American Journal of Multidisciplinary Research and Innovation (AJMRI)

ISSN: 2158-8155 (ONLINE), 2832-4854 (PRINT)

VOLUME 5 ISSUE 5 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Multidisciplinary Research and Innovation (AJMRI)

Volume 5 Issue 5, Year 2026
ISSN: 2158-8155 (Online), 2832-4854 (Print)
DOI: https://doi.org/10.54536/ajmri.v5i5.8599
https://journals.e-palli.com/home/index.php/ajmri

The Application of SOLO Taxonomy in Secondary Science Education: A Scoping Review

Neil Bryant I. Lucas[1*], Lilibeth G. Abrogena[2], Jayferson G. Panilo[3]

Article Information

Received: July 30, 2026

Accepted: September 12, 2026

Published: September 24, 2026

Keywords

Cognitive progression, Conceptual Understanding, SOLO Taxonomy, Secondary Science Education, Scoping Review

ABSTRACT

The SOLO (Structure of Observed Learning Outcomes) Taxonomy is a widely used framework for characterizing the structural complexity of student understanding, yet no systematic mapping of its application in secondary science education currently exists. This scoping review aimed to map and synthesize the literature on SOLO's application in secondary science education: where it has been studied, for what purposes, and with what reported cognitive and conceptual impacts. Following PRISMA-ScR guidelines and a Population-Concept-Context (PCC) eligibility framework, this review searched ERIC, ScienceDirect, Web of Science, and PsycINFO, yielding 2,826 unique records for empirical studies involving secondary students (Grades 7–12) who used SOLO in science education. Twenty-nine studies met the inclusion criteria and were charted and synthesized narratively and thematically. The evidence base was geographically and disciplinarily concentrated: most studies originated from Asia (48.3%), Europe (24.1%), and Oceania (20.7%), with one study each from North and South America and none from Africa. Some disciplines' evidence bases were concentrated within a small number of countries rather than independent replication. SOLO was used most often to evaluate pedagogical interventions' impact on cognitive-level gains, followed by qualitative coding, instrument development, formative/summative assessment, and curriculum alignment. Targeted interventions were consistently associated with reported gains in SOLO-classified reasoning, though progression frequently plateaued at the relational level, with the extended-abstract level rarely reached, a recurring "relational ceiling." Some studies reported persistently lower-order reasoning despite instruction, and found curriculum documents were not always structured to support higher-order progression. SOLO-based practices are associated with genuine, though bounded, gains in cognitive progression, particularly when activities emphasize conceptual integration; however, this synthesis reflects what study authors reported rather than an independent appraisal of study quality. Given this concentration, findings characterize the contexts where SOLO has been most studied, rather than establishing its general applicability across secondary science education.

INTRODUCTION

Rationale

Secondary science education is still struggling to find frameworks that facilitate deeper conceptual understanding, higher-level thinking, and better alignment among curriculum, instruction, and assessment. In this regard, the Structure of Observed Learning Outcomes (SOLO) Taxonomy has emerged as a powerful frame to describe the complexity of student learning at progressively more integrated levels of learning: from basic and surface-level responses to relational and extended-abstract learning (Arceo, 2024; Davies & Mansour, 2022). SOLO is not only about the number of correct answers, but is also about the quality, structure, and integration of students' responses, and it is especially relevant for science education, where understanding is closely tied to how students are able to generalize knowledge, not only recall facts (Arceo, 2024).

Based on the literature, SOLO has been applied for teaching and research purposes in many international secondary schools. The taxonomy has also been used to assess the effectiveness of instructional scaffolds - for example, haptic feedback (Minogue & Jones, 2009), digital storytelling (Chubko, Morris, McKinnon, Slater, & Lummis, 2019), and virtual laboratories (Padiotis & Mikropoulos, 2010) - in guiding students through abstract scientific concepts, and it has been used to frame "productive questions" leading students from unistructural data retrieval to critical analysis (Lake, 1999). This represents a broad spectrum of educational research in which clear, well-structured competency and assessment frameworks are typically used to design objectives, teaching tools, and assessment tools across cultures, rather than being confined to a single intervention model.

Although there is a substantial body of evidence supporting this line of thinking, the broader research on SOLO in science education is mixed. In one study, clear correlations were reported between SOLO level, cognitive learning, and student achievement, with improvement in performance evident from early in the intervention (Arceo, 2024). At the same time, a curriculum analysis of science learning outcomes found that outcomes were concentrated mostly in the multistructural and relational

[1] Graduate School Student, Mariano Marcos State University-Graduate School, Laoag City, Philippines
[2] Associate Professor, Mariano Marcos State University-Graduate School, Laoag City, Philippines
[3] Assistant Professor, Mariano Marcos State University-Graduate School, Laoag City, Philippines
[*] Corresponding author's e-mail: mmsugsneilbryantlucas@gmail.com

range, that extended-abstract outcomes were not well developed at higher grade levels, and that progress in cognitive complexity was not consistently maintained across the curriculum (Çabuk & Özyürek, 2026). This pattern recurs at the level of individual studies: even under a targeted inquiry-based intervention, only 1.14% of students reached the extended-abstract level by post-test (Damopolii, Nunaki, Nusantari, & Kandowangko, 2020). Longitudinal and cross-sectional work similarly shows that students do not necessarily exhibit significant shifts in operational scientific skills across the secondary years, suggesting that school science may not always provide the cognitive support necessary to apply these skills in real-world contexts (Soobard, Rannikmäe, & Reiska, 2015). Taken together, these findings suggest both instructional potential and longstanding structural gaps in the way higher-order SOLO levels are supported and realized in science education.

Despite the growing body of work, no systematic review has identified how the SOLO Taxonomy has been applied in secondary science disciplines, geographic locations, and implementation purposes, or synthesized what this literature collectively tells us about how students do and do not progress to higher-order reasoning. Indeed, without such a synthesis, it is unclear whether the patterns reported to date represent genuine limits on SOLO-based practice, or rather reflect the concentration of the field within a few disciplines, countries, and research groups in the present day.

This difference has implications for the type of questions the review is intended to address. A systematic effectiveness review asks a comparative, outcome-focused question: how effective is SOLO-based instruction compared to other approaches? Answering that question well requires broad agreement on the intervention studies included, a common outcome metric for SOLO, and formal assessment of methodological quality and bias—none of which can be established with confidence, given the inconsistent application of SOLO across curriculum analysis, classroom intervention, instrument development, and diagnostic evaluation in the literature to date. A scoping review asks a mapping question: how is the use of SOLO in secondary science education? This is exactly the kind of question the current evidence base is well suited to answer, as the literature has its own scope: curriculum analysis, classroom instruction, large-scale assessment, instrument development, teacher-facing applications, and not only a set of comparable interventions that can be pooled or ranked by effect size. Framed this way, a scoping review is not a lesser or preliminary substitute for an effectiveness review — it is the appropriate design for the question this literature is currently capable of answering. Before it is possible to ask how effective SOLO-based instruction is relative to alternatives, it is necessary to establish where and how SOLO has actually been deployed, the objectives for which it has been used, the outcomes that have been reported, and the conceptual and empirical gaps that remain, including underexplored areas such as its application to affective domains. These mapping questions call for evidence from diverse study designs, educational purposes, implementation contexts, and outcome types, rather than for comparing a homogeneous set of interventions. A scoping review is accordingly the design best suited to surveying where and how the SOLO Taxonomy is used, mapping the full extent of the available evidence, identifying underrepresented and persistent gaps, and establishing the boundaries of the field - work that must precede, rather than substitute for, any future systematic effectiveness review.

Objectives

This scoping review aimed to map and synthesize existing literature on the application of the Structure of Observed Learning Outcomes (SOLO) Taxonomy in secondary science education.

Specifically, the study sought to (1) characterize the geographic and disciplinary distribution of studies applying the SOLO Taxonomy; (2) determine the primary purposes of implementing the SOLO Taxonomy; and (3) describe the impacts of SOLO-based practices on students' cognitive progression and conceptual understanding in science.

MATERIALS AND METHODS

The methodology for this scoping review was guided by the Preferred Reporting Items for Systematic reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) (Tricco *et al.*, 2018). The review was organized to find, select, and synthesize relevant literature.

Eligibility Criteria

Studies were selected for inclusion according to the Population, Concept, and Context (PCC) methodology.

Population: Studies of secondary school student, typically Grades 7–12, were included in the review. Studies of primary education or undergraduate university students was excluded unless providing insights in relation to secondary education or on primary education context.

