Test Theory supports that the instrument minimizes measurement error and accurately assesses students' abilities (Thompson, 2023).

Table 4. *Item-correlation test results of the Contextualized PISA-based Assessment Items using Item Analysis*

Item Placement	Item-Correlation	Remarks	Item Placement	Item-Correlation	Remarks
Item_1	0.279	Acceptable	Item_26		No correlation
Item_2	0.086	Poor	Item_27	0.701	Good to Excellent
Item_3	0.693	Good to Excellent	Item_28	0.524	Good to Excellent
Item_4	0.674	Good to Excellent	Item_29	0.693	Good to Excellent
Item_5	0.250	Acceptable	Item_30	0.524	Good to Excellent
Item_6	0.426	Good to Excellent	Item_31	0.692	Good to Excellent
Item_7	0.444	Good to Excellent	Item_32	0.510	Good to Excellent
Item_8	0.157	Poor	Item_33	0.523	Good to Excellent
Item_9	0.389	Good to Excellent	Item_34	0.428	Good to Excellent
Item_10	0.558	Good to Excellent	Item_35		No correlation
Item_11	0.552	Good to Excellent	Item_36	0.009	Poor
Item_12	0.481	Good to Excellent	Item_37	-0.060	Very Problematic
Item_13	0.693	Good to Excellent	Item_38	0.057	Poor
Item_14	0.693	Good to Excellent	Item_39	0.093	Poor
Item_15	0.693	Good to Excellent	Item_40	0.233	Acceptable
Item_16	0.175	Poor	Item_41	0.125	Poor

Item_17	0.693	Good to Excellent	Item_42	0.340	Good to Excellent
Item_18	0.110	Poor	Item_43	0.411	Good to Excellent
Item_19	0.486	Good to Excellent	Item_44	0.572	Good to Excellent
Item_20	0.447	Good to Excellent	Item_45	-0.155	Acceptable
Item_21	0.585	Good to Excellent	Item_46	0.237	Acceptable
Item_22	0.654	Good to Excellent	Item_47	0.269	Acceptable
Item_23	0.363	Good to Excellent	Item_48	0.155	Poor
Item_24	0.395	Good to Excellent	Item_49	0.437	Good to Excellent
Item_25	0.701	Good to Excellent	Item_50	0.165	Poor

Interpretation Scale. 0.30 and above (Good to Excellent), 0.20 – 0.29(Acceptable), below 0.20(Poor), Negative (Very Problematic)

Table 4 shows the item-correlation analysis of the 50-item contextualized PISA-based assessment, indicating that most items performed well. Thirty-one items were rated "Good to Excellent," effectively distinguishing high- and low-performing students and aligning with learning objectives. Six items were "Acceptable," requiring minor revisions, while 10 items were rated "Poor," suggesting misalignment, insufficient challenge, or clarity issues that need to be reviewed. Item 26 and Item 35 showed no correlation because all participants answered correctly or incorrectly, preventing analysis, and Item 37 was "Very Problematic" due to a negative correlation, indicating structural or clarity problems. Overall, the results demonstrate that the majority of items are effective, though lower-performing items should be revised to improve the assessment of quality, fairness, and ability to measure students' mathematical competencies.

Table 5. *Difficulty index results of the Contextualized PISA-based Assessment items*

