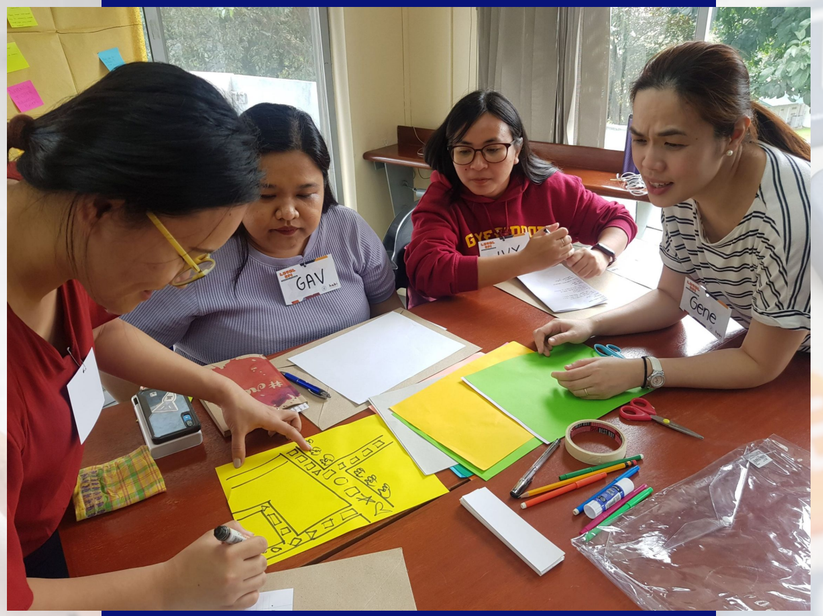
ATENEO DE MANILA UNIVERSITY

Leveraging International Large-Scale Assessments:

Insights from an Item-Writing Professional Development Program for High School Mathematics Teachers



Mark Lester B. Garcia, Ph. D. (Cand.)
Catherine P. Vistro-Yu, Ed. D.
Ateneo de Manila University

NCEME 2024, De La Salle University – Manila
August 31, 2024

OUTLINE OF PRESENTATION:

ATENEO DE MANILA UNIVERSITY

1. Introduction
2. Related Literature
3. Methodology
4. Results and Discussion
5. Conclusion

INTRODUCTION

ATENEO DE MANILA UNIVERSITY

Background

Alderson & Wall (1993, as cited in Cheng, 2000):

- Washback effect: "*Influence of testing on teaching and learning, which is rooted in the notion that tests should and could drive teaching and hence learning*"
- What and how teachers teach, what and how learners learn; the rate, sequence, degree, and depth of teaching and learning usually receive washback

Vrikki et al. (2024): Developed a comprehensive model of effective teaching based on PISA 2018 results

INTRODUCTION

ATENEO DE MANILA UNIVERSITY

Rationale

- This study is part of the primary author's dissertation project about the item-writing behaviors of high school mathematics teachers

- It is crucial to understand teachers' assessment knowledge and practices (Leong, 2015) given that they spend a significant amount of time and effort in assessment-related activities (MacBeath & Galton, 2004, as cited in Leong, 2015)

INTRODUCTION

ATENEO DE MANILA UNIVERSITY

Rationale

- PH commits to continued participation in ILSAs as part of its efforts to measure quality of education among learners and benchmark with other education systems
- Materials were developed for teachers and students in the hopes of increased familiarity with PISA, as part of PH's field preparations for PISA 2022 (Senate of the Philippines, 2024)
- PISA items meant to measure students' abilities to solve problems related to real life (Sjøberg & Jenkins, 2022)

INTRODUCTION

ATENEO DE MANILA UNIVERSITY

Conceptual Framework

- Teacher Professional Development (TPD): programs and strategies designed to change the beliefs and practices of teachers in order to improve the achievement of their students (Guskey, 2002, as cited in Lauer et al., 2014)

Research Question:

- What is the quality of PISA-like math items created by secondary mathematics teachers?

RELATED LITERATURE

ATENEO DE MANILA UNIVERSITY

Studies related to TPD

- Positive effect of TPD: “incentives such as promotion or salary implications, having a specific subject focus, incorporating lesson enactment in the training, and including initial face-to-face training—are positively associated with student test score gains” (Popova et al., 2022)

- TPD on Assessment for Learning for math teachers in the Netherlands resulted to significantly higher student outcomes for students of Grades 7-9 (5 TPD sessions over a year) (de Vries et al., 2022)

RELATED LITERATURE

ATENEO DE MANILA UNIVERSITY

Studies related to TPD

- Lauer et al. (2014): it is possible for short-term PD to improve participant outcomes, PD content and method are as important as duration

- Kyaruzi et al. (2020): 1 day intensive PD on analyzing student errors and developing strategies to prevent errors in math had a positive effect on student perceptions of their math teacher's behavior

RELATED LITERATURE

ATENEO DE MANILA UNIVERSITY

Studies related to Item-Writing Training

- OECD (2016a): Item writers are trained in the PISA framework and develop items with real-world contexts as part of PISA-D Capacity Building Plan in Zambia
- OECD (2016b): Item development process based on PISA frameworks as part of technical capacity building in Cambodia
- Kurniasih et al. (2023): Proposed a teacher-training model for writing PISA-like reading test items (Indonesian context)

RELATED LITERATURE

ATENEO DE MANILA UNIVERSITY

Studies related to Item-Writing Training

- Yurdakul et al. (2020): PD program designed to upskill math teachers in writing items concerning higher-order thinking skills similar to those in PISA and TIMSS; 100 secondary math teachers in Turkey, 4 three-hour sessions over 2 days

METHODOLOGY

ATENEO DE MANILA UNIVERSITY

Research Design

- Concurrent mixed-methods study (Descriptive quantitative study and exploratory qualitative study)

Research Sample

- Nine high school mathematics teachers from the junior high school unit of a Catholic university in Mindanao

METHODOLOGY

ATENEO DE MANILA UNIVERSITY

Data Collection

- Teachers underwent a three-day item-writing PD as part of their summer in-service training for SY 2022-23
- First day was allotted for a seminar about PISA mathematics assessment framework, and an orientation about the study
- Succeeding days were for the item-writing workshop, where they were observed via screencast videography (Garcia & Hao, 2023)

METHODOLOGY

ATENEO DE MANILA UNIVERSITY

Data Collection

- Each teacher was assigned to write 8 PISA-like math items each

Table of Specifications (TOS #1):

Content Category \ Process Category	Item Format	Reasoning (25%)	Formulate (25%)	Employ (25%)	Interpret (25%)	Subtotal	Total No. of Items
Quantity (25%)	MC		Q1			1	2
	CR			Q2		1	
Change & Relationships (25%)	MC	Q3				1	2
	CR				Q4	1	
Space & Shape (25%)	MC	Q5				1	2
	CR		Q6			1	
Uncertainty & Data (25%)	MC			Q7		1	2
	CR				Q8	1	
Total No. of Items		2	2	2	2		8

Table of Specifications (TOS #2):

Content Category \ Context Category	Personal (25%)	Occupational (25%)	Societal (25%)	Scientific (25%)	Total
Quantity (25%)	Q1	Q2			2
Change & Relationships (25%)			Q4	Q3	2
Space & Shape (25%)	Q5	Q6			2
Uncertainty & Data (25%)			Q7	Q8	2
Total	2	2	2	2	8

Summary of Item Specifications
(Please refer to the legend below)

Item	Item Format	Content Category	Context Category	Process Category
Q1	MC	QNT	PER	PSF
Q2	CR	QNT	OCC	PSE
Q3	MC	C&R	SCI	RSN
Q4	CR	C&R	SOC	PSI
Q5	MC	S&S	PER	RSN
Q6	CR	S&S	OCC	PSF
Q7	MC	U&D	SOC	PSE
Q8	CR	U&D	SCI	PSI

METHODOLOGY

ATENEO DE MANILA UNIVERSITY

Data Collection

- All items were collected and forwarded to expert evaluators (Instrument: Expert Evaluation Form)

EVALUATION FORM FOR EXPERT EVALUATORS

PART A:
Please rate each test item of the participant based on its alignment with PISA's Content Category, Process Category and Context Category. Please use the 6-point Likert scale shown below.

NOTE: Comments and feedback must be based on the Summary of Item Specifications. Thank you!

5 – Perfectly Aligned
4 – Moderately Aligned
3 – Fairly Aligned
2 – Somewhat Aligned
1 – Poorly aligned
0 – Not aligned at all

Item	Rating based on the alignment with **CONTENT Categories**	Rating based on the alignment with **CONTEXT Categories**	Rating based on the alignment with **PROCESS Categories**
Q1	5	5	3
Q2	5	5	5
Q3	5	5	5
Q4	5	5	5
Q5	5	4	5
Q6	5	5	3
Q7	5	4	5
Q8	5	4	3

PART B:
Please rate each test item of the participant based on the appropriateness of the use of context. Base your scores on the level of agreement with the following statements. Please use the 6-point Likert scale shown below.

5 – I strongly agree with this statement.
4 – I agree with this statement.
3 – I slightly agree with this statement.
2 – I slightly disagree with this statement.
1 – I disagree with this statement.
0 – I strongly disagree with this statement.

Statements	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
The context used in this item is stimulating and interesting for 15-year-old Filipino students.	5	5	5	5	5	5	5	3
The context used in this item is realistic as it accurately reflects daily life realities and real-world problems.	5	5	5	5	5	4	5	5
The problem arising from the context is plausible and believable.	5	5	5	5	5	4	5	5
The context used in this item is compatible with the measured content.	5	5	5	5	5	5	5	5
The context used in this item is equally accessible regardless of gender, religion, culture or living conditions in the Philippines.	3	3	5	5	4	5	5	4
The context used in this item has minimal vocabulary with unnecessary linguistic complexity.	5	4	5	5	5	5	5	4

METHODOLOGY

ATENEO DE MANILA UNIVERSITY

Data Collection

Statements (continuation)	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
The context used in this item has a minimal amount of irrelevant information.	5	4	5	4	5	5	5	5
The context used in this item is socially and culturally sensitive- that is, it is free from offensive or controversial references.	5	5	5	5	5	5	5	5
The context used in this item shows language accessibility in the sense that the intentions of the task and the expectations of the test writer are communicated clearly.	5	5	5	5	4	5	5	5
The context used in this item shows conservative novelty- that is, a novelty that gives an item just enough unpredictability, giving the test-takers a "different but familiar" feeling which is neither overwhelming nor alienating.	4	5	5	5	5	4	5	4

PART C:
Please provide constructive qualitative feedback for the item writer to make the item: (a) become more PISA-like and (b) become more aligned* with the content, context and process categories indicated in the Tables of Specifications and Summary of Item Specifications.

***The test items of the respondents must conform to the TOS 1, TOS 2 and Summary of Item Specifications, NOT the other way around.**

Item Q1	Feedback: This requires more than recognizing and identifying mathematics information or data that can be used to solve problem situations. This also needs performing computations and applying the mathematical concepts to arrive at a solution to the problem. Context may be accessible only to those who are into basketball.
Item Q2	Feedback: Try to lessen irrelevant information in the context.
Item Q3	Feedback: Improve sentence construction. For example: Suppose in a barangay, one person was identified positive for the disease in one day, 3 more persons were infected in the following day, and on the third day 9 patients were afflicted by the disease.
Item Q4	Feedback: Try to lessen irrelevant information in the context.

METHODOLOGY

ATENEO DE MANILA UNIVERSITY

Data Collection

Item Q5	Feedback: Check the context.
Item Q6	Feedback: This requires more than recognizing and identifying mathematics information or data that can be used to solve problem situations. This also needs performing computations and applying the mathematical concepts to be able to answer the question. Properly label the figure.
Item Q7	Feedback:
Item Q8	Feedback: This is more than interpreting. This requires one to reason.
OVERALL FEEDBACK:	

METHODOLOGY

ATENEO DE MANILA UNIVERSITY

Data Analysis

- Descriptive statistics for the Likert scale ratings
- Content analysis and descriptive statistics for the qualitative feedback on the items

RESULTS AND DISCUSSION

ATENEO DE MANILA UNIVERSITY

Results from Quantitative Data

- Descriptive statistics of alignment ratings

Content Categories	Mean Rating	SD	Context Categories	Mean Rating	SD	Process Categories	Mean Rating	SD
QNT	4.22	1.56	PER	4.33	1.03	RSN	4.11	1.53
C&R	4.17	1.34	OCC	4.28	1.13	PSF	3.50	1.29
S&S	4.17	1.69	SOC	4.22	1.17	PSE	4.33	1.14
U&D	4.22	1.52	SCI	3.72	1.53	PSI	4.11	1.18
Overall	4.19	1.50	Overall	4.14	1.23	Overall	4.01	1.31

RESULTS AND DISCUSSION

ATENEO DE MANILA UNIVERSITY

Results from Quantitative Data

- Descriptive statistics of context appropriateness ratings

Desirable context attribute	Mean	SD	Desirable context attribute	Mean	SD
Stimulating and interesting for 15 y/o students	4.35	0.79	Minimal unnecessary linguistic complexity	4.64	0.81
Realistic, reflects real-world scenarios	4.57	0.77	Minimal amount of irrelevant information	4.64	0.89
Problem is plausible and believable	4.56	0.84	Free from offensive or controversial references	4.90	0.51
Context is compatible with content	4.13	1.50	Intentions of task and expectations are clear	4.53	0.92
Accessible regardless of gender, religion or culture	4.75	0.64	Presence of conservative novelty	3.68	0.87
Overall Mean: 4.47			**Overall SD: 0.94**		

RESULTS AND DISCUSSION

ATENEO DE MANILA UNIVERSITY

Coding of Qualitative Data

- Feedback on Alignment (FA), Item component (FC), Linguistics (FL), Mathematical Aspect (FM), Accessibility (FAcc), Overall (FO)

RESULTS AND DISCUSSION

ATENEO DE MANILA UNIVERSITY

Sample Coding:

- Check sentence construction, grammar, and proper use of capitalization of words. (FL*Grammar), (FL*Ortho) The cost for constructing an apartment is not realistic. (FAcc)

- There is no specified years for Jenny to save the needed amount. If the pattern is based on the arithmetic sequence (which reflects that Jenny has an increase every year but her salary seems to be constant at 240,000), the 8% seems irrelevant also. (FM) The context is personal and not occupational. (FA*Context)

RESULTS AND DISCUSSION

ATENEO DE MANILA UNIVERSITY

Sample Coding:

- Check the content again. (FA*Content)
- Item construction can still be improved. (FO)
- This item is good. (FO)

RESULTS AND DISCUSSION

ATENEO DE MANILA UNIVERSITY

Results from Qualitative Data

- There were a total of 103 instances of feedback, with major feedback types: Alignment with categories (34.95%) and Linguistic feedback (23.3%)
- Most of the feedback about alignment were on process categories (44.4%) and content categories (33.3%) (36 instances of feedback on alignment)
- Most of the feedback about linguistics were on grammar (45.83%) and orthography (33.33%) (24 instances of feedback on linguistics)

CONCLUSION

ATENEO DE MANILA UNIVERSITY

- Results show that the teachers produced items of satisfactory quality and are compliant with the PISA mathematics assessment framework

- Teachers, as primary assessment implementors, are capable of writing items that follow standards from ILSAs, although there is a need to understand process categories and content categories more

- Conduct follow-up item-writing PD sessions to encourage sustained training, feedback, and improvement

END OF PRESENTATION

ATENEO DE MANILA UNIVERSITY

ACKNOWLEDGEMENTS

ATENEO DE MANILA UNIVERSITY

The primary author wishes to express his gratitude to the Ateneo de Manila University (through the Office of the Assistant Vice President for Research, Creative Work, Innovation or OAVP-RCWI) for the funding support (conference subsidy for graduate students).

Additionally, he also wishes to express his gratitude to the Office of the Assistant Vice President for Graduate Education (formerly Office of the Associate Dean for Graduate Programs) of the Ateneo de Manila University which provided the primary author a scholarship grant for his PhD.

References: (1 of 3)

Cheng, L. (2000). Washback or backwash: A review of the impact of testing on teaching and learning. *Education Resources Information Center*, *1*(4), 1-33. https://eric.ed.gov/?id=ED442280

De Corte, E. (2003). Intervention research: A tool for bridging the theory-practice gap in mathematics education. In A. Rogerson (Ed.), *Proceedings of the International Conference: The decidable and the undecidable in mathematics education* (pp. 45–55). The Mathematics Education into the 21st Century Project.

de Vries, J. A., Dimosthenous, A., Schildkamp, K., & Visscher, A. J. (2022). The impact on student achievement of an assessment for learning teacher professional development program. *Studies in Educational Evaluation*, *74*, 101184. https://doi.org/10.1016/j.stueduc.2022.101184

Garcia, M. L. B., & Hao, L. C. (2023). Learnings from the use of Screencast Videography in mathematics education research on item-writing. In W.-C. Yang, M. Majewski, D. Meade, & W. K. Ho (Eds.), *Proceedings of the 28th Asian Technology Conference in Mathematics* (pp. 277-286). Mathematics and Technology, LLC. https://atcm.mathandtech.org/EP2023/regular/22055.pdf

Gutierez, S. B., & Kim, H.-B. (2017). Becoming teacher-researchers: teachers' reflections on collaborative professional development. *Educational Research*, *59*(4), 444–459. https://doi.org/10.1080/00131881.2017.1347051

References: (2 of 3)

Kurniasih, N., Emilia, E., & Sujatna, E. (2023). Using item analysis to evaluate a model of PISA-like reading test developed by teachers. *International Journal of Language Testing*, *13*(1), 188-205. https://doi.org/10.22034/ijlt.2023.345275.1169

Kyaruzi, F., Strijbos, J.-W., & Ufer, S. (2020). Impact of a short-term professional development teacher training on students' perceptions and use of errors in mathematics learning. *Frontiers in Education*, *5*, 559122. https://doi.org/10.3389/feduc.2020.559122

Lauer, P. A., Christopher, D. E., Firpo-Triplett, R., & Buchting, F. (2014). The impact of short-term professional development on participant outcomes: a review of the literature. *Professional Development in Education*, *40*(2), 207–227. https://doi.org/10.1080/19415257.2013.776619

Leong, W. S. (2015). Teachers' assessment literacies and practices: Developing a professional competency and learning framework. *Advances in Scholarship of Teaching and Learning*, *2*(2), 1-20.

Organisation for Economic Co-operation and Development. (2016a*). PISA for Development capacity building plan: Zambia*. https://www.oecd.org/pisa/aboutpisa/CBP%20Zambia_FINAL.pdf

Organisation for Economic Co-operation and Development. (2016b). *PISA for Development capacity needs analysis: Cambodia*. https://www.oecd.org/pisa/aboutpisa/Cambodia-Capacity-Needs-Analysis.pdf

References: (3 of 3)

Popova, A., Evans, D. K., Breeding, M. E., & Arancibia, V. (2022). Teacher professional development around the world: The gap between evidence and practice. *The World Bank Research Observer*, *37*(1), 107–136. https://doi.org/10.1093/wbro/lkab006

Senate of the Philippines. (2024, February 20). *Gatchalian eyes funding for PISA 2025 preparations*. https://legacy.senate.gov.ph/press_release/2024/0220_gatchalian1.asp

Sjøberg, S., & Jenkins, E. (2022). PISA: a political project and a research agenda. *Studies in Science Education*, *58*(1), 1–14. https://doi.org/10.1080/03057267.2020.1824473

Vrikki, M., Kyriakides, L., & Dimosthenous, A. (2024). The potential of following-up international large-scale assessment studies: Using PISA 2018 to develop a comprehensive model of effective teaching. *Educational Research and Evaluation*, *29*(5–6), 249–273. https://doi.org/10.1080/13803611.2024.2344094

Yurdakul, B., Başokçu, T. O., & Yazıcılar, Ü. (2020). Evaluation of the professional development program for secondary math teachers on item writing related to higher order thinking skills. *Journal of Teacher Education and Educators*, *9*(1), 83-106. https://dergipark.org.tr/en/pub/jtee/issue/54091/591395