Concepts: The key concept of this review is the application of the SOLO taxonomy. To be eligible, a study must have used the SOLO framework as a fundamental tool for instructional design, assessment, curriculum development, or as a theoretical lens for student learning outcomes.

Context: The context is science education. That covers basic science disciplines like Biology, Chemistry, Physics, Earth Science, and integrated science courses or subjects.

The following are the inclusion and exclusion criteria for screening

Inclusion Criteria

This review included scholarly publications and academic theses or dissertations containing empirical data, including journal articles, conference papers, and doctoral dissertations or theses, published from 1982 (the introduction of the SOLO taxonomy) to July 30, 2026. Eligible studies were published in the English language or had available English translations, and were empirical in nature, employing qualitative, quantitative, or mixed-methods designs, providing data on the application or impact of the SOLO taxonomy. Studies also had to be conducted within the science education context, specifically biology, chemistry, physics, earth science, environmental science, life science, integrated science, or natural science courses or subjects, and had to involve secondary-level students (Grades 7–12). Studies that also included students below Grade 7 or above Grade 12 were eligible provided that secondary-level participants (Grades 7–12) were included in the study.

Exclusion Criteria

Studies were excluded if they mentioned the SOLO taxonomy only in passing, without using it as a central analytical or instructional framework, or if they consisted of editorials, commentaries, or book reviews. Studies conducted outside the science education context, such as those situated in mathematics, history, language arts, or other non-science disciplines, were also excluded. Finally, studies that were not in the target grade range (Grades 7-12), such as preschool, elementary/primary (below Grade 7), tertiary/higher education, or adult learners were excluded unless also including Grades 7-12.

Information Sources

A systematic search was carried out on the electronic databases ERIC, Science Direct, Web of Science, and PsycINFO, which were selected for their coverage of the education and cognitive/developmental science literature relevant to this review. ERIC and Web of Science were selected as core databases for education research as they cover science education and curriculum studies in depth, and Science Direct for its high level of coverage of discipline-specific science education journals. PsycINFO was included to capture SOLO-related research published in cognitive development and educational psychology journals since SOLO's theoretical origins in Piagetian cognitive-developmental psychology mean that relevant research in those fields is not always indexed in education-specific databases. Scopus was not included in the search, as institutional access to this database was not available to the screening team; its multidisciplinary coverage overlaps substantially with Web of Science and Science Direct, which were expected to offset this gap to some extent. The most recent search across all databases was executed on July 30, 2026.

To ensure a minimum standard of academic rigor, this scoping review was restricted to scholarly publications and academic theses or dissertations that reported empirical data on the application of the SOLO Taxonomy, including journal articles, conference papers, and doctoral dissertations or theses. Consequently, non-scholarly and non-empirical materials, such as informal classroom lesson plans, blog posts, institutional reports, editorials, and other forms of grey literature, were excluded to maintain the rigor, credibility, and quality of the evidence base. This boundary was also established to ensure that both screening teams conducting the title-and-abstract and full-text reviews could evaluate the retrieved records thoroughly and consistently.

Covidence, a systematic review management platform, was used to remove duplicate records, facilitate title-and-abstract and full-text screening, and manage the retrieval and inclusion of eligible studies.

Search Strategy

The search strategy was developed together with a research librarian. It was a combination of keywords, controlled vocabulary terms and Boolean operators. The search string used was: (SOLO Taxonomy OR Structure of Observed Learning Outcomes) AND ('science') AND (middle school OR high school OR secondary school).

A supplementary search strategy was also employed to identify relevant studies that the primary search string failed to capture, particularly those indexed under discipline-specific science subject terms or alternative regional terminology for secondary education. This supplementary strategy was necessary because indexing practices vary across databases, and a single search string constructed around general terms may not retrieve studies that classify their subject matter under more specific descriptors (e.g., "biology" or "chemistry" rather than "science") or under population terms common in other educational systems (e.g., "adolescent" or "lower secondary"). The supplementary search string used was:

("SOLO Taxonomy" OR "Structure of Observed Learning Outcomes") AND (science OR "science education" OR biology OR chemistry OR physics OR "earth science" OR "environmental science" OR "life science" OR "integrated science" OR "natural science") AND ("secondary school" OR "secondary education" OR "middle school" OR "high school" OR adolescent* OR "lower secondary" OR "upper secondary").

This supplementary string was applied across the same databases as the primary search to maximize sensitivity and reduce the likelihood of omitting eligible studies due to indexing inconsistencies. The search terms were refined through an iterative process to ensure high sensitivity, and search strategies were adapted for each database to accommodate differences in indexing systems and controlled vocabularies.

Selection of Sources

The study selection process was conducted in two main phases

Level 1 (Title and Abstract Screening): All identified citations were uploaded to Covidence. Initially, 3,044 records were identified from the selected databases. Before screening, 218 duplicate records were identified and removed using the platform's automatic duplicate detection tool, leaving 2,826 unique records for screening. This screening process is depicted in Figure 1.

Prior to full screening, the reviewers independently screened a pilot set of records to establish consistency in application of the eligibility criteria. Two reviewers then independently screened the titles and abstracts of all 2,826 records against the pre-defined inclusion and exclusion criteria. Records were categorized as "yes" or "include," "no" or "exclude," or "maybe." Disagreements were resolved through discussion and consensus, with a third member of the research team consulted when consensus could not be reached. Following this independent dual screening of the full record set, 2,788 records were excluded and 38 records were identified as potentially eligible and sought for full-text retrieval, of which 2 could not be retrieved.

Level 2 (Full-Text Screening): The full-text versions of all articles that passed Level 1 (including "maybe" records) were retrieved and independently reviewed by the same two reviewers against the eligibility criteria. Of the 36 reports assessed for eligibility, 7 were excluded: 5 for ineligible educational level and 2 for ineligible language. This resulted in 29 studies included in the review.

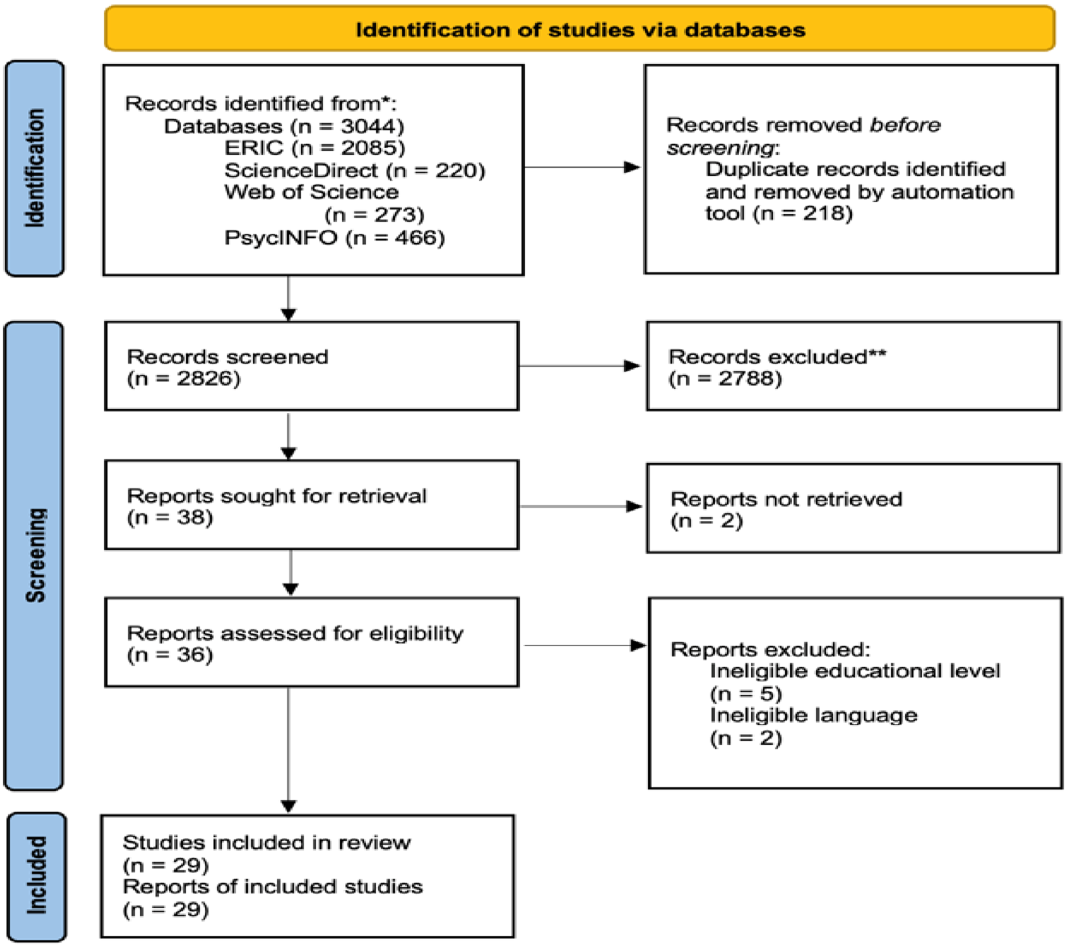

Figure 1: Articles included and excluded at each stage of the screening process, according to the PRISMA framework.

Data Charting Process

A data charting form was developed in a spreadsheet to systematically extract information from the included studies. The form was first used on a sample of five studies for consistency between reviewers. One reviewer carried out the initial data extraction and was verified by the other for accuracy and completeness.

Data Items

The following information was extracted from the included studies in a structured data charting form: study identification number, author(s), year of publication, title, country, and publication type. The characteristics of general study were the study aim, research design, participants, grade level, the sample size, the science discipline and educational context.

To address the review objectives, data were extracted on the implementation of the Structure of Observed Learning Outcomes (SOLO) Taxonomy, including how it was applied (e.g., lesson design, assessment, rubric construction, analysis of student responses, curriculum alignment, or examination classification), its primary purpose, and, where applicable, the duration of implementation. Information on the outcomes that were recorded and key results were collected as well as regarding the impact of the SOLO Taxonomy on students' cognitive development, conceptual understanding, academic achievement, and educational outcomes (e.g., critical thinking, problem-solving, scientific reasoning, and instructional effectiveness). This study also highlighted the research gaps identified by the authors and limitations and reviewer notes related to the screening and synthesis process.

The final inclusion decision was also recorded for each full-text article to keep it as open as possible during the selection process. The studies excluded during full text screening were given a specific exclusion reason based on the criteria.

Synthesis of Results

The results were synthesized by narrative and thematic analysis, and summary tables and charts were used. The synthesis revolved around the three research questions: implementation patterns (geographic and disciplinary), purposes of implementation, and impact analysis.

A thematic analysis of the studies was conducted to identify some common themes across the studies for implementation and application of the SOLO Taxonomy in secondary science education.

Descriptive statistics were used to report the frequency of study types, geographic and disciplinary distribution, and purposes of SOLO Taxonomy implementation. The synthesis concluded with a summary of the implications for science education practice and future research.

RESULTS AND DISCUSSION

Results

Study characteristics

The study analyzed 29 empirical studies. 25 of the 29 included studies (86.2%) were published in academic journals. The remaining 4 comprised two conference papers (Damopolii *et al.*, 2020; Soobard *et al.*, 2015), one pre-published manuscript reporting a quantitative pilot-test analysis of an assessment instrument (Panizzon *et al.*, 2006), and one doctoral dissertation (Lebo *et al.*, 2013).

The results were organized according to the study objectives to address each objective systematically.

Objective 1: Characterize the geographic and disciplinary distribution of studies applying the SOLO Taxonomy.

Geographic Distribution

As shown in Figure 2, the included studies originate from 13 countries across five continental regions. Asia (k = 14, 48.3%) and Europe (k = 7, 24.1%) together account for over seven in ten studies, followed by Oceania (k = 6, 20.7%); North and South America are represented by a single study each (3.4% each). Within Asia, Southeast Asia, and Indonesia in particular (k = 5), is the single most productive setting, followed by East Asia (China, k = 4), Türkiye (k = 2), and single studies from the Philippines, Singapore, and Kyrgyzstan. Within Europe, Sweden (k = 3) and Estonia (k = 3) account for the majority of studies,

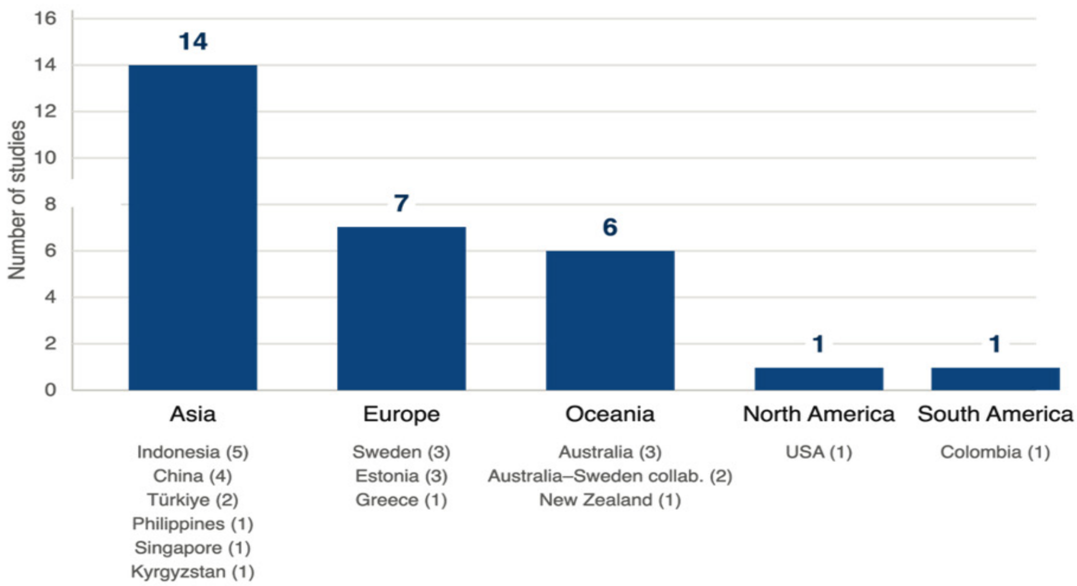

Figure 2: Included studies by geographic region (n = 29)

with a single study from Greece rounding out the region. Oceania is predominantly represented by Australia (k = 3 solely-Australian studies plus 2 Australia–Sweden collaborative studies, accounting for 5 of the region's 6 studies), with one study from New Zealand. No studies from Africa were identified in this dataset.

Disciplinary Distribution

Figure 3 illustrates the distribution of the included studies by science discipline across geographic regions. SOLO has been implemented across every major secondary science subject. Biology is the most common discipline (k = 8), followed by Chemistry (k = 7) and General/

Integrated Science (k = 7), which includes curriculum-wide, environmental, and scientific-literacy applications. Earth Science (k = 4), treating astronomy and cosmology as sub-domains, consistent with the categorization used throughout this review, and Physics/Technology (k = 3) represent smaller but still meaningfully represented

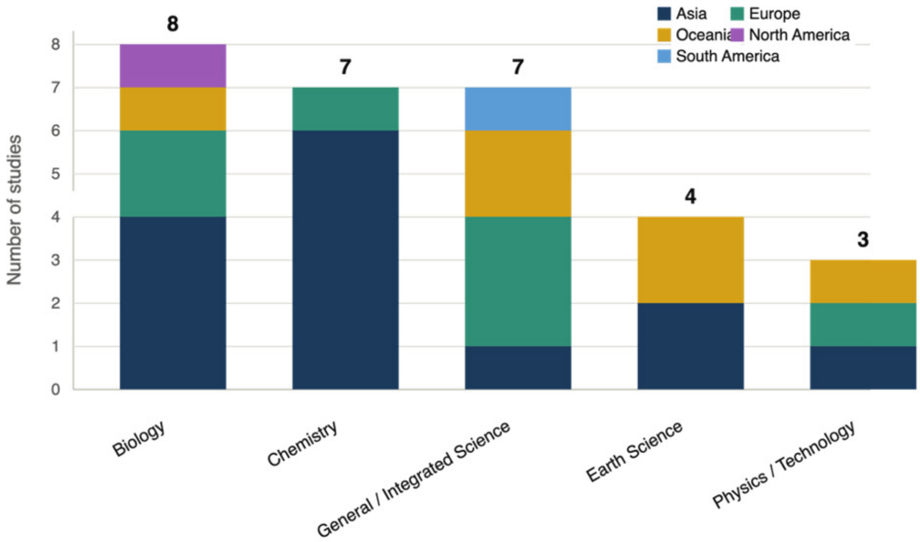

Figure 3: Science discipline by geographic region (n = 29)

Table 1: Primary purposes of SOLO Taxonomy implementation in secondary science education (n = 29)

#	Purpose category	k	%	Illustrative studies
1	Evaluate the impact of a pedagogical intervention on cognitive-level gains (pre-/post-test SOLO coding)	9	31.0%	Damopolii *et al.*, 2020; Damopolii *et al.*, 2021; Fägerstam & Blom, 2013; Minogue & Jones, 2009; Mahanani *et al.*, 2020; Padiotis & Mikropoulos, 2010; Guzman-Lenis *et al.*, 2025; Lebo *et al.*, 2013; Sudaria & Sanchez, 2024
2	Serve as an analytical/ coding framework for qualitative data (interviews, argumentation, discourse)	5	17.2%	Luo *et al.*, 2020; Salimpour *et al.*, 2023; Rundgren *et al.*, 2012; Toh *et al.*, 1999; Eskilsson, 2008
3	Develop or validate a new assessment instrument (often with Rasch modelling)	4	13.8%	Chi *et al.*, 2024; Salimpour *et al.*, 2022; Fan *et al.*, 2025; Panizzon *et al.*, 2006
4	Assess student learning in formative/summative contexts	4	13.8%	Tian *et al.*, 2024; Chubko *et al.*, 2019; Soobard *et al.*, 2015; Soobard & Rannikmäe, 2015
5	Analyze or align curriculum documents/intended learning outcomes	3	10.3%	Çabuk & Özyürek, 2026; Kışoğlu, 2025; Kõlamets *et al.*, 2023
6	Guide instructional design/ pedagogical planning	2	6.9%	Siregar *et al.*, 2025; Lake, 1999
7	Diagnose student misconceptions/learning difficulties	1	3.4%	Haka *et al.*, 2025
8	Serve as a theoretical lens for interpreting alternative conceptions	1	3.4%	Jones *et al.*, 1993
	Total	**29**	**100%**	

categories, with Physics/Technology comprising studies addressing optics, vision, and technology-mediated instruction. Three studies span multiple disciplines within a single curriculum or literacy assessment (Fägerstam & Blom, 2013; Kõlamets *et al.*, 2023; Soobard *et al.*, 2015) and were categorized here according to their predominant disciplinary focus.

The distribution of disciplines also varies by region. Chemistry applications are predominantly concentrated in China (k = 4 of 7 Chemistry studies), with the remainder distributed across Indonesia and Sweden. Earth Science research is similarly concentrated, with half of the category's studies (k = 2 of 4) originating from the Australia–Sweden research collaboration examining cosmological reasoning. Curriculum- and literacy-oriented General/Integrated Science studies span four regions, with the clearest concentration in Europe via Estonia (k = 3); a single study from Asia (Türkiye), single studies from each of two Oceania countries (New Zealand and Australia), and a single study from South America (Colombia) round out the remaining k = 4. Biology remains the most geographically dispersed discipline, appearing across Asia (Indonesia, Türkiye), Europe (Sweden), Oceania (Australia), and North America (USA).

Objective 2: Determine the primary purposes of implementing the SOLO Taxonomy.

Purpose of Implementation

Table 1 describes the main purposes for which the SOLO Taxonomy was used across the 29 included studies. The most common (k = 9, 31.0%) purpose was to evaluate the impact of a pedagogical intervention (i.e., inquiry-based, outdoor learning, technology-enhanced, or scaffolded instruction) on the cognitive growth of students through pre- and post-test coding. Five studies (17.2%) used SOLO as an analytical or coding framework for qualitative data, such as interview transcripts, argumentation, or classroom discussion, as opposed to numerical test scoring. Four studies (13.8%) used SOLO to develop or validate assessment instruments, often in conjunction with Rasch modelling. A further four studies (13.8%) employed SOLO to assess student learning in formative or summative contexts, such as evaluating written responses, classroom assessments, or performance tasks across grade levels. Three studies (10.3%) applied SOLO to analyze or align curriculum documents and intended learning outcomes. Two studies (6.9%) used SOLO primarily as an instructional design or pedagogical planning tool, structuring lessons, worksheets, or learning modules according to increasing cognitive complexity. One study (3.4%) used SOLO for diagnostic assessment to identify students' misconceptions or learning difficulties, and one further study (3.4%) used SOLO as a ground-up theoretical lens for understanding students' alternative conceptions.

Objective 3: Describe the impacts of SOLO-based practices on students' cognitive progression and conceptual understanding in science.

Impacts of SOLO-based practices on students' cognitive progression and conceptual understanding

Theme 1: Marked cognitive-level advancement through targeted intervention

The most significant, and perhaps the most consistent, effect in the data was a sustained improvement in SOLO levels in the context of an explicitly designed pedagogical intervention. Damopolii *et al.* (2020) reported a statistically significant improvement in students' thinking-skill levels after an inquiry-based learning intervention ($p = 0.000$), with the proportion of students at pre-structural and unistructural levels falling sharply while relational-level responses rose to 25%. Similarly, Damopolii *et al.* (2021) found that inquiry-based instruction produced markedly greater cognitive-level gains than conventional teaching (a 96.67% increase versus 42.42%), and Padiotis and Mikropoulos (2010) documented that the majority of students moved to higher SOLO levels after interacting with an educational virtual environment. Technology-mediated interventions also produced measurable gains: Minogue and Jones (2009) found that haptic (touch-based) feedback significantly improved students' progression to relational and extended-abstract levels compared to a visual-only condition ($p = 0.043$), while Sudaria and Sanchez (2024) reported a large effect size (Cohen's $d = 3.402$) for a formative-assessment intervention on earthquake concepts. Comparable, if more moderate, gains were reported by Chubko *et al.* (2019), whose integrated astronomy task for EFL (English-as-a-foreign-language) students produced a moderate effect ($d = 0.41$), and by Luo *et al.* (2020), whose reasoning-flow-scaffold instruction significantly improved students' evidence-use and rebuttal skills in argumentation. Guzman-Lenis *et al.* (2025) and Eskilsson (2008) similarly found sustained movement toward relational-level reasoning, the latter documenting long-term development in the structural quality of students' science talk over repeated instructional cycles. It is worth noting that Chubko *et al.* (2019), Luo *et al.* (2020), and Eskilsson (2008) are classified in Table 1 under their primary methodological purpose - formative/summative assessment and qualitative coding, respectively - rather than under the pedagogical-intervention-impact category; their inclusion here reflects that each study also involved a structured instructional activity whose association with cognitive-level gains is relevant to this theme, even though intervention evaluation was not the primary purpose for which SOLO was applied in the original study.

Theme 2: A recurring relational-level ceiling in the reviewed evidence

A second key pattern was that students showed genuine but bounded cognitive progression. Across intervention studies, many students advanced from prestructural or unistructural thinking to multistructural and relational levels, whereas progression to the extended-abstract level was comparatively uncommon. This ceiling was observed in Chemistry (Tian *et al.*, 2024), where students could move from single-point structures to multiple-point structures relatively easily, but average score ratios declined significantly at higher relational sub-levels (from 80% at unistructural to 28.5% at the most complex relational sub-level), with integration and application of knowledge the hardest part. Haka *et al.* (2025) also found a "breakdown" towards the extended abstract level, even though most students achieved multistructural and relational levels. Rundgren *et al.* (2012) found that conceptual depth and terminology use increased with educational level but that protein-function reasoning from upper-secondary students was still more multistructural and partially relational, and Lebo *et al.* (2013) found systems and sustainability thinking to be evolving and very variable among students. Siregar *et al.* (2025) also reported a qualitative shift towards relational reasoning, but illustrated this through practicality and alignment ratings rather than pre-/post-test effect sizes, tempering the claim.

Theme 3: Persistent lower-order reasoning and unresolved misconceptions

A substantial subset of studies found little or no meaningful cognitive progression, with students' reasoning remaining concentrated at lower SOLO levels despite instruction. Soobard and Rannikmäe (2015) found no significant progress in most items between Grade 10 and Grade 11 cohorts, with only one extended-abstract item showing a modest effect (Cohen's d = 0.365). Soobard *et al.* (2015) likewise reported minimal change in interdisciplinary scientific-literacy skills, with older students only marginally outperforming younger peers. Salimpour *et al.* (2023) found that most student reasoning about cosmology clustered in the pre-structural-to-multistructural region, with a strong presence of pre-structural responses. Fägerstam and Blom (2013) found no significant immediate difference in SOLO levels between indoor and outdoor biology instruction, with both groups remaining around a multistructural level - though outdoor learners showed better long-term retention (discussed further under Theme 5). Toh *et al.* (1999) documented a persistent failure among students to connect the concepts of light, eye, and object in explanations of vision, with over half of respondents not fully understanding the underlying mechanism. Jones *et al.* (1993), reanalyzing existing questionnaire data, found that 74% of student responses relied on ikonic (concrete, non-abstracted) functioning rather than concrete-symbolic reasoning.

Theme 4: Curriculum-level barriers to progression

Several studies shifted the unit of analysis from individual students to the curriculum itself, and consistently found that opportunities for cognitive progression were unevenly or inconsistently built into official learning outcomes. Çabuk and Özyürek (2026) found inconsistencies in the 2024 revised Turkish science curriculum, with some higher grades showing lower cognitive-level outcomes than earlier grades, undermining the intended gradual progression. Kışoğlu (2025) found that the majority of learning outcomes in the high school biology curriculum sat at the relational (30.77%) and extended-abstract (47.44%) levels, with far fewer at unistructural (8.97%) and multistructural (12.82%) levels. Kõlamets *et al.* (2023) found that although energy-related concepts were well represented in the curriculum (35% of learning outcomes), coverage was unevenly distributed across subjects - with Physics contributing the largest share (45.7%) - which may compromise coherent, integrated conceptualization despite high nominal coverage.

Theme 5: Conceptual integration over accumulation - and a durability effect

Where cognitive development did take place, the qualitative texture of the change was less about accumulating more facts and more about students learning to connect previously isolated ideas into coherent, relational explanations. Damopolii *et al.* (2020) characterized this as a shift from factual recall toward correlating relationships and building new knowledge. Minogue and Jones (2009) found that students were able to build more integrated understandings of structure–function relationships in cell biology with tactile feedback. Fägerstam and Blom (2013) also divorced immediate change in SOLO level from long-term retention: outdoor and indoor groups did not differ in SOLO level immediately after instruction, but outdoor learners retained significantly more of the activity and content at long-term follow-up, suggesting that contextualized, hands-on learning can support durable conceptual understanding even when it does not produce an immediate improvement in cognitive level.

Theme 6: Secondary and enabling impacts

Beyond the direct cognitive or conceptual change, a smaller number of studies reported impacts that could plausibly mediate or accompany SOLO-connected learning gains. Siregar *et al.* (2025) reported high practicality ratings from both teachers and students, along with a high alignment rating from teachers, alongside the reported shift toward relational reasoning. Panizzon *et al.* (2006) and Mahanani *et al.* (2020) both highlight the role of instructional support - assessment design that incorporated identification of student difficulties in the former (a secondary feature of an instrument-development study, distinct from this review's diagnostic-assessment category), and inquiry-based scaffolding through a socioscientific-issues context in the latter - suggesting that cognitive-level gains are shaped by instructional design choices and not solely by internal student cognition. The two technology-mediated interventions captured in this review - haptic feedback (Minogue & Jones, 2009) and virtual-environment instruction (Padiotis & Mikropoulos, 2010) - were both associated with measurable gains; however, with only two such studies in the dataset, this review offers limited basis for judging how consistently technology-mediated approaches produce these gains.

Discussion

Geographic and Disciplinary Concentration of SOLO Research

A Regionally Uneven Evidence Base: Dominance of Asia and Europe

The included studies were concentrated in three regions - Asia (k = 14, 48.3%), Europe (k = 7, 24.1%), and Oceania (k = 6, 20.7%) - which together account for over 93% of the evidence base, while North and South America are each represented by only a single study (Minogue & Jones, 2009, USA; Guzman-Lenis *et al.*, 2025, Colombia) and no eligible studies were identified from Africa. Within Asia, Indonesia (k = 5) and China (k = 4) contributed the largest number of studies, followed by Türkiye (k = 2) and single studies from the Philippines, Singapore, and Kyrgyzstan. Within Europe, Sweden (k = 3) and Estonia (k = 3) were the leading contributors, with a single study from Greece. Oceania was virtually dominated by Australia, either independently (k = 3 only Australian studies) or jointly with Sweden on cosmology-oriented studies (k = 2 studies), and a single study from New Zealand filled in the region. This suggests that current understanding of SOLO implementation is from a relatively small number of countries among these three, and that a more diverse evidence base would be needed to assess whether the observed implementation patterns and reported results translate well to other educational contexts in the future.

The near-total absence of the Americas and the complete absence of studies from Africa is clearly a limitation of the current evidence base and not a side effect. This geographic pattern might be due to differences in research activity, publication visibility, and the dominance of English-language indexed literature across regions, but this review did not investigate these factors directly. As a consequence, the absence of studies from Africa and the lack of studies from the Americas may reflect the diversity of the literature reviewed in those territories rather than the absence of SOLO-related pedagogical practices in those regions. This distinction is relevant to how the findings of this review should be interpreted: conclusions about the purpose and impact of SOLO's impact on education and education policy are based on what has been published and indexed within the review's search parameters, not on standard educational practice applicable to all educational systems.

Disciplinary Clustering and the Absence of Cross-National Replication

The disciplinary findings reinforce the geographic concentration identified above, though unevenly across disciplines. Biology (k = 8), the most frequently studied discipline, is also the most geographically dispersed, spanning Asia (Indonesia, Türkiye), Europe (Sweden), Oceania (Australia), and North America (USA) - providing some indication that SOLO's applicability has been examined across varied educational contexts for at least this discipline. This is not the case for Chemistry or Earth Science, the review's next most- and least-studied disciplines, respectively. Chemistry (k = 7, tied with General/Integrated Science as the second most frequently examined discipline) is predominantly concentrated in China (k = 4 of 7), with the remainder distributed across Indonesia and Sweden. Earth Science (k = 4) - comprising the astronomy-, cosmology-, and earthquake-focused studies identified in this review — is similarly concentrated, with half of its studies (Salimpour *et al.*, 2022, 2023) originating from a single Australia–Sweden research collaboration examining cosmological reasoning.

This pattern is notable for somewhat different reasons across the two disciplines. Chemistry's concentration within a single country constrains the generalizability of a comparatively substantial portion of the evidence base, given that it accounts for nearly a quarter of all included studies. Earth Science, by contrast, comprises only four studies in total, half of which come from the same collaborative team; in a category this small, that concentration means the discipline's apparent breadth of evidence largely reflects repeated investigation within one research partnership rather than independent replication across national contexts. Physics/Technology (k = 3) presents a different picture: its three studies originate from Australia (Jones *et al.*, 1993), Greece (Padiotis & Mikropoulos, 2010), and Singapore (Toh *et al.*, 1999) - three unrelated national contexts with no shared authorship or research team. This discipline therefore does not show the single-country or single-team concentration seen in Chemistry and Earth Science; its principal limitation is simply the very small number of studies (k = 3), which is too few to draw firm conclusions regardless of their geographic spread.

This low cross-country replication in Chemistry and Earth Science is significant because SOLO's main theoretical claim - that it captures a domain-general structure of increasing response complexity across diverse areas of learning - is an abstract claim that needs to be tested separately in different educational, linguistic, and curricular contexts to be empirically valid. In the absence of replication of Chemistry- or Earth Science-focused SOLO applications outside of their concentrated research settings, it remains uncertain whether the reported implementation patterns and outcomes are transferable beyond those specific contexts. A related, if smaller-scale, concern arises from the review's curriculum-alignment studies (k = 3): two of the three (Çabuk & Özyürek, 2026; Kışoğlu, 2025) originate from Türkiye, with the third (Kõlamets *et al.*, 2023) from Estonia. Because these studies share a common analytic purpose and, in two of three cases, a common national curriculum context, it is difficult to determine whether curriculum-alignment applications of SOLO are genuinely more prevalent in Türkiye specifically, or whether this review's search strategy and source indexing were more likely to identify this cluster of publications than comparable work that may exist elsewhere but was not captured within the parameters of this review.

Taken together, these findings indicate that the geographic and disciplinary distribution identified in this review characterizes the regions and disciplines in which SOLO research has been most visible and productive within the accessible literature, rather than the full extent of contexts in which SOLO may be applicable. This distinction has direct implications for the interpretation of the findings presented in the sections that follow, both of which should be understood as characterizing a research base that remains geographically and disciplinarily concentrated, rather than as a comprehensive account of SOLO's applicability across the broader landscape of secondary science education.

The Predominance of Intervention-Impact and Instrument-Development Purposes

SOLO as an Evaluative Tool: Why Intervention Studies Dominate

The predominance of intervention-impact studies is consistent with the methodological orientation of science education research, where pre-test/post-test and quasi-experimental designs remain the most common approaches for demonstrating the effectiveness of a pedagogical intervention. SOLO's five-level structure provides a systematic, ordinal outcome variable well-suited to such an approach, and this is probably why it is used as an important dependent measure in studies assessing inquiry-based, technology-based, or scaffolded instruction (e.g., Damopolii *et al.*, 2020; Minogue & Jones, 2009; Sudaria & Sanchez, 2024) and others in this category. SOLO is used primarily as a coding and scoring framework for pre-existing instructional designs rather than as a framework that itself generates or informs the designs.

Instrument-development studies (k = 4) can be viewed in a similar light. Several studies pair SOLO with Rasch modelling to construct psychometrically robust, ordinal measures of cognitive complexity (Chi *et al.*, 2024; Salimpour *et al.*, 2022; Fan *et al.*, 2025; Panizzon *et al.*, 2006). This is consistent with the current trend in science education measurement, where taxonomies like SOLO are turned into quantitative instruments suitable for large-scale quantitative analysis. This methodological compatibility between SOLO's ordinal structure and psychometric modelling may explain why instrument development, rather than instructional application, is one of the more frequently represented purposes in this review.

Combining these two, it follows that the most frequently reported application of SOLO in the literature is as an outcome measure: assessing how a student responds to instruction after it has occurred, rather than shaping the instruction itself. This reflects the volume and consistency of evaluative applications within the reviewed studies, rather than an independent confirmation of this application's validity or effectiveness, which this review did not assess. This finding is consistent with SOLO's original formulation by Biggs and Collis (1982) as a taxonomy for classifying the structure of learning outcomes, and suggests that subsequent research has predominantly extended this original evaluative function rather than substantially expanding SOLO's application to earlier stages of the instructional process.

Underused Applications: Instructional Design, Diagnosis, and Theory-Building

By comparison, SOLO was used directly to guide instructional design, diagnose student difficulties before teaching, or provide a theoretical basis for interpreting student reasoning in only a small fraction of the studies included. The instructional design category (k = 2) consisted of studies that used SOLO to structure lessons, worksheets, or learning modules according to increasing cognitive complexity: Siregar *et al.* (2025) and Lake (1999). This is distinct from the evaluation of lesson design described above, since the taxonomy is here used prospectively to guide the timing and sequencing of teaching materials rather than to assess their effectiveness after the fact. The small number of studies of this kind indicates that SOLO has more often been used to evaluate instructional outcomes than to actively guide the design of instruction itself.

The diagnostic assessment category (k = 1) consisted of a single study, Haka *et al.* (2025), which used SOLO to identify student misconceptions or learning difficulties before teaching or independent of any formal intervention. This application differs from the formative and summative assessment uses discussed elsewhere in this review in that its purpose is not to evaluate or measure the extent of learning already achieved, but to inform subsequent classroom decisions before instruction begins. That only one study used SOLO in this way suggests that its diagnostic potential — in contrast to its evaluative application — has been substantially less explored in the secondary science education literature captured by this review.

The only study in the theoretical-lens category, Jones *et al.* (1993), used SOLO as a general framework for understanding the structure of students' alternative conceptions, rather than as part of a specific instructional or assessment instrument. This single instance suggests that SOLO's application as a standalone theoretical framework, rather than as a component embedded within an evaluative or instrument-development study, is uncommon within the reviewed literature.

The relatively low representation of these three categories — together accounting for only four of the 29 included studies — has implications for how SOLO's role in science education should be understood. The findings in this review indicate that SOLO's usefulness in the literature is mostly in detecting and categorizing learning outcomes after instruction while its use to guide instructional design, diagnose learning needs before teaching or serve as a stand-alone explanatory theory is relatively rare. This distinction should be taken into account when interpreting the cognitive-progression and conceptual-understanding findings discussed next as those findings are based on studies that used SOLO as an evaluative tool and not as an instructional method on its own.

Cognitive Progression and Conceptual Understanding: Promise and Persistent Limits

Evidence for SOLO-Aligned Interventions as Drivers of Cognitive Growth

Across the reviewed studies, pedagogical interventions were consistently reported to be associated with measurable improvements in students' SOLO-classified reasoning levels. This was observed across different interventions, including inquiry-based (Damopolii *et al.*, 2020; Damopolii *et al.*, 2021; Mahanani *et al.*, 2020), technology-based (Minogue & Jones, 2009; Padiotis & Mikropoulos, 2010), and structured task-based interventions (Sudaria & Sanchez, 2024; Chubko *et al.*, 2019; Luo *et al.*, 2020). The consistency of this reported pattern across differing instructional formats and disciplinary contexts suggests a shared design feature across these studies: each intervention was explicitly structured to require students to integrate multiple pieces of information or construct relationships between concepts, corresponding to the criteria the SOLO framework uses to distinguish multistructural from relational responses.

This pattern of reported outcomes is consistent with SOLO's theoretical basis as an outcome-based taxonomy, in that instructional tasks explicitly structured to elicit higher-order responses were more frequently associated with reported shifts toward those response categories, compared to instruction not explicitly targeting this structural change. Several of the reviewed studies reported statistically significant results, including comparatively large effect sizes as calculated by the study authors (e.g., Cohen's d = 3.402 in Sudaria & Sanchez, 2024).

Consistent with standard scoping review methodology, this review did not conduct a formal risk-of-bias or quality appraisal of the included studies, and the patterns described above should be read as a synthesis of what was reported by study authors rather than as an independent evaluation of intervention effectiveness. It is nonetheless worth noting, as a descriptive feature of the evidence base rather than a judgment of its validity, that the majority of these studies were described by their authors as employing single-group pre-test/post-test or non-equivalent control-group designs rather than randomized controlled designs.

The "Relational Ceiling": A Recurring Barrier to Extended-Abstract Reasoning

Across multiple studies, student progression through the SOLO hierarchy was frequently reported up to the multistructural or relational level, but rarely extended to the extended-abstract level. This pattern was documented in Chemistry (Tian *et al.*, 2024) and Biology (Haka *et al.*, 2025; Rundgren *et al.*, 2012), suggesting the limitation was not confined to a single discipline among those in which it was examined — though it is worth noting that the Physics/Technology and Earth Science studies in this review did not report this specific pattern, so its generality across disciplines cannot be assumed beyond the contexts in which it was observed. Given that the extended-abstract level requires students to generalize a principle beyond the immediate context in which it was learned, this recurring pattern is consistent with — though not sufficient on its own to establish — a specific point in the reasoning-development process at which student progression may be more difficult to achieve. This interpretation should be treated as a provisional pattern rather than an established developmental finding, since it is drawn from a limited number of studies employing varied instruments, samples, and definitions of extended-abstract performance, rather than from designs capable of directly testing where in the reasoning process this difficulty arises.

Despite these limitations, the recurrence of this pattern across the disciplines and study populations in which it was documented suggests that it is a noteworthy feature of the current evidence base, although it should not be interpreted as confirming a general cognitive-developmental constraint. This has a bearing on how the findings discussed above should be read. While targeted interventions were reported as producing progression from unistructural to relational levels of reasoning, the studies reviewed here reported comparatively less evidence of, or did not evaluate, progression to the extended-abstract level specifically. This raises the possibility, though it does not confirm, that the instructional conditions supporting generalization and abstraction differ from those supporting integration of related concepts — a distinction that the intervention designs represented in this review do not appear to have directly tested.

Curriculum Design as a Structural Constraint on Progression

Limitations in cognitive progression are not necessarily attributable solely to student reasoning or classroom-level instructional design. A small number of document-analysis studies included in this review (k = 3) found that intended learning outcomes within official curriculum frameworks were themselves inconsistently structured to support progression toward higher-order reasoning. Çabuk and Özyürek (2026) and Kışoğlu (2025) each found that the proportion of learning outcomes classified at the extended-abstract level did not increase consistently across grade levels within their respective national curricula, and in some cases decreased at higher grade levels. Kõlamets *et al.* (2023) found a similar structural inconsistency: although energy-related concepts were well represented in the Estonian curriculum overall (35% of learning outcomes), their distribution across subjects was uneven, with Physics contributing the largest share (45.7%) – a fragmentation the authors warned could destroy coherent, integrated conceptualization despite high nominal coverage.

Given that only three studies in this review examined curriculum documents directly, these findings should be read as an indicative pattern rather than a generalizable claim about curriculum design across the reviewed literature as a whole. With that qualification, the findings suggest that the extended-abstract limitation discussed above may not be attributable solely to constraints originating within individual students or classroom-level instruction. If curriculum documents do not consistently specify learning outcomes at the extended-abstract level, particularly at higher grade levels where such reasoning would be developmentally appropriate, the opportunity for students to engage in and be assessed on this form of reasoning may be constrained prior to any classroom-level instructional intervention. This raises the possibility that curriculum design contributes to the patterns of limited progression identified across the intervention-based studies reviewed in this discussion, and that instructional interventions alone may be insufficient to address this limitation where curriculum frameworks do not provide corresponding opportunities for extended-abstract reasoning — though confirming this relationship would require studies that directly examine both curriculum structure and classroom outcomes within the same educational system.

Beyond Test Scores: Motivation, Retention, and Teacher Scaffolding as Enabling Conditions

Several factors reported alongside cognitive-level outcomes appear to function as conditions that support or moderate the progression discussed above, rather than as outcomes independent of it. Fägerstam and Blom (2013) showed that while outdoor and indoor instructional groups did not differ significantly in the level of SOLO-level reasoning immediately after classes were taught, the outdoor group retained much more of the instructional content five months later. This result suggests that immediate SOLO-level classification and longer-term conceptual retention may not be identical, and that instructional methods producing similar immediate cognitive-level classifications may nonetheless differ in how well those gains are maintained.

Additional studies identified teacher-mediated scaffolding as a contributing factor in students' attainment of higher SOLO levels. Mahanani *et al.* (2020) reported that teacher scaffolding during classroom discussion was associated with students reaching higher SOLO levels, and Panizzon *et al.* (2006) similarly noted that information generated by their SOLO-based assessment instrument helped inform teacher practice. These results indicate that cognitive-level progression as measured in the reviewed studies was frequently reported as a function of the interaction between student reasoning and teacher-provided instructional support rather than as an outcome of instructional materials or task design separately.

Moreover, increased motivation and engagement were also found to be associated with cognitive-level progress in several studies (Damopolii *et al.*, 2020; Siregar *et al.*, 2025), but the reviewed studies did not necessarily show a direct or causal relationship between motivational change and cognitive-level enhancement. The two technology-mediated interventions captured in this review — haptic feedback (Minogue & Jones, 2009) and virtual-environment interaction (Padiotis & Mikropoulos, 2010) — were both associated with measurable progression toward higher SOLO levels; however, with only two such studies in the dataset, this review offers limited basis for judging how consistently technology-mediated approaches produce these gains, or which features of their implementation account for the reported effectiveness. This is an area where the small number of included studies constrains the strength of any conclusion more than the direction of the reported findings themselves.

Implications, Limitations, and Future Directions

Theoretical and Pedagogical Implications

The findings demonstrate that the SOLO Taxonomy has been applied consistently across multiple science disciplines, educational systems, and assessment contexts, indicating its broad applicability as a framework for classifying the structural complexity of student responses. At the same time, the findings indicate that SOLO's demonstrated practical value within the reviewed literature has been concentrated in its evaluative function — namely classifying and measuring the outcomes of instruction — rather than in its potential generative function as a tool for designing instruction in advance. The comparatively limited number of studies applying SOLO to instructional design suggests that this application, while theoretically consistent with SOLO's hierarchical structure, has not been extensively tested or reported in the literature captured by this review.

Across the reviewed studies, instructional interventions designed to promote conceptual integration were commonly associated with progression toward relational-level reasoning, whereas progression to the extended-abstract level was reported less consistently. This suggests that practitioners seeking to use SOLO to inform instructional planning should consider not only the design of learning activities but also whether curriculum-level learning outcomes provide corresponding opportunities for extended-abstract reasoning, given the curriculum-level constraints identified earlier in this discussion. Teacher scaffolding and the durability of conceptual understanding over time were also reported as relevant to cognitive-level outcomes in the reviewed studies, suggesting that instructional planning informed by SOLO may benefit from considering these factors in addition to the structure of instructional tasks themselves.

Limitations of the Review

Several limitations should be considered when interpreting the findings of this review. First, the included studies were geographically and disciplinarily concentrated, with most evidence originating from Asia, Europe, and Oceania and only a single study each representing North and South America, and none from Africa. In addition, some science disciplines were represented primarily by a limited number of countries or closely related research groups: Chemistry evidence was concentrated predominantly in China, and Earth Science evidence was concentrated substantially within a single Australia–Sweden research collaboration. Physics/Technology, on the other hand, was based on studies in three different national contexts (Australia, Greece and Singapore), but the overall evidence base (k = 3) was too small to support strong conclusions regardless of what regions were considered. The mapped evidence may not be representative of the total amount of SOLO Taxonomy used in education systems and science fields.

Second, this review is limited by the search strategy and eligibility criteria used. Four databases were searched—ERIC, ScienceDirect, Web of Science, and PsycINFO—covering studies published in English or available in English translation up to July 30, 2026; the work indexed elsewhere, published after that date or available only in other languages may not have been included. Notably, Scopus was not accessible to the screening team and so could not be searched; although its multidisciplinary coverage overlaps substantially with Web of Science and Science Direct, some relevant records indexed uniquely in Scopus may not have been captured by the present review. Additionally, despite efforts to develop a sensitive search string in collaboration with a research librarian, it remains possible that eligible studies using alternative terminology for the SOLO framework or describing science disciplines differently were not captured by the search. To assess this possibility, the reviewers conducted a supplementary search using an alternative string with different discipline- and framework-related terms; this supplementary search did not identify any additional eligible studies beyond those already captured, providing some reassurance regarding the comprehensiveness of the primary search strategy. Relatedly, of the 38 records identified as potentially eligible at the title/abstract stage, two could not be retrieved for full-text review and so could not be assessed for inclusion, though this represents a small proportion of the overall pool. Nonetheless, the possibility of missed studies due to indexing, terminology differences, or retrieval limitations across databases cannot be fully ruled out. The review's restriction to scholarly, empirical publication types (journal articles, conference papers, and doctoral theses or dissertations) — while intended to maintain methodological rigor and was feasible for the two-member core screening team — also means that relevant grey literature, such as institutional reports or non-empirical practitioner accounts of SOLO implementation, was excluded by design and is not reflected in this synthesis.

Finally, as a scoping review conducted in line with the PRISMA-ScR guidance (Tricco *et al.*, 2018), this study was designed to map and synthesize the extent, characteristics, and distribution of the existing literature rather than to critically appraise the methodological quality or risk of bias of the included studies. Accordingly, the findings should be interpreted as a descriptive synthesis of the current evidence base rather than as an assessment of the effectiveness or methodological rigor of SOLO-based practices.

Future Research Directions

The findings of this review identify several priorities for future research. First, more studies in underrepresented areas (especially Africa and Latin America) would be required to verify whether the reported implementation patterns and results of the SOLO Taxonomy could be generalized to other countries and cultural contexts. Redrawing the geographic distribution of studies would be a better tool for promoting a more balanced and representative evidence base.

Second, the SOLO Taxonomy needs to be more widely studied in the disciplines in which it is relevant, and research in the disciplines should focus only on the specific problem that it is addressing. Chemistry and Earth Science evidence, while reasonably sized (Chemistry) and informative though it is small (Earth Science), are concentrated in a small number of countries and in the case of Earth Science a small research team, so there needs to be a replication in other countries to explore whether their patterns apply to other countries. Physics/Technology, in fact, needs to be explored more overall and its evidence is rather small and is obtained from three countries. Greater disciplinary diversity and independent replication across educational settings would enhance our understanding of the use of the SOLO Taxonomy in secondary science education.

Third, the predominance of evaluative applications suggests a need for further research on underexplored uses of the SOLO Taxonomy. Future studies could examine its application in instructional design, curriculum planning and alignment, diagnostic assessment, and teacher professional development, where relatively few studies were identified in the current evidence base.

Finally, the limited progression to the extended-abstract level reported across many intervention studies indicates an important area for further investigation. Longitudinal studies examining how sustained SOLO-informed instruction, curriculum alignment, and classroom implementation influence higher-order reasoning and conceptual understanding would provide valuable insights into students' cognitive development over time.

Funding

This scoping review received no external funding from any agency in the public, commercial, or not-for-profit sectors.

CONCLUSION

This scoping review synthesizes how the SOLO Taxonomy has been applied across secondary science education, mapping where research has occurred, which disciplines it has addressed, and the purposes and outcomes reported. Evidence was concentrated in Asia, Europe, and Oceania, with entire regions, notably Africa, absent. SOLO has been used primarily as an evaluation tool, with comparatively little application to teaching, diagnosis, or theory-building. Included studies reported cognitive-level gains following interventions integrating across concepts, alongside a persistent ceiling at the relational level attributed to curriculum structure and instructional conditions.

Because this review did not appraise methodological quality or evaluate SOLO's effectiveness, these patterns should be read as a summary of what study authors reported, not as evidence of SOLO's validity. The evidence base remains methodologically fragmented, relying largely on single-group or non-equivalent control designs rather than randomized trials, and progression to extended-abstract reasoning remains an unresolved challenge across studies.

Future research should prioritize underrepresented regions and disciplines, independent replication, and longitudinal work on extended-abstract reasoning, including direct evaluations of effectiveness. For practice, positive outcomes were linked to instructional design choices and structured scaffolding at the curriculum level. Addressing these gaps would strengthen the evidence base for SOLO's role in secondary science education.

REFERENCES

Arceo, K. M. M. (2024). Structure of observed learning outcomes (SOLO) taxonomy based teaching on the students' cognitive learning outcome and performance. *International Journal of Research Publications*. https://doi.org/10.47119/ijrp1001491520246512

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the observed learning outcome)*. Academic Press.

Çabuk, C., & Özyürek, C. (2026). An analysis of the science curriculum within the Türkiye century education model based on the SOLO taxonomy. *Pedagogical Perspective, 5*(1), 37–56. https://doi.org/10.29329/pedper.2026.136

Chi, M., Zheng, C., & He, P. (2024). Assessing high school students' chemical thinking using an essential questions-perspective framework. *Chemistry Education Research and Practice, 25*(4), 1143–1158. https://doi.org/10.1039/d4rp00106k

Chubko, N., Morris, J. E., McKinnon, D. H., Slater, E. V., & Lummis, G. W. (2019). SOLO taxonomy as EFL students' disciplinary literacy evaluation tool in technology-enhanced integrated astronomy course. *Language Testing in Asia, 9*(1). https://doi.org/10.1186/s40468-019-0095-6

Damopolii, I., Nunaki, J. H., Nusantari, E., & Kandowangko, N. Y. (2020). The effectiveness of inquiry-based learning to train students' thinking skill based on SOLO taxonomy. *Journal of Physics Conference Series, 1567*(4), 042025. https://doi.org/10.1088/1742-6596/1567/4/042025

Damopolii, I., Nusantari, E., Kristianto, D. D., Nunaki, J. H., & Kandowangko, N. Y. (2021). The effect of inquiry-based science learning on students' level of thinking. *Journal of Physics Conference Series, 1731*(1), 012089. https://doi.org/10.1088/1742-6596/1731/1/012089

Davies, O., & Mansour, N. (2022). Exploring the use of cognitive science approaches alongside SOLO taxonomy as a pedagogical framework to build deeper knowledge in science and foundation subjects at primary schools in UK. *Education Sciences, 12*(8), 1–15. https://doi.org/10.3390/educsci12080523

Eskilsson, O. (2008). The quality of lower secondary students' discussions during labwork in chemistry. *Eurasia Journal of Mathematics Science and Technology Education, 4*(3). https://doi.org/10.12973/ejmste/75346

Fägerstam, E., & Blom, J. (2013). Learning biology and mathematics outdoors: effects and attitudes in a Swedish high school context. *Journal of Adventure Education & Outdoor Learning, 13*(1), 56–75. https://doi.org/10.1080/14729679.2011.647432

Fan, Y., Tian, P., Ma, Y., & Wu, L. (2025). Evaluating students' chemistry quantitative reasoning skills (QRSC) in lower-secondary schools. *International Journal of Science Education, 1*–23. https://doi.org/10.1080/09500693.2025.2540619

Guzman-Lenis, A. R., Perez-Mesa, M. R., & Porras-Contreras, Y. A. (2025). The impact of learning approaches and performance levels on pro-environmental behavior in inquiry contexts. *International Journal of Education in Mathematics Science and Technology, 13*(5), 1064–1082. https://doi.org/10.46328/ijemst.5697

Haka, N. B., Ningsing, D. W., Pratama, A. O. S., Masya, H., & Rakhmawati, I. (2025). Mapping high level thinking abilities using SOLO taxonomy in Class VIII students in middle schools in Bandar Lampung City. *ISEJ Indonesian Science Education Journal, 6*(3), 53–63. https://doi.org/10.62159/isej.v6i3.2087

Jones, B. L., Collis, K. F., & Watson, J. M. (1993). Towards a theoretical basis for students' alternative frameworks in science and for science teaching. *Research in Science Education, 23*(1), 126–135. https://doi.org/10.1007/bf02357053

Kışoğlu, M. (2025). Analysis of the learning outcomes of the Türkiye Century Education Model high school biology course curriculum according to SOLO taxonomy. *International Journal of Educational Research Review, 10*(4), 235–253. https://doi.org/10.24331/ijere.1751884

Kõlamets, L., Kasuk, H., Holbrook, J., & Mamlok-Naaman, R. (2023). The relevance of learning outcomes included in Estonian grade 7-9 science subject curricula associated with the concept of energy. *Journal of Baltic Science Education, 22*(4), 653–667. https://doi.org/10.33225/jbse/23.22.653

Lake, D. (1999). Helping students to go SOLO: teaching critical numeracy in the biological sciences. *Journal of Biological Education, 33*(4), 191–198. https://doi.org/10.1080/00219266.1999.9655664

Lebo, N., Eames, C., Coll, R., & Otrel-Cass, K. (2013). Toward ecological literacy: A permaculture approach to junior secondary science. *University of Waikato.* http://researchcommons.waikato.ac.nz/handle/10289/7392

Luo, X., Wei, B., Shi, M., & Xiao, X. (2020). Exploring the impact of the reasoning flow scaffold (RFS) on students' scientific argumentation: based on the structure of observed learning outcomes (SOLO) taxonomy. *Chemistry Education Research and Practice, 21*(4), 1083–1094. https://doi.org/10.1039/c9rp00269c

Mahanani, I., Rahayu, S., & Fajaroh, F. (2020). The effect of inquiry-based learning with socioscientific issues context on critical thinking skills and scientific explanation. *Jurnal Kependidikan Penelitian Inovasi Pembelajaran, 3*(1), 53–68. https://doi.org/10.21831/jk.v3i1.20972

Minogue, J., & Jones, G. (2009). Measuring the impact of haptic feedback using the SOLO taxonomy. *International Journal of Science Education, 31*(10), 1359–1378. https://doi.org/10.1080/09500690801992862

Padiotis, I., & Mikropoulos, T. A. (2010). Using SOLO to evaluate an educational virtual environment in a technology education setting. *Educational Technology & Society, 13*(3), 233–245. https://www.researchgate.net/publication/220374666_Using_SOLO_to_Evaluate_an_Educational_Virtual_Environment_in_a_Technology_Education_Setting

Panizzon, D., Arthur, D., & Pegg, J. (2006). ESSA: Developing and scoping a test to explore scientific literacy and achievement in NSW. *RUNE (Research UNE).* https://hdl.handle.net/1959.11/4276

Rundgren, C., Hirsch, R., Rundgren, S. C., & Tibell, L. A. E. (2012). Students' communicative resources in relation to their conceptual understanding—The role of non-conventionalized expressions in making sense of visualizations of protein function. *Research in Science Education, 42*(5), 891–913. https://doi.org/10.1007/s11165-011-9229-2

Salimpour, S., Tytler, R., Doig, B., Fitzgerald, M. T., & Eriksson, U. (2022). Conceptualising the cosmos: Development and validation of the Cosmology Concept Inventory for high school. *International Journal of Science and Mathematics Education, 21*(1), 251–275. https://doi.org/10.1007/s10763-022-10252-y

Salimpour, S., Tytler, R., Fitzgerald, M. T., & Eriksson, U. (2023). Is the universe infinite? Characterising a hierarchy of reasoning in student conceptions of cosmology concepts using open-ended surveys. *Journal for STEM Education Research, 6*(1), 102–129. https://doi.org/10.1007/s41979-023-00088-8

Siregar, R., Ubaidah, Huda, Z., Romadon, & Lubis, A. (2025). Practicality of an inquiry-based module on atomic structure and nanomaterials to enhance relational understanding in vocational high school students. *Tadris Jurnal Keguruan Dan Ilmu Tarbiyah, 10*(2), 835–844. https://doi.org/10.24042/tadris.v10i1.28930

Soobard, R., & Rannikmäe, M. (2015). Examining curriculum related progress using a context- based test instrument - a comparison of Estonian grade 10 and 11 students. *Science Education International, 26*(3), 263–283. http://files.eric.ed.gov/fulltext/EJ1074874.pdf

Soobard, R., Rannikmäe, M., & Reiska, P. (2015). Upper secondary schools students' progression in operational scientific skills – A comparison between grades 10 and 12. *Procedia - Social and Behavioral Sciences, 177*, 295–299. https://doi.org/10.1016/j.sbspro.2015.02.342

Sudaria, C. M., & Sanchez, J. M. (2024). SOLO-based formative assessments in teaching and learning earthquakes. *Science Education International, 35*(4), 408–420. https://doