



American Journal of Interdisciplinary Research and Innovation (AJIRI)

ISSN: 2833-2237 (ONLINE)

VOLUME 5 ISSUE 4 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

Science Teachers' Awareness of International Large-Scale Assessments and Related Competencies: A Systematic Review of Evidence, Professional Development, and Classroom Practice

Shandie V. Carbonell^{1*}, Romelyn T. Lagura¹

Article Information

Received: July 02, 2026

Accepted: September 28, 2026

Published: October 14, 2026

Keywords

*Assessment Literacy, ILSAs,
International Large-Scale
Assessments, Professional
Development, Science Teachers*

ABSTRACT

International large-scale assessments (ILSAs) such as the Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS) have been established as benchmarks to evaluate educational systems, but there is limited evidence on the degree of knowledge and use of ILSA-related frameworks by science teachers. The systematic review synthesized empirical evidence on science teachers' ILSA-related knowledge, competencies, classroom practices, and professional development. Across five databases, 48,606 records were checked for eligibility, 161 were screened, and seven studies published between 2019 and 2025 were included in the sample (in Indonesia, France, Saudi Arabia, Western Balkan countries, Croatia, and the US). None of the studies used a validated instrument to objectively measure teachers' knowledge of ILSA purposes, frameworks, results, and uses. Instead, the evidence focused on teachers' knowledge of PISA or TIMSS tasks, cognitive demands, ILSA-related pedagogical knowledge, higher-order thinking skills-oriented planning, professional development, and teaching practices. Professional development was associated with better ILSA-related knowledge and assessment skills, but causal conclusions were limited as there were no randomized comparison groups. The evidence on science teachers' ILSA awareness is limited, fragmented, and largely indirect. Future work should validate ILSA awareness, evaluate classroom assessment practices, and evaluate professional development using comparative and longitudinal approaches.

INTRODUCTION

International Large-Scale Assessments (ILSAs) are now considered as a benchmark for educational performance. In terms of school performance, the Trends in International Mathematics and Science Study (TIMSS) and Program for International Student Assessment (PISA) are the most widely used indicators to assess student achievement and influence national education policy, curriculum and teacher development (Alrwaythi & Al-Otaibi, 2020).

The findings of international assessment have also highlighted the problems in science education. In Indonesia, TIMSS 2015 performance was characterized by knowing and applying domains over higher-order reasoning (Suwarma & Apriyani, 2022). Science literacy research has also shown the difficulty in gaining international competencies (Triwahyuni *et al.*, 2025).

Teachers are instrumental in translating curriculum and expectations for assessment into classroom instruction and assessment. However, the research presented in this review indicates that teachers may not be able to cope with the cognitive requirements of ILSA items, recognize competencies that are assessed, and create higher-order assessment tasks (Hebel *et al.*, 2019; Prihastuti & Widodo, 2019). ILSA-related professional development has thus been explored as a way to enhance pedagogical content

knowledge and assessment design (Alrwaythi & Al-Otaibi, 2020; Triwahyuni *et al.*, 2025).

One of the major conceptual issues is that 'awareness of ILSAs' in the literature is often not recognized. This review focused on ILSA purposes, frameworks, assessed competencies, results, or applications. Proxy or ILSA-related evidence refers to constructs that reflect teachers' knowledge or use of assessment demands (such as understanding PISA or TIMSS tasks, classifying cognitive levels, creating scientific literacy tests and adopting HOTS methodologies of assessment) or ILSA-related pedagogical knowledge. The papers that have been published have largely addressed proxy constructs and not ILSA awareness in general.

This review reviews what the empirical literature is able and cannot provide about secondary science teachers' ILSA-related awareness and competencies. The evidence, factors and context that are reported along with these competencies, the classroom teaching and assessment processes, interventions for teachers in professional development, and the remaining research gaps will be considered in this review.

Research Questions

General question: What is the current state of empirical evidence regarding secondary science teachers' awareness,

¹ Mariano Marcos State University, Philippines

* Corresponding author's e-mail: shandievicentecarbonell@gmail.com

understanding, and use of International Large-Scale Assessments (ILSAs)?

RQ1. What is known about secondary science teachers' awareness, knowledge, and understanding of ILSAs and ILSA-related competencies?

RQ2. What factors or contextual conditions have been reported in relation to science teachers' ILSA-related awareness, understanding, or competencies?

RQ3. What evidence exists regarding the relationship between science teachers' ILSA-related awareness, knowledge, or competencies and their classroom teaching and assessment practices?

RQ4. What professional development strategies or interventions have been reported to strengthen science teachers' ILSA-related awareness, assessment literacy, or competencies?

RQ5. What research gaps and priorities for future investigation are evident in the literature on science teachers and ILSAs?

LITERATURE REVIEW

ILSAs and Their Relevance to Science Teaching

PISA and TIMSS are based on different assessment criteria and concepts of learning. TIMSS is intimately tied to mathematics and science achievement in relation to intended and implemented curricula whereas PISA focuses on how knowledge and competencies are applied in the real world (Orkodashvili, 2018). This difference matters to the teachers in a way that the ILSA framework is very different from the tests: it is not only about what students should know, but also about how knowledge should be applied and performance interpreted. ILSAs are also technically complex. They can be based on item response theory, sampling weights, plausible values and procedures to help cross-system comparisons (Arıkan *et al.*, 2020; Ouyang *et al.*, 2026). Besides achievement tests, contextual questionnaires gather information from students, teachers and schools. Van Staden (2026) suggests that the challenge is not only to build system-level evidence but to translate that evidence into practice-level change. For teachers, this creates a translation problem: how to understand what ILSA results mean, to know what we can learn and what is relevant to education and the learning environment to teach without reducing teaching to test preparation.