Item Placement	Df	Remarks	Item Placement	Df	Remarks
Item_1	0.80	Easy	Item_26	1.00	Very Easy
Item_2	0.93	Very Easy	Item_27	0.90	Easy
Item_3	0.95	Very Easy	Item_28	0.73	Average
Item_4	0.80	Easy	Item_29	0.60	Average
Item_5	0.88	Easy	Item_30	0.70	Average
Item_6	0.88	Easy	Item_31	0.73	Average
Item_7	0.83	Easy	Item_32	0.88	Easy
Item_8	0.88	Easy	Item_33	0.78	Easy
Item_9	0.85	Easy	Item_34	0.75	Average
Item_10	0.90	Easy	Item_35	0.00	Very Difficult
Item_11	0.88	Easy	Item_36	0.18	Difficult
Item_12	0.78	Easy	Item_37	0.73	Average
Item_13	0.95	Very Easy	Item_38	0.63	Average
Item_14	0.95	Very Easy	Item_39	0.78	Easy
Item_15	0.95	Very Easy	Item_40	0.53	Average
Item_16	0.93	Very Easy	Item_41	0.38	Average
Item_17	0.95	Very Easy	Item_42	0.75	Average
Item_18	0.83	Easy	Item_43	0.58	Average
Item_19	0.93	Very Easy	Item_44	0.68	Average
Item_20	0.88	Easy	Item_45	0.23	Difficult
Item_21	0.85	Easy	Item_46	0.65	Average
Item_22	0.85	Easy	Item_47	0.58	Average
Item_23	0.78	Easy	Item_48	0.63	Average
Item_24	0.88	Easy	Item_49	0.30	Average
Item_25	0.88	Easy	Item_50	0.65	Average

Scale. 0.91-0.95 (Very Easy), 0.76- 0.90(Easy), 0.25-0.75(Average), 0.10 – 0.24(Difficult), 0.09 and below (Very Difficult)

Table 5 shows the difficulty index of the 50-item contextualized PISA-based assessment, indicating that most items were Easy or Very Easy, with difficulty factors (Df) of 0.76–0.95, suggesting that the questions were clear, understandable, and well-aligned with the PISA framework. Several items fell in the Average range (Df 0.25–0.75), providing a moderate level of challenge and helping differentiate student performance. A few items—such as Item 35 (Df 0.00), Item 36 (Df 0.18), and Item 45 (Df 0.23)—were classified as Difficult or Very Difficult, indicating they may be too challenging or misaligned with the intended objectives. Overall, the instrument demonstrates strong potential, with most items performing effectively, while a small number may require revision to improve balance and better assess the targeted learning competencies.

Table 6 presents the discrimination index (Dx) results of the 50-item contextualized PISA-based assessment in Functions for Grade 11 Mathematics. Thirteen items were rated "Very Good" (Dx ≥ 0.40), effectively differentiating high- and low-performing students and suitable for retention. Five items were reasonably good (Dx 0.30–0.39) and generally acceptable. Eight items were marginal (Dx 0.20–0.29) and need improvement, while 24 items were poor (Dx ≤ 0.19), indicating weak discrimination and requiring major revision or removal. Overall item analysis across item-total correlation, difficulty index, and discrimination index showed that most items were clear, answerable, and measured student performance effectively, but several items were problematic, too easy or difficult, or weak in discriminating performance. These findings suggest that while many items are strong, refinement is necessary to ensure the assessment

is balanced, fair, and accurately measures intended learning outcomes, identifying which items are suitable for final use.

Table 6. *Discrimination index results of the Contextualized PISA-based Assessment items*

Item Placement	Dx	Remarks	Item Placement	Dx	Remarks
Item_1	0.30	Reasonably Good	Item_26	0.00	Poor Item
Item_2	0.05	Poor Item	Item_27	0.00	Poor Item
Item_3	0.10	Poor Item	Item_28	0.55	Very Good Item
Item_4	0.40	Very Good Item	Item_29	0.70	Very Good Item
Item_5	0.05	Poor Item	Item_30	0.50	Very Good Item
Item_6	0.05	Poor Item	Item_31	0.45	Very Good Item
Item_7	0.15	Poor Item	Item_32	0.25	Marginal Item
Item_8	-0.05	Poor Item	Item_33	0.35	Reasonably Good
Item_9	0.00	Poor Item	Item_34	0.50	Very Good Item
Item_10	0.20	Marginal Item	Item_35	0.00	Poor Item
Item_11	0.15	Poor Item	Item_36	-0.05	Poor Item
Item_12	0.35	Reasonably Good	Item_37	0.05	Poor Item
Item_13	0.10	Poor Item	Item_38	-0.05	Poor Item
Item_14	0.10	Poor Item	Item_39	0.25	Marginal Item
Item_15	0.10	Poor Item	Item_40	0.45	Very Good Item
Item_16	-0.05	Poor Item	Item_41	0.15	Poor Item
Item_17	0.10	Poor Item	Item_42	0.50	Very Good Item
Item_18	0.05	Poor Item	Item_43	0.25	Marginal Item
Item_19	0.05	Poor Item	Item_44	0.45	Very Good Item
Item_20	0.25	Marginal Item	Item_45	-0.15	Poor Item
Item_21	0.30	Reasonably Good	Item_46	0.40	Very Good Item
Item_22	0.30	Reasonably Good	Item_47	0.15	Poor Item
Item_23	0.25	Marginal Item	Item_48	0.45	Very Good Item
Item_24	0.25	Marginal Item	Item_49	0.50	Very Good Item
Item_25	0.25	Marginal Item	Item_50	0.50	Very Good Item

Scale. 0.40 and above (Very Good Item), 0.30 – 0.39 (Reasonably Good – Usually Retained 0.20 – 0.29(Marginal- Subject for Improvement), 0.19 and below (Poor – To be revised or discarded)

Table 7. *Results of Item Decision Matrix Based on Discrimination and Difficulty Indices on Contextualized PISA-based Assessment items in Functions for Grade 11 Mathematics*

Item Placement	Dx	Df	Remarks	Item Placement	Dx	Df	Remarks
Item_1	Reasonably Good	Easy	Revise	Item_26	Poor Item	Very Easy	Reject
Item_2	Poor Item	Very Easy	Reject	Item_27	Poor Item	Easy	Reject
Item_3	Poor Item	Very Easy	Reject	Item_28	Very Good Item	Average	Retain
Item_4	Very Good Item	Easy	Revise	Item_29	Very Good Item	Average	Retain
Item_5	Poor Item	Easy	Reject	Item_30	Very Good Item	Average	Retain
Item_6	Poor Item	Easy	Reject	Item_31	Very Good Item	Average	Retain
Item_7	Poor Item	Easy	Reject	Item_32	Marginal Item	Easy	Revise
Item_8	Poor Item	Easy	Reject	Item_33	Reasonably Good	Easy	Revise
Item_9	Poor Item	Easy	Reject	Item_34	Very Good Item	Average	Retain
Item_10	Marginal Item	Easy	Revise	Item_35	Poor Item	Very Difficult	Reject
Item_11	Poor Item	Easy	Reject	Item_36	Poor Item	Difficult	Reject
Item_12	Reasonably Good	Easy	Revise	Item_37	Poor Item	Average	Reject
Item_13	Poor Item	Very Easy	Reject	Item_38	Poor Item	Average	Reject
Item_14	Poor Item	Very Easy	Reject	Item_39	Marginal Item	Easy	Revise
Item_15	Poor Item	Very Easy	Reject	Item_40	Very Good Item	Average	Retain
Item_16	Poor Item	Very Easy	Reject	Item_41	Poor Item	Average	Reject
Item_17	Poor Item	Very Easy	Reject	Item_42	Very Good Item	Average	Retain
Item_18	Poor Item	Easy	Reject	Item_43	Marginal Item	Average	Revise
Item_19	Poor Item	Very	Reject	Item_44	Very Good	Average	Retain

		Easy				Item		
Item_20	Marginal Item	Easy	Revise	Item_45	Poor Item	Difficult	Reject
Item_21	Reasonably Good	Easy	Revise	Item_46	Very Good Item	Average	Retain
Item_22	Reasonably Good	Easy	Revise	Item_47	Poor Item	Average	Reject
Item_23	Marginal Item	Easy	Revise	Item_48	Very Good Item	Average	Retain
Item_24	Marginal Item	Easy	Revise	Item_49	Very Good Item	Average	Retain
Item_25	Marginal Item	Easy	Revise	Item_50	Very Good Item	Average	Retain

Scale. 0.40 and above (Very Good Item), 0.30 – 0.39 (Reasonably Good – Usually Retained 0.20 – 0.29(Marginal- Subject for Improvement), 0.19 and below (Poor – To be revised or discarded)

Discrimination Index / Difficulty Index	Very Easy (0.91–0.95)	Easy (0.76–0.90)	Average (0.25–0.75)	Difficult (0.10–0.24)	Very Difficult (≤ 0.09)
Very Good (≥ 0.40)	Revise	Revise	Retain	Revise	Reject
Reasonably Good (0.30–0.39)	Revise	Revise	Retain	Revise	Reject
Marginal (0.20–0.29)	Revise	Revise	Revise	Revise	Reject
Poor (≤ 0.19)	Reject	Reject	Reject	Reject	Reject

Table 7 presents the item decision matrix for the 50-item contextualized PISA-based assessment in Functions for Grade 11 Mathematics, showing a range of item quality and effectiveness. Several items (28, 29, 30, 31, 34, 40, 42, 44, 46, 48, 49, 50) rated "Very Good" in discrimination and "Average" in difficulty performed well and are recommended for retention, as they effectively differentiate students' performance while maintaining balanced challenge. Many items, particularly those with poor discrimination or extreme difficulty levels (2, 3, 5–9, 11, 13–19, 26, 27, 35–38, 41, 45, 47), are recommended for rejection, as they do not meaningfully contribute to the assessment's reliability. Items rated Marginal or Reasonably Good but paired with Easy or Average difficulty (1, 4, 10, 12, 20–25, 32, 33, 39, 43) are suggested for revision to improve clarity, fairness, and effectiveness. Overall, while several items are well-constructed, targeted revisions and removals are necessary to enhance the tool's ability to measure students' mathematical understanding accurately. These findings align with Situated Learning Theory, highlighting that assessments should engage students in real-life, cognitively demanding tasks that differentiate performance and support meaningful learning through experience.

Table 8. *Suitable Contextualized PISA-based Assessment Items in Functions for final use*

Item Placement	Pilot Testing Item Placement	Original Item Placement	Item Placement	Pilot Testing Item Placement	Original Item Placement
Item_1	Item_1	Item_4	Item_16	Item_31	Item_35
Item_2	Item_4	Item_7	Item_17	Item_32	Item_36
Item_3	Item_6	Item_9	Item_18	Item_33	Item_37
Item_4	Item_9	Item_12	Item_19	Item_34	Item_40
Item_5	Item_10	Item_13	Item_20	Item_39	Item_45
Item_6	Item_12	Item_15	Item_21	Item_40	Item_46
Item_7	Item_19	Item_22	Item_22	Item_42	Item_51
Item_8	Item_20	Item_23	Item_23	Item_43	Item_52
Item_9	Item_21	Item_24	Item_24	Item_44	Item_63
Item_10	Item_22	Item_25	Item_25	Item_45	Item_64
Item_11	Item_24	Item_28	Item_26	Item_46	Item_65
Item_12	Item_25	Item_29	Item_27	Item_47	Item_66
Item_13	Item_28	Item_32	Item_28	Item_48	Item_67
Item_14	Item_29	Item_33	Item_29	Item_49	Item_68
Item_15	Item_30	Item_34	Item_30	Item_50	Item_70

Table 8 shows the finalized Contextualized PISA-based assessment items in Functions for Grade 11 Mathematics, highlighting items selected for final use after careful review and validation. Each item corresponds to its original version, with references to initial item numbers to ensure transparency and traceability in development. This mapping confirms that only items meeting the criteria for clarity, relevance, cognitive demand, and alignment with the PISA framework were retained. Overall, the table reflects a rigorous selection process, resulting in a set of assessment items that are contextually appropriate, valid, and capable of effectively evaluating students' understanding of functions.

Table 9. *Students' achievement based on the Pretest and Posttest Results of the Contextualized PISA-based Assessment Items in Functions*

Test	Sample Size	Mean	Minimum Score	Maximum Score	Std. Deviation	Mean Percentage Score (MPS)	Level of Achievement
Pretest	72	11.70	4	25	4.15	23.39%	Not Proficient
Posttest	72	15.72	6	30	6.45	31.03%	Low Proficient

MPS Scale. 0% – 24% = Not Proficient; 25% – 49% = Low Proficient; 50% – 74% = Nearly Proficient 75% – 89% = Proficient; 90% – 100% = Highly Proficient

Table 9 presents the students' achievement on the pretest and posttest of the Contextualized PISA-based assessment items in Functions for Grade 11 Mathematics. In the pretest, students had a weighted mean of 11.70 (MPS = 23.39%), categorized as Not Proficient, while in the posttest, their weighted mean increased to 15.72 (MPS = 31.03%), falling under Low Proficient. Although the posttest showed improvement in scores, the overall achievement level remained low, indicating minimal gains in understanding Functions. These results suggest that students need more sustained, targeted interventions such as differentiated instruction, scaffolded practice, and contextualized problem-solving activities aligned with PISA frameworks to strengthen conceptual understanding and performance. The findings are consistent with previous studies showing that Grade 11 learners often struggle with mastery in mathematics (Mamolo & Sugano, 2021) but can benefit from adaptive, learner-centered strategies (Fernandez & Tangalin, 2020).

Table 10. *Paired-Samples t-test on the significant difference on students' achievement per learning competency between the pretest and posttest results of the Contextualized PISA-based Assessment*

Paired Differences	Mean	Std. Deviation	Df	t- value	p- value	Decision	Results
Pretest and Post Test	11.70 15.72	5.07	71	-6.736	0.001	Reject Null Hypothesis	Significant Difference
Paired Differences	Mean	Std. Deviation	Df	t- value	p- value	Decision	Results

Table 10 presents the paired-samples t-test comparing pretest and posttest scores of students on the Contextualized PISA-based assessment in Functions. The mean difference of -4.03 (SD = 5.07, SE = 0.60) indicates that posttest scores were higher than pretest scores. The computed t-value of -6.736 with 71 degrees of freedom and a p-value of 0.001 (< 0.05) shows a statistically significant improvement, leading to the rejection of the null hypothesis. This demonstrates that the intervention had a positive effect on students' achievement, even though the overall performance remained within the Low proficiency level.

These findings align with previous studies (Fernandez & Tangalin, 2020; Project STAIR) showing that instruction and targeted interventions can improve student performance, but mastery may still be limited. The results highlight the need for scaffolded, differentiated, and continuous instructional strategies, coupled with frequent formative assessments and feedback, to help students progress toward higher levels of competency in Functions.

Feedback from Students on Contextualized PISA-Based Assessment Items

The students' feedback revealed four main themes regarding their experiences with the Contextualized PISA-based assessment items in Functions: Ease and Manageability, Moderate and Balanced Difficulty, High Difficulty and Challenge, and Mixed and Varied Perceptions.

Theme 1: Ease and Manageability. Some students found the test easy or only slightly challenging, indicating alignment with their prior knowledge and confidence in answering the items. This theme included subcategories of Expressing Ease in Answering and Downplaying Difficulty, in which learners reported that tasks were either comfortable or only mildly challenging. Their responses suggest the assessment was fair and manageable, supporting Vygotsky's Zone of Proximal Development and Bandura's Self-Efficacy Theory, as students felt capable and motivated to engage with the tasks.

Theme 2: Moderate and Balanced Difficulty. A few students found the test "just right," indicating a fair and reasonable level of challenge. Subcategories Perceiving Balance in Difficulty and Viewing Assessment as Moderate indicated that items encouraged critical thinking without causing frustration. This aligns with Piaget's concept of equilibration and Csikszentmihalyi's Flow Theory, suggesting that a balanced level of difficulty promotes engagement and persistence.

Theme 3: High Difficulty and Challenge. Several students admitted struggling with the test, particularly with complex items that exceeded their current mastery of Functions. Subcategories Experiencing Strong Test Difficulty and Balancing Difficulty with Manageability showed that, while some items were challenging, they remained within reach for motivated learners. This highlights the need for scaffolding and targeted interventions to support higher-order problem-solving skills.

Theme 4: Mixed and Varied Perceptions. Some students reported a combination of experiences, noting that while several questions were manageable, others were confusing or more demanding. Subcategories Expressing Mixed Difficulty, Acknowledging Item Variation, and Recognizing Minor Confusion demonstrate that students navigated different levels of cognitive demand within the same assessment.

The findings indicate that students' experiences ranged from manageable to highly challenging, reflecting the diversity in readiness, prior knowledge, and problem-solving skills. These results underscore the importance of designing assessments that are contextually relevant, cognitively balanced, and supportive of varied learning needs, aligning with Bronfenbrenner's ecological model of human development and international research on factors influencing mathematics achievement (Wang, 2023; Akay & Oskonbaeva, 2024). Targeted instructional support and scaffolding are recommended to help students effectively engage with complex PISA-based tasks and improve their conceptual understanding in Functions.

Conclusions

Based on the findings of this study, it is concluded that the developed Contextualized PISA-based assessment items in Functions for

Eloiza Jane O. Serdeña

Grade 11 Mathematics are valid, reliable, and effective tools for evaluating students' mathematical competencies in real-world contexts. Most items were rated excellent or good in content validity, demonstrating strong alignment with learning outcomes, Bloom's Taxonomy, and the PISA framework, while only a few items required revision or were rejected. The instrument showed high reliability, with a Cronbach's Alpha of 0.889, confirming strong internal consistency and dependability in assessing student performance. Item analyses revealed that 12 items were suitable for final use, exhibiting balanced difficulty, clear wording, and strong discriminating power, while other items were revised or discarded to ensure fairness, clarity, and effectiveness. Students' achievement in the pretest and posttest indicated slight improvement, with a statistically significant difference between the two, yet overall mastery remained at a Low Proficiency level, highlighting the need for sustained, scaffolded, and differentiated instructional strategies to enhance understanding and application of Functions. Qualitative feedback from students revealed varied perceptions of item difficulty, ranging from ease and manageability to high challenge, emphasizing the importance of targeted support, balanced assessment design, and the incorporation of contextualized, real-world problems to foster engagement and critical thinking. Overall, the study concludes that the Contextualized PISA-based assessment items are suitable for final implementation as an assessment tool, but their full potential in improving student mastery relies on continuous instructional interventions and contextually responsive teaching practices.

References

Akay, E. Ç., & Oskonbaeva, Z. (2024). Investigating the factors affecting the PISA-based test performance of Asian students. Economic Journal of Emerging Markets, 16(1), 38–49. https://doi.org/10.20885/ejem.vol16.iss1.art4

Akhter, N., & Akhter, N. (2018). Learning in mathematics: Difficulties and perceptions of students. Journal of Educational Research, 21(1), 147–162. https://jer.iub.edu.pk/journals/JER-Vol-21.No-1/11.pdf

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. New York, NY: Longman.

Andrade, H. L., & Brookhart, S. M. (2020). Classroom assessment and grading to assure learning. New York, NY: Routledge. https://doi.org/10.4324/9780429440646

Bandura, A. (1997). Self-efficacy: The exercise of control. New York, NY: W. H. Freeman

Barba, A. T. P., Valeria, Z. M. A., Caserial, I. J. L., & Gaylo, D. (2024, March). Alignment of General Mathematics curriculum to the PISA 2022 Mathematics Framework. Bukidnon State University. https://www.researchgate.net/publication/385353147_Alignment_of_General_Mathematics_Curriculum_to_the_PISA_2022_Mathematics_Framework

Bart, D. (2022). Research discourse in the programme for international student assessment: A critical perspective. European Educational Research Journal, 23(1), 145-162. https://doi.org/10.1177/14749041221127758 (Original work published 2024)

Bernardo, A. B. I. (2020). Socioeconomic status moderates the relationship between growth mindset and learning in mathematics and science: Evidence from PISA 2018 Philippine data. Asia-Pacific Education Researcher, 29(3), 208–222. https://doi.org/10.1080/21683603.2020.1832635

Besar, P. H. S. N. b. P. H. B. (2018). Situated learning theory: The key to effective classroom teaching? HONAI: International Journal for Educational, Social, Political & Cultural Studies, 1(1), 49 60. https://www.journals.mindamas.com/index.php/honai/article/view/1022

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa[3](https://psycnet.apa.org/record/2006-06991-https://academic-writing.uk/braun-clarke-thematic-analysis/)

Borgonovi, F., Ferrara, A., & Schleicher, A. (2020). Education responses to COVID-19: Implementing a way forward. OECD Education Working Papers, No. 253. https://doi.org/10.1787/6f8d19e0-en

Branch, R. M. (2009). Instructional design: The ADDIE approach. New York, NY: Springer.https://doi.org/10.1007/978-0-387-09506-6.

Culatta, R. (2020). ADDIE model. InstructionalDesign.org. Retrieved from https://www.instructionaldesign.org/models/addie/.

Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. New York, NY: Harper & Row.

DeBell, A. (2020, January 6). What is the ADDIE model of instructional design? Water Bear Learning. Retrieved from https://waterbearlearning.com/addie-model instructional-design/.

Fernandez, L. S., & Tangalin, I. A. (2020). Effects of differentiated instruction on the Grade 11 students' academic performance in mathematics. International Journal of Advanced Research in Engineering and Technology (IJARET), 11(9), 207–215. https://doi.org/10.34218/IJARET.11.9.2020.021

Grek, S 2015, 'Governing by Numbers: The PISA 'Effect' in Europe', Journal of Education Policy, vol. 24, doi.org/10.1080/02680930802412669

Herrera, C. D. A., Jr., & Dio, R. V. (2016). Extent of readiness of Grade 10 students for General Mathematics of Senior High School in Sorsogon City, Philippines. Asia Pacific Journal of Education, Arts and Sciences, 3(4), 1–8. https://www.apjeas.apjmr.com/wp-content/uploads/2016/10/APJEAS-2016.3.4.01.pdf

Imasuen, K. (2022). Sample size determination in test-retest and Cronbach alpha reliability estimates. British Journal of Contemporary Education, 2(1), 17–29. https://doi.org/10.52589/BJCE-FY266HK9

Jaudinez, A. S. (2019). Teaching Senior High School Mathematics: Problems and Interventions. Pedagogical Research, 4(2), em0031. https://doi.org/10.29333/pr/5779

Kane, M. T. (2016). Validating the interpretations and uses of test scores. Journal of Educational Measurement, 53(1), 1–14. https://doi.org/10.1111/jedm.12117

Kholid, M. N., & Nissa, M. (2022). Students' math literacy in solving PISA-like problems in Papuan local context. Al-Ishlah: Jurnal Pendidikan, https://journal.staihubbulwathan.id/index.php/alishlah/article/view/1234

Lewis, S. (2017). Governing schooling through 'what works': The OECD's PISA for Schools. Journal of Education Policy, 32(3), 281–302. https://doi.org/10.1080/02680939.2016.1252855

Lewis, S., & Lingard, B. (2015). The multiple effects of international large-scale assessment on education policy and research. Discourse: Studies in the Cultural Politics of Education, 36(5), 621–637. https://doi.org/10.1080/01596306.2015.1034460

Mamolo, L. A., & Sugano, S. G. C. (2021). Development of an achievement test to measure students' competency in General Mathematics. Anatolian Journal of Education, 6(1), 79–90. https://doi.org/10.29333/aje.2021.61.79-90.

McLeod, S. (2025). Constructivism as a theory for teaching and learning. Simply Psychology. Retrieved from https://www.simplypsychology.org/constructivism.html.

Metsämuuronen, J. (2020). Generalized discrimination index. International Journal of Educational Methodology, 6(2), 237–257. https://doi.org/10.12973/ijem.6.2.237

Mitani, H. (2021). Test score gaps in higher order thinking skills: Exploring instructional practices to improve the skills and narrow the gaps. AERA Open, 7(1), 1–23. https://doi.org/10.1177/23328584211016470

Mitari, O., & Zulkardi, Z. (2019). PISA-like problems on students' mathematical literacy using the context of Jakabaring Sport City. Journal of Physics: Conference Series, 1315(1), 012014. https://doi.org/10.1088/1742-6596/1315/1/012014

Mohammed, M., & Omar, N. (2020). Question classification based on Bloom's taxonomy cognitive domain using modified TF-IDF and Word2Vec. PLoS One, 15(3), e0230442. https://doi.org/10.1371/journal.pone.0230442

Morrison, G. R., Ross, S. M., Kemp, J. E., & Kalman, H. (2010). Designing effective instruction (6th ed.). Hoboken, NJ: John Wiley & Sons.

OECD. (2019). PISA 2018 assessment and analytical framework. Paris: OECD Publishing. https://doi.org/10.1787/b25efab8-en

Padilla, C. L., & Songco, M. T. (2016). Contextualizing mathematics education in the Philippines: Impact on student learning and engagement. Philippine Journal of Education Research, 3(1), 45–58. https://ejournals.ph/article.php?id=10456

Piaget, J. (1970). The science of education and the psychology of the child. New York, NY: Viking

Piaget, J. (1972). The psychology of the child. New York, NY: Basic Books.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Research in Nursing & Health, 29(5), 489–497. https://doi.org/10.1002/nur.20147

Phan, H. P. (2010). Students' academic performance and various cognitive processes of learning: An integrative framework and empirical analysis. Educational Psychology, 30(3), 297–322. https://doi.org/10.1080/01443410903573297

Reynolds, C. R., Altmann, R. A., & Allen, D. N. (2021). Item analysis: Methods for fitting the right items to the right test. In C. R. Reynolds, R. A. Altmann, & D. N. Allen (Eds.), Mastering modern psychological testing: Theory and methods (pp. 263–289). Springer. https://doi.org/10.1007/978-3-030-59455-8_7

Salkind, N. J. (2017). Tests & measurement for people who (think they) hate tests & measurement (3rd ed.). SAGE Publications, Inc. https://www.amazon.com/Tests-Measurement-People-Think-They/dp/1506368387

Schleicher, A. (2017). Can competencies assessed by PISA be considered the fundamental school knowledge 15-year-olds should

possess? Journal of Educational Change, 8(4), 349–357. https://doi.org/10.1007/s10833-007-9042-x

Schleicher, A. (2020). The power and promise of the OECD's PISA. International Summit on the Teaching Profession. OECD Publishing. https://www.oecd.org/education/the-power-and-promise-of-the-oecd-pisa-9789264454904-en.htm

Shabatura, J. (2013). Using Bloom's taxonomy to write effective learning objectives. University of Arkansas, USA, https://tips. uark. edu/using-blooms-taxonomy/,(Zadnji pristup 12.02. 2021.)[7] C. Ou, DA Joyner i AK Goel,(2019.), Designing and developing video lessons for online learning: A seven-principle model, Online Learning, 23(2), 89-90.

Shafa, S., Zulkardi, & Putri, R. I. I. (2023). Students' creative thinking skills in solving PISA-like mathematics problems related to quantity content. Jurnal Elemen, 9(1), 271-282. https://doi.org/10.29408/jel.v9i1.6975

Schunk, D. H. (2020). Learning theories: An educational perspective (8th ed.). New York, NY: Pearson.

Thompson, N. (2023). Classical Test Theory vs. Item Response Theory. Assessment Systems. Retrieved from https://assess.com/classical-test-theory-vs-item-response-theory/

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press

Wang, X. S., Perry, L. B., Malpique, A., & Ide, T. (2023). Factors predicting mathematics achievement in PISA: A systematic review. Large-scale Assessments in Education, 11(24), 1–42. https://doi.org/10.1186/s40536-023-00174-8

Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. Western Journal of Nursing Research, 25(5), 508–518. https://doi.org/10.1177/0193945903252998

Affiliations and Corresponding Information

Eloiza Jane O. Serdeña
Saint Columban College – Philippines