Direct evidence from PISA and TIMSS science-teacher studies

There is little direct evidence for how teachers in secondary science understand ILSAs, but there are a number of studies have examined related aspects of ILSAs. Pinto and El Boudamoussi (2009) conducted an open-ended questionnaire based on the PISA 2003 scientific literacy framework on 30 experienced in-service secondary science teachers (natural and physical science). Teachers did not understand the scientific processes they discovered in the PISA items and their expectations for students' difficulties were not consistent

with what the science competencies PISA was supposed to measure. This research provides direct evidence that science teachers may know science assessment but do not understand the science competencies in the same way from the PISA science competencies. Glynn (2012) develops another connection between science teachers and an ILSA framework. A focus group of science teachers assessed TIMSS science achievement items in a representative way with respect to the TIMSS assessment framework in terms of content, cognitive demand, item format, wording, and clarity. Most of the items selected were of good quality, but teachers highlighted that items could be improved for clarity and alignment with the framework specifications. As a result of this study, the study demonstrated that published or representative TIMSS items that can serve as measurement tools for examining reasoning as a professional resource in international assessment.

Assessment and data literacy as supporting competencies

Science-teacher assessment literacy is good evidence for understanding what competencies are required for ILSA engagement. Hung and Wu (2024) studied 40 high school science teachers in Taiwan who had worked with inquiry-based science teaching. The professionals looked at the objectives of assessment, learning outcomes, assessment strategies and scoring. Teachers used many sources of evidence and formative assessment throughout the course of teaching to support inquiry and self-regulated learning and to show that assessment literacy involves sharing what is assessed, why it is assessed, and how evidence is gathered and interpreted. This research had not been aimed at PISA or TIMSS awareness but science-based findings would help make the process of moving competency-based assessment frameworks from theory to practice clear.

The Philippine evidence also supports the context of assessment. Tesorio and Canizares (2018) interviewed secondary science teachers from 30 public and private schools in Cebu and showed the teachers positively correlated assessment with student learning, diagnosing student abilities and instruction. Walag *et al.* (2022) also found that Filipino science teachers' scientific literacy was related to their science teaching ability and knowledge about Science, Technology & Society and other science content domains was positively related to confidence in science teaching. These studies do not measure ILSA awareness in itself but they do suggest assessment conceptions and scientific literacy are at the centre of educators' ability to interpret and act on ILSA-related scientific literacy objectives.

ILSA knowledge also requires a degree of data literacy. Teachers need not necessarily have to do more advanced secondary analyses of international databases but they have to know what large-scale results can and cannot show. Arıkan *et al.* (2020) show sample weights and plausible values are relevant for large-scale assessment analysis and

Freddano and Pandolfini (2022) describe data literacy as an emerging competency for teachers who engage in school evaluation and improvement. These studies suggest that the teachers' knowledge of ILSAs should not only be about PISA or TIMSS but also about how to read results in a cautious and meaningful manner and see how to relate them to school decisions. Interpretation problems can arise from the instruments themselves. Oz (2021) studied TIMSS teacher background items in Turkey and found comparability and equivalency issues in teacher questionnaire data. This evidence does not indicate that teachers are generally wrong about ILSAs, but rather that international instruments may carry concepts not interpreted in a similar way in different educational contexts. These problems only serve as a reminder of the need for teacher orientation to ILSA frameworks to be contextually informed.

What the current evidence allows us to conclude

As a group the literature is more nuanced but cautious. Direct PISA/TIMSS studies with science teachers have demonstrated that teachers can identify international assessment items and scientific literacy constructs, and that their own interpretation of assessed competencies is not always consistent with the frameworks (Glynn, 2012; Pinto & El Boudamoussi, 2009). Research also shows that assessment literacy, scientific literacy, and data interpretation are essential professional competencies (Hung & Wu, 2024; Tesorio & Canizares, 2018; Walag *et al.*, 2022). However, there are too few and context-dependent studies to establish a common ILSA awareness among secondary science teachers. In the future there must be more empirical research that distinguishes the knowledge of PISA or TIMSS from a deeper understanding of assessment purposes, scientific literacy competencies, item characteristics, result interpretation, and appropriate classroom usage.

Factors Influencing Teachers' Ability to Implement ILSA-Related Frameworks

Teacher experience and professional knowledge
Teacher characteristics will shape the way that new assessment and curriculum expectations are understood and adopted. Watson *et al.* (2020) found that teaching experience was associated with the knowledge of standards and resources for implementing large-scale curriculum. Diaz *et al.* (2026) also reported differences in teachers' experiences in implementing the MATATAG Curriculum in a Philippine curriculum setting. Such findings are contextual, but not ILSA-based, but they underscore the need for continued exposure to curriculum implementation and professional practice to raise teachers' confidence in the ability to interpret and adapt new frameworks.

Experience alone, however, shouldn't be conflated with ILSA expertise. Teachers may be very experienced in classroom teaching but not familiar with international assessment frameworks and how to interpret the results

of large-scale assessments. Knowledge of the profession should also be a mixture of pedagogical experience and training and knowledge of the skills international assessments are looking for in many areas. Teacher scientific literacy can also influence skills to implement the curriculum.

In the Philippine context, Walag *et al.* (2022) found a correlation between science teachers' scientific literacy and science teaching success. They found that knowledge of the content (science, technology and society) can help them teach effectively. If the research is ILSA-informed, then this matters because PISA science is about teachers working with disciplinary content and how science is used, in relation to evidence, context and society. When assessing teachers' readiness for ILSA-related competencies, scientific literacy should be considered along with assessment literacy.

Curriculum alignment

The curriculum alignment is also important. Pestovs *et al.* (2019) have shown that large-scale assessment constructs may not match up with new competency-based curriculum expectations. If teachers see a gap between what the curriculum requires and what an external assessment measures, the classroom value of the assessment may be unclear. Orkodashvili (2018) similarly emphasises that learning from international assessment should be taken into account along with local testing, curriculum and international benchmarks.

Alignment is particularly important for science teaching, as ILSA competencies might include inquiry, reasoning, application and interpretation rather than simply recall. Teig (2020) assessed scientific inquiry in TIMSS and PISA 2015 and found that inquiry is not just an instructional approach but an assessed outcome. The relationship between teaching and science literacy is also complex in the context of PISA science data by McConney and Woods-McConney (2019), but we would suggest that one approach, like frequent inquiry, doesn't automatically translate into ILSA-aligned teaching.

Institutional and policy support

Teachers' ability to implement new frameworks is very much up to the school environment. This relationship is well established in ILSA-based science-teacher research. Ogegbo *et al.* (2024) examined TIMSS 2019 data for 537 Grade 9 science teachers in South Africa and found that inquiry-based teaching practices were influenced by the educational environment (e.g., teacher job satisfaction, lack of instructional resources, teachers' view of assessment strategies and professional development). These findings are of great relevance to the implementation of ILSA because they demonstrate that teachers' classroom engagement with inquiry is not just about pedagogical knowledge but also what works in the classroom, assessment beliefs, resources and career opportunities they have.

Cross-country TIMSS evidence also indicates that

implementation is contextual. Chen (2026) investigated science-teacher reports from TIMSS 2019 in 14 countries and found different cognitive activation and inquiry-related teaching profiles. The distribution and achievement associations of these profiles were different in different countries. Rather than providing an overall pattern of good science teaching, the study indicates that science teaching is embedded in national and school contexts. This is in line with the view that ILSA-informed recommendations should be seen as professional judgment recommendations and not just a good copy from better-performing systems.

Policy coherence is also important. Kiamco (2025) states the differences in teacher quality, curriculum, and assessment style are also responsible for the differences in PISA performance in Singapore and the Philippines. Although a system-level comparison can not identify the ILSA level of the individual teachers, it is an indication of teacher practice and policy practice being integrated into a more general context. Teachers are more likely to take on ILSA-related competencies when the curriculum, assessment guidance, professional development and school expectations are consistent.

Barriers to Effective ILSA-Informed Practice

Technical and data-related barriers

One major obstacle is the technical difficulty of large-scale assessment. Weighting, plausible values and other statistical features can lead to secondary analysis that is inaccurate (Arikan *et al.*, 2020). At the classroom level, the more immediate issue is that teachers may obtain rankings or summary results that don't explain how or why they were obtained and how they are interpreted or what interpretations are made. It's a shortcoming for teachers to make ILSA evidence a basis for professional reflection.

Item-writing and assessment-design challenges

Translating an international assessment framework into classroom assessment is challenging, and science-teacher studies reveal a knowledge-to-practice gap. Pitot and Balgopal (2021) examined scientific literacy among secondary science teachers, their perceptions of scientific literacy skills, and teacher-crafted common assessments. The teachers were scientifically literate, but their perceptions were not consistent with the science education reform documents they were provided and their assessments were not representative of the scientific literacy subconstructs. The research isn't so much about PISA as it is about the translation of ILSA scientific literacy frameworks into classroom assessment tasks.

Knowing a construct does not imply that the assessment tools will accurately measure it. Teegaza *et al.* (2015) also provide more specific TIMSS-based evidence from 350 intermediate science teachers in Saudi Arabia. Using TIMSS 2011 performance classifications and a teacher survey, they examined teachers' actual assessment practices and the value teachers were assigned to

different assessment methods. Teachers' assessment practices did not vary only by whether schools were classified as higher or lower-performing in TIMSS, but also the teachers' perception of the value of assessment practices when it came to implementation. The findings indicate that participation in an international assessment or even school performance in that assessment does not guarantee ILSA-aligned classroom assessments. Teachers' beliefs about assessment and their assessment methods remain relevant for assessing the assessment process.

Questionnaire interpretation is also a barrier. Oz (2021) shows that even translated teacher-background items in TIMSS can raise comparability questions. The implication is not that international frameworks can't be used, but that teachers and education systems should be given opportunities to see how the language and constructs of ILSAs connect with local curriculum and professional practice.

Resource, time, and implementation barriers

The Philippine evidence suggests that the barrier may not be a lack of positive assessment beliefs. Tesorio and Canizares (2018) found that secondary science teachers in the field generally viewed assessment as a way to enhance learning, identify students' abilities, and improve the quality of instruction. The question is then if these positive assessment beliefs are also true for the more specific capabilities of ILSAs such as interpretation of evidence, scientific investigation, applying knowledge in new situations and reasoning from data. This is important because positive beliefs about assessment don't necessarily mean that we are familiar with international assessment practices.

A lack of resources and implementation barriers can prevent teachers from applying competency frameworks into practice. Diaz *et al.* (2026) discuss problems in curriculum implementation in geographically isolated and disadvantaged Philippine settings, including poor access to support and resources. Kisirko and Hough (2025) also discuss technology, materials, and institutional support as vital and relevant. The curriculum coverage pressure can make deeper learning difficult, Orkodashvili points out (2018). These barriers are not unique to ILSAs, but they have a significant effect on teachers' ability to use assessment and instruction to assess and apply reasoning, application, and higher-order thinking.

Policy-to-practice barriers

Such large scale assessment policy can have unintended consequences if it is translated into accountability rather than professional learning. Elyashiv and Avidov-Ungar (2023) studied a national large scale assessment system and found that a top-down assessment regime might restrict teacher autonomy and instead lead to teaching to the test rather than real pedagogical change. Because this study is national rather than international, it should be seen as similar evidence. However, it is a very real risk: ILSA-informed reform is not likely to enhance classroom

practice if teachers have to be subjected to assessment results only as an incentive to raise scores.

Professional Development Models for ILSA-Related Competencies

Targeted assessment-literacy development

One of the most common messages in the literature is that professional development should be targeted rather than generic and linked to actual science-teaching practice. You *et al.* (2025) evaluated 66 studies on professional development in science education and found a strong positive impact on teacher and student outcomes (Hedges' $g = 0.772$). Effects on outcomes varied by program dosage, duration, and opportunities for active teaching. While not ILSA-specific, this meta-analysis is stronger than generic teacher development studies in supporting the proposition that well-designed science-teacher professional development can enhance the knowledge and practices on which ILSA-informed teaching depends.

Collaborative professional learning

The practice-oriented and inquiry-focused professional development is also promising. Yang *et al.* (2020) investigated an in-service Interdisciplinary Science Inquiry professional development program that involved 204 teachers and 5,581 students in 12 public schools in the context of the study. They found that professional development, school factors, and teacher factors were related to pedagogical content knowledge and inquiry learning, and inquiry teaching was more important for students to understand interdisciplinary science concepts. Professional development plans that bridge teacher learning with classroom action and the learning experience are better than just awareness training. Hung and Wu's (2024) work with 40 high school science teachers suggests a useful content focus as well which can be used for such training: assessment and learning outcomes in the context of inquiry, assessment strategies, scoring, and evidence-based approaches. For an ILSA-specific program, these components could be paired with close examination of PISA or TIMSS science frameworks, released items, scoring guides, and examples of how scientific reasoning and evidence use can be assessed in ordinary classroom tasks.

Data-literacy and needs-based models

Data literacy training is another example of such. Freddano and Pandolfini (2022) describe teacher professional development based on the interpretation and use of assessment data for school evaluation and improvement. Needs-based professional development has shown promise for improving pre-service teachers' assessment-for-learning literacy (Oo *et al.*, 2023). These studies are evidence-based, not directly assessing ILSA-specific programs, and together they suggest that professional development is most useful when it is designed to address identified competency gaps and to

give teachers a concrete opportunity to utilize what they learn.

Integrating ILSA competencies into teacher preparation

Teacher education is one of the best routes for ILSA-related competence development. Espinosa *et al.* (2024) propose to integrate key ILSA skills in teacher education curricula in the Philippines by incorporating content preparation, pedagogical content knowledge, and ILSA-like assessment experiences. This is especially important, particularly as professional development after teachers are in the workforce in order to overcome an absence of exposure to international assessment frameworks. In addition, science-based professional development in assessment frameworks can also relate to pedagogical content knowledge. Teig (2020) emphasizes the need for a distinction between inquiry as an instructional method and inquiry as a scientific competency that is to be assessed. ILSA-related development for science teachers must therefore not only encourage teachers to use more specific strategies to inform their assessment but also help them gain a grasp of how scientific reasoning, evidence analysis, inquiry, and application are represented in assessment tasks and how these competencies can be gained through instruction.

MATERIALS AND METHODS

Eligibility Criteria

Studies were eligible if they were for (a) secondary level science teachers or science educators; (b) considered an International Large-Scale Assessment (TIMSS, PISA or a related cross-national assessment), (c) concerned teacher awareness, knowledge, understanding, familiarity, recognition or ILSA-related competency or practice; (d) reported empirical quantitative, qualitative, mixed-methods, quasi-experimental, case-study or secondary data evidence; and (e) published in English as journal articles or theses with enough empirical proof to be considered. Studies were excluded if they related to elementary/primary teachers, pre-service teachers or non-science teachers; covered only national or local assessments; were editorials, commentaries or conference abstracts if they did not have original empirical evidence; or if they provided no detailed information for assessing them to be done. Although ILSA awareness was not well measured in terms of quantitative information, proxy constructs were considered relevant to ILSA knowledge, assessment literacy, professional development or classroom practice.

Information Sources and Search Strategy

Systematic searches were conducted on five sites (ERIC, Google Scholar, PubMed, Semantic Scholar, and E-Palli). 13 search queries combined population terms: 'science teacher', 'science educator' and 'secondary science teacher' with ILSA terms like 'PISA' and 'TIMSS' and 'International Large-Scale Assessment' and construct

terms like awareness, knowledge, understanding, professional development and teaching practices. The same broad concept structure was used in all five sources to create the same way, and we adjusted the syntax for each.

Selection Process

This review found 48606 records in five databases. 40971 records were excluded from the search based on exclusion

criteria. 6602 of them were not ILSA related and thus were removed in terms of their eligibility, leaving a total of 1033 articles. This resulted in 1033 articles being screened manually, and 161 articles were evaluated for eligibility (full text). These included 154 articles, while 7 articles were included in the review. The removal of the data without abstracts before the full text is retrieved is a limitation as some studies that might have been eligible may have been missed.

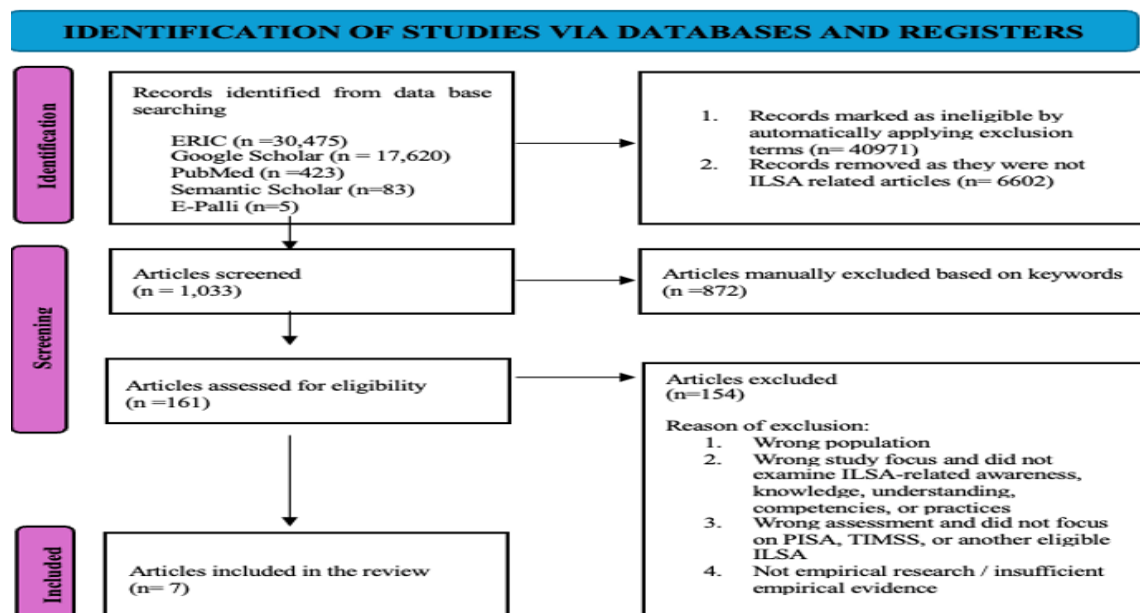


Figure 1: PRISMA Flow Diagram of the Study Selection Process

Data Extraction, Synthesis and Methodological Appraisal

Data were collected from author and year, country or region, research design, participant characteristics, ILSA focus, construct or outcome examined, key findings and methodological limitations. Since the design and outcome of the study were heterogeneous, findings were synthesized narratively and not statistically. Evidence was classified as it reflected ILSA awareness or ILSA-related knowledge, skills, development or classroom practice.

In the study design, we looked at limitations such as the choice of participants, whether a comparison or

control was appropriate, the presence of outcome data, outcome measurement, and selective reporting. This review considered the domain that is not relevant for a specific design to be not applicable. Appraisal results are used to qualify the interpretation of the results and not imply that a single level of certainty is held for different outcomes and study designs.

RESULTS AND DISCUSSION

Included Study Characteristics

This review included seven studies, published from 2019 to 2025, among them Indonesia, France, Saudi Arabia,

Table 1: Summary of Studies Included in the Review

Study	Context and design	Main construct examined	Relation to ILSA awareness
Suwarma & Apriyani (2022)	Indonesia; case study / teacher development context	HOTS-oriented lesson planning and assessment development	Indirect/proxy: classroom and assessment competency related to higher-order demands
Hebel <i>et al.</i> (2019)	France; questionnaire study; 125 science and technology teachers	Prediction of student difficulty and recognition of competencies in PISA science inquiry tasks	Proxy: item-level understanding of PISA task demands and competencies
Alrwaythi & Al-Otaibi (2020)	Saudi Arabia; one-group pretest-posttest; 20 female science teachers	TIMSS-requirements-based PLC and pedagogical content knowledge	Proxy/intervention: TIMSS-related professional knowledge

Bikić <i>et al.</i> (2024)	Western Balkan countries and Croatia; secondary analysis of TIMSS 2019	Teacher professional development and student achievement patterns	Indirect/contextual: PD indicators associated with TIMSS outcomes
Prihastuti & Widodo (2019)	Indonesia; 30 secondary science teachers	Classification of cognitive levels in TIMSS science items and teacher-made assessments	Proxy: understanding of cognitive demands reflected in TIMSS items
Triwahyuni <i>et al.</i> (2025)	Indonesia; blended training intervention with PLC components	Development of scientific-literacy assessment instruments	Proxy/intervention: assessment design competency related to scientific literacy
Haridza & Ding (2024)	United States; hierarchical linear modeling; 456 science teachers and 8,739 Grade 8 students	Evidence-based explanations, instructional practice, and student reasoning	Indirect/practice: TIMSS-based evidence on instructional practices rather than teacher awareness

Western Balkan countries and Croatia and the United States. Designs included questionnaire-based studies, case studies, one-group quasi-experimental interventions and secondary analyses of TIMSS data. The studies varied substantially in how closely they addressed ILSA awareness itself.

Nature of the Evidence: Direct and Proxy Measures

The clearest cross-study finding was the absence of direct measurement. None of these seven studies was based on a validated instrument to assess the overall level of ILSA awareness of science teachers' knowledge of the purposes, frameworks, results or uses they are studying. Most of them were also proxy constructs such as cognitive-level classification, interpretation of PISA or TIMSS items, pedagogical content knowledge, scientific literacy assessment design, HOTS planning, professional development and instructional practice. As a result, the review can better describe ILSA-related knowledge and competencies than it can say to what extent a general level of ILSA awareness is.

The lack of context also makes it harder to predict correlations between variables. For example, a TIMSS study based on instructional practice data might inform the review of ILSA-associated classroom patterns, but it does not mean that a teacher who is more familiar with TIMSS will teach differently. So, the results below are also based on the construct that each study actually measured.

Science Teachers' Awareness, Understanding, and ILSA-Related Knowledge

Hebel *et al.* (2019) presented evidence of teachers' knowledge of PISA science inquiry tasks. French science and technology teachers were able to predict that tasks would be challenging for medium-low-achieving students, but they were not able to identify the competencies that were assessed by the PISA items. That is, understanding the difficulty of a task is not enough to know what competencies are needed.

Prihastuti and Widodo (2019) found that Indonesian

science teachers did not correctly categorize cognitive levels with TIMSS science items. Only 37% of respondents correctly classified the remembering-level item and 63% placed it at higher cognitive levels. For the understanding-level item, only 7% correctly classified it and 93% placed it at higher cognitive levels (70% at the analyzing level). Teacher-developed summative assessments (in conjunction with experts) also found that items that were considered higher-level were sometimes less cognitively demanding.

Alrwaythi and Al-Otaibi (2020) reported that science teachers did not have enough knowledge of TIMSS-related pedagogical skills. Their intervention aimed at developing TIMSS-related pedagogical skills knowledge. These studies, together, suggest gaps in some ILSA-related knowledge and assessment competencies, but they do not provide a complete picture of teachers' general ILSA awareness.

Factors and Contextual Conditions Related to ILSA-Related Competencies

The studies included had a number of contextual variables relevant to the ILSA-related competencies development (professional development exposure, teaching experience, institutional support and access to assessment resources). However, these variables were not tested as predictors of ILSA awareness in general and so they should be considered as less than that as a matter of local and situational context rather than specific causes of ILSA awareness.

Bikić *et al.* (2024) in TIMSS 2019 data from Western Balkan countries and Croatia reported differences in teacher experience, age profiles, reported professional development needs and associations between various professional development indicators and student performance across countries for teacher experience and student achievement. Professional development in problem-solving skills was strongly associated with student performance in Bosnia and Herzegovina while the other professional development indicators were

different in different countries. These findings support the importance of teacher development context to TIMSS-related outcomes but do not provide definitive data that more teaching experience or professional development is associated with higher ILSA awareness.

Also Haridza and Ding (2024) highlighted the importance of science resources and professional learning in relation to science content, pedagogy and critical thinking as well as evidence-based explanations for science lessons. This evidence is relevant to the environment in which ILSA-related classroom competencies can develop, but not directly related to teachers' awareness of TIMSS itself.

ILSA-Related Competencies and Classroom Teaching and Assessment Practices

The correlations between ILSA-related constructs and classroom practice were relatively small and indirect. Haridza and Ding (2024) found that, based on TIMSS 2019 data, teachers' need to ask students to give detailed evidence-based explanations was related to students' ability to perform reasoning-related tasks. The study does shed light on teaching procedures related to reasoning but does not explain what teachers do with TIMSS in general. Suwarma and Apriyani (2022) found that teachers' HOTS-based lesson-planning skills are intermediate, and vary by lesson plan. Some plans had tasks that were predictive and analytical while others had tasks that were easy to identify. We also found assessment development very difficult for many teachers. Teachers may have interpreted HOTS assessment as simply asking "hard questions." Teacher-developed questions were not clearly higher-order in meaning.

Prihastuti and Widodo (2019) also reported a disconnect between teachers' intended cognitive levels and the cognitive demands identified by experts. The strongest conclusion from these studies is that some science teachers have difficulty translating higher-order or ILSA assessment needs into classroom assessment tasks. There is no clear statistical evidence for a relation between ILSA awareness and classroom assessment practice.

Professional Development Interventions

Three studies have investigated the professional development of ILSA skills and practices. Alrwaythi and Al-Otaibi (2020) investigated a Professional Learning Community (PLC) based on TIMSS requirements among 20 female science teachers in a one-group pretest-posttest setting. The post-intervention pedagogical content knowledge scores were significantly higher than the pre-intervention scores, with an eta squared of 0.9039 for the total PCK scale. They also reported improvements in scientific content knowledge, educational purposes, curriculum knowledge, teaching practice knowledge, and educational context knowledge. In PLC activities, reflective practice, peer visits, action research, workshops, focus sessions, and facilitated discussion were conducted. Triwahyuni *et al.* (2025) also reported improvements in the BLAST-SCI blended-learning training model

and PLC activities. N-Gain scores were 73% to 84% and Aiken's V of 0.817 for instrument validity. The intervention involved face-to-face and online learning and was designed to evaluate teachers' ability to develop scientific literacy assessments.

Suwarma and Apriyani (2022) and Prihastuti and Widodo (2019) also found that teachers need more experience building and executing higher-order assessment tasks. Prihastuti and Widodo (2019) reported that previous district-level training was brief (about five to six hours) and did not have direct practice of generating valid and reliable HOTS items. These intervention results are promising but should be taken with caution. The most significant improvements were reported in one-group pretest-posttest studies without randomized comparison groups, which limits causal inference and makes it hard to distinguish the effects of the interventions from other reasons for improvement.

Research Gaps and Future Priorities

The most common gap among the seven studies is measurement. The literature reviewed here did not provide a validated instrument that directly measures science teachers' overall awareness of ILSA purposes, frameworks, results, assessed competencies, and classroom uses. Instead, studies generally measured related constructs such as cognitive-level understanding, PCK, assessment-development skills, scientific literacy assessment, or instructional practices.

There are multiple studies that described ILSA-related instructional or assessment competencies but none of these studies directly tested whether science teachers with more general ILSA awareness practice different classroom assessment practices. This is a crucial distinction since understanding assessment frameworks and actual classroom practice are closely related but not interchangeable.

Professional development research needs to be more robust and longer-term. The intervention studies presented in this section show the benefits of working together and blended approaches, but it is currently not clear if the impact is long-term sustainable and transferable to the classroom. Continued mentoring and practice-based support is also very important (Triwahyuni *et al.*, 2025; Prihastuti & Widodo, 2019).

Of course, more cross-context evidence is needed. The seven studies are diverse in scope but the constructs, interventions, and outcomes are not uniform across settings. Future research in this area should be done to support awareness measures in different contexts and how institutional resources, professional learning opportunities, teaching experience, and school-level support relate to ILSA-related knowledge and practice.

Discussion

Summary and Interpretation of Findings

There were seven studies in this review on knowledge, understanding, and use of ILSA-related ideas in

secondary science teachers. The main finding is not that a particular proportion of science teachers are 'aware' or 'unaware' of ILSAs because studies that were included in the review did not assess that construct directly with a validated overall instrument. Rather, there is a lack of clear evidence of core competencies that might constitute ILSA understanding, such as interpretation of inquiry tasks, recognition of cognitive demand, higher-order assessment, pedagogical content knowledge, scientific literacy assessment design, and evidence-based instruction.

What item-based studies demonstrate is that teachers may not see the real meaning of what an assessment task is measuring. Hebel *et al.* (2019) showed that teachers had a good understanding of some aspects of PISA task difficulty, but they could not identify competencies. Prihastuti and Widodo (2019) found systematic overestimation of cognitive levels in TIMSS items and teacher-developed assessments. These results suggest that cognitive or competency needs might be difficult to identify but are teachable.

And professional development studies suggest ILSA-related competencies are teachable as well. TIMSS-based PLC work and combined scientific literacy assessment training related to training lead to large pre-post improvements (Alrwaythi & Al-Otaibi, 2020; Triwahyuni *et al.*, 2025). In addition, the absence of randomized control groups means that any causal effect is not certain. The most reasonable answer is that collaborative and practice-oriented professional development is actually very promising and needs further investigation and evaluation.

This review also fills an important gap in the knowledge of classroom practice. Higher-order assessment and reasoning-centered instruction are known to be present, but the current studies do not directly assess how ILSA awareness is related to classroom assessment practice. In future research, awareness should be formally measured and how teachers design, select, and use classroom assessments should be studied.

Strengths and Limitations

The strengths of this review are the focus on a relatively narrow and underexamined topic and the fact that Asian, European and North American studies are included. Some studies used authentic TIMSS or PISA items or TIMSS datasets so that the synthesis is linked to real ILSA frameworks and not just researcher-developed tasks. In addition, the review distinguishes direct awareness from proxy knowledge and therefore classroom/assessment competencies are not automatically translated to general ILSA awareness in the present study.

However, there are a few limitations in the evidence base. None of the studies directly measured overall ILSA awareness using a validated instrument. Most designs were cross-sectional, case-based, secondary-data, or one-group quasi-experimental designs, and causal interpretation was limited. Intervention studies

did not contain randomized comparison groups, and the outcomes of the interventions were not uniform, so it was difficult to pool.

The review process was also limited. The study was only made up of five sources and English-language studies and only 118 records were removed without abstracts before the full-text was retrieved. This could have resulted in missing eligible studies. Moreover, broad proxy constructs were necessary as direct awareness studies were scarce, but this increased conceptual heterogeneity. In this regard, the results should be considered as a map of the existing evidence and not as an accurate estimate of the prevalence or level of ILSA awareness among science teachers.

Implications for Practice and Research

The evidence reviewed supports a professional development model that goes beyond information-only sessions and offers teachers the opportunity to interact with real or ILSA-aligned items, categorize cognitive demands, develop assessment tasks, receive feedback, and discuss the implementation with colleagues. PLC activities, reflective practice, peer observation, action research, and blended training can all be good things to do, although stronger assessment will be necessary before causal conclusions can be drawn (Alrwaythi & Al-Otaibi, 2020; Triwahyuni *et al.*, 2025).

Instrument development is the first focus for research. An indicator of science teachers' ILSA knowledge should be the knowledge of what those ILSAs do, the way they are used, how they work, and why and how they do so. The second focus is to see how this knowledge is related to how well they know the classroom and how they use it. At the same time, longitudinal and comparative studies might look at whether professional development leads to longer-term changes in knowledge and practice.

Future systematic updates should also enhance search coverage and record full database-specific search strings and full-text exclusion reasons so the review process can be fully replicated.

CONCLUSION

The systematic review uncovered a small and very different evidence base for secondary science teachers and International Large-Scale Assessments. The most significant finding is that the direct measurement of teachers' ILSA awareness is not found in the seven studies identified. Teachers' interpretation of PISA and TIMSS tasks and cognitive level classification, higher-order and scientific literacy assessment, ILSA-related content knowledge, professional development and instructional practices are the indirect evidence that is provided. These studies clearly indicate that cognitive and competency needs are not well understood and that developing interdisciplinary and hands-on professional development in a professional context might lead to enhancement in such competencies. The evidence for intervention outcomes is not strong enough to demonstrate any causal

effect. Future research needs to develop validated ILSA-awareness instruments, directly compare awareness with classroom assessment practices, and more closely examine professional development using more comparative and longitudinal models.

REFERENCES

- Alrwaythi, E. M., & Al-Otaibi, R. T. (2020). Effectiveness of TIMSS requirements-based professional learning community in developing pedagogical content knowledge among science teachers. *Journal of Educational and Social Research*, 10(2), 24. <https://doi.org/10.36941/jesr-2020-0024>
- Arıkan, S., Özer, F., Şeker, V., & Ertaş, G. (2020). The Importance of Sample Weights and Plausible Values in Large-Scale Assessments. *Journal of Measurement and Evaluation in Education and Psychology*, 11(1), 43-60. <https://doi.org/10.21031/epod.602765>
- Baah, D., Yunus, H., & Vlahek Štok, A. (2025). Analyzing the contribution of teacher professional development to mathematics outcomes: A Ghanaian perspective from TIMSS 2022. *Interval: Indonesian Journal of Mathematical Education*, 3(1), 102–108. <https://doi.org/10.37251/ijome.v3i1.1732>
- Bikić, N., Buzadija, N., & Nuić, I. (2024). Impact of teacher professional development on student achievement with a focus on Bosnia and Herzegovina: Analyzing TIMSS 2019 results in Western Balkan countries and Croatia. *International Journal of Professional Development, Learners and Learning*, 6(1), ep2405. <https://doi.org/10.30935/ijpdll/14388>
- Chen, Z., Carrasco, D., & Sandoval-Hernández, A. (2026). Profiles of teachers built on cognitive-activation teaching strategies and its effect on student's outcomes in TIMSS 2019. *Studies in Educational Evaluation*, 89, 101610. <https://doi.org/10.1016/j.stueduc.2026.101610>
- Diaz, K. P., Galendez, M. L. A. F., & Lawas, D. D. (2026). Teachers' perception in the implementation of MATATAG curriculum: Approach to content, delivery, and impact measurement. *International Journal of Innovative Science and Research Technology*, 11(3), 2654–2678. <https://doi.org/10.38124/ijisrt/26mar940>
- Elyashiv, R. A., & Avidov-Ungar, O. (2023). Teachers' perceptions of national large-scale assessment: The pedagogical dimension. *Educational Review*, 76(6), 1691–1707. <https://doi.org/10.1080/00131911.2023.2256996>
- Espinosa, A. A., Gomez, M. A. C., Miranda, P. A., David, A. P., Macahilig, H. B., Reyes, A. S., Cortez, L. A. S., Palomar, B. C., de Vera, J. L., Patal, M. C., & San Juan, M. P. C. (2024). Integrating the key competencies of international large-scale assessment (ILSA) into the teacher-education curriculum in the Philippines. *Journal of Teaching and Learning*, 18(2), 173–193. <https://doi.org/10.22329/jtl.v18i2.8898>
- Freddano, M., & Pandolfini, V. (2022). Teacher training on data literacy for school evaluation and improvement: A teacher professional development experience in Italy. *Education and Self Development*, 17(1), 89–98. <https://doi.org/10.26907/esd.17.1.08>
- Glynn, S. M. (2012). International assessment: A Rasch model and teachers' evaluation of TIMSS science achievement items. *Journal of Research in Science Teaching*, 49(10), 1321–1344. <https://doi.org/10.1002/tea.21059>
- Haridza, R., & Ding, L. (2024). Enhancing students' reasoning skills by asking students to provide evidence-based explanations in science classrooms: Findings from TIMSS 2019. *International Journal of Science Education*, 2302–2327. <https://doi.org/10.1080/09500693.2024.2385067>
- Hebel, F. L., Tiberghien, A., Montpied, P., & Fontanieu, V. (2019). Teacher prediction of student difficulties while solving a science inquiry task: Example of PISA science items. *International Journal of Science Education*, 41(11), 1517–1540. <https://doi.org/10.1080/09500693.2019.1615150>
- Hung, C. S., & Wu, H. K. (2024). High school science teachers' assessment literacy for inquiry-based science instruction. *International Journal of Science Education*, 46(7), 621–642. <https://doi.org/10.1080/09500693.2023.2251657>
- Kiamco, K. (2025). Teacher quality, curriculum, and assessment in comparative perspective: Explaining PISA performance gaps between Singapore and the Philippines. *International Journal of Education Science*, 2(2), 104–121. <https://doi.org/10.46328/ijofes.5624>
- Kisirkoi, F., & Hough, T. (2025). The teacher, technology, and the implementation of the competency-based curriculum in Kenya: A case study. *Journal of Contemporary Issues in Open Distance and E-Learning*, 2(1), 54–88. <https://doi.org/10.71060/3kwe8k36>
- McConney, A., & Woods-McConney, A. (2019). Inquiry-based, teacher-directed and adaptive instruction in secondary science: A cross-national analysis of associations with science literacy and interest using PISA 2015. In *ICERI2019 Proceedings* (pp. 6045–6050). IATED. <https://doi.org/10.21125/iceri.2019.1464>
- Ogebo, A. A., Ramnarain, U., & Krajcik, J. (2024). Factors predicting teachers' implementation of inquiry-based teaching practices: Analysis of South African TIMSS 2019 data from an ecological perspective. *Journal of Research in Science Teaching*, 61(9), 2069–2103. <https://doi.org/10.1002/tea.21943>
- Oo, C. Z., Alonzo, D., & Davison, C. (2023). Using a needs-based professional development program to enhance pre-service teacher assessment for learning literacy. *International Journal of Instruction*, 16(1), 781–800. <https://doi.org/10.29333/iji.2023.16144a>
- Orkodashvili, M. (2018). Learning from successful examples: From local testing to international assessments. In A. W. Wiseman & P. M. Davidson (Eds.), *Cross-nationally comparative, evidence-based educational policymaking and reform* (Vol. 35). Emerald Publishing Limited. [<https://journals.e-palli.com/home/index.php/ajiri>](https://doi.org/10.1108/S1479-</p>
</div>
<div data-bbox=)

- 367920180000035007
- Oz, E. (2021). Comparability of teachers' educational background items in TIMSS: A case from Turkey. *Large-scale Assessments in Education*, 9, Article 4. <https://doi.org/10.1186/s40536-021-00097-2>
- Ouyang, J., Chen, Y., Li, C., & Xu, G. (2026). Accounting for measurement bias: A new framework for reliable country ranking in large-scale educational assessments. *Journal of the American Statistical Association*. <https://doi.org/10.1080/01621459.2026.2670732>
- Pestovs, P., Namsone, D., Čakāne, L., & Saleniece, I. (2019). Alignment of 6th grade large-scale assessment constructs with the revised curriculum framework. Society. Integration. Education. *Proceedings of the International Scientific Conference*, 2, 387–398. <https://doi.org/10.17770/SIE2019VOL2.3811>
- Pinto, R., & El Boudamoussi, S. (2009). Scientific processes in PISA tests observed for science teachers. *International Journal of Science Education*, 31(16), 2137–2159. <https://doi.org/10.1080/09500690802559074>
- Pitot, L. N., & Balgopal, M. (2021). Science education reform conundrum: An analysis of teacher developed common assessments. *School Science and Mathematics*, 121(5), 299–309. <https://doi.org/10.1111/ssm.12472>
- Prihastuti, I., & Widodo, A. (2019). Teachers understanding about cognitive level on science test items. *Journal of Physics: Conference Series*, 1157(2), 022054. <https://doi.org/10.1088/1742-6596/1157/2/022054>
- Suwarma, I., & Apriyani, S. (2022). Explore Teachers' Skills in Developing Lesson Plan and Assessment That Oriented on Higher Order Thinking Skills (HOTS). *Journal of Innovation in Educational and Cultural Research*, 3(2), 106–113. <https://doi.org/10.46843/jiecr.v3i2.66>
- Teegaza, E., Abelfattah, F., Al-Saadawi, A., & Al-Turki, O. (2015). Assessment practices of science teachers in middle school based on classification of performance on international examinations. *Journal of Educational and Psychological Studies*, 9(1), 160–178. <https://doi.org/10.53543/jeps.vol9iss1pp160-178>
- Teig, N. (2020, May 26). *Scientific inquiry in TIMSS and PISA 2015: Inquiry as an instructional approach and the assessment of inquiry as an instructional outcome in science*. OSF. <https://doi.org/10.31237/osf.io/5xutk>
- Tesorio, P. J. C., & Canizares, M. J. F. (2018). Philippine K-12 secondary science teachers' assessment conceptions in promoting student learning. *The Normal Lights*, 12(2), 145–167. <https://doi.org/10.56278/tnl.v12i2.1002>
- Triwahyuni, N., Prasetyorini, P., & Rubini, B. (2025). Development of a training model to improve science teachers' competence in designing science literacy assessments for junior high schools. *Journal of Science Education and Practice*, 9(2), 128–143. <https://doi.org/10.33751/jsep.v9i2.13017>
- Van Staden, S. (2026). From system-level usefulness to practice-level transformation: A sociological perspective on international large-scale assessments. *Frontiers in Education*, 11, 1844778. <https://doi.org/10.3389/educ.2026.1844778>
- Walag, A. M. P., Fajardo, M. T. M., Bacarrisas, P. G., & Guimary, F. M. (2022). A canonical correlation analysis of Filipino science teachers' scientific literacy and science teaching efficacy. *International Journal of Instruction*, 15(3), 249–266. <https://doi.org/10.29333/iji.2022.15314a>
- Watson, G., Liang, G., & Sottile, J. (2020). Implementing a large-scale curriculum: Educators' self-assessments and beliefs. *Journal of School Administration Research and Development*, 5(2), 52–62. <https://doi.org/10.32674/jsard.v5i2.2470>
- Yang, Y., Liu, X., & Gardella, J. A., Jr. (2020). Effects of a professional development program on science teacher knowledge and practice, and student understanding of interdisciplinary science concepts. *Journal of Research in Science Teaching*, 57(7), 1028–1057. <https://doi.org/10.1002/tea.21620>
- You, H., Park, S., Hong, M., & Warren, A. (2025). Unveiling effectiveness: A meta-analysis of professional development programs in science education. *Journal of Research in Science Teaching*, 62(4), 971–1005. <https://doi.org/10.1002/tea.21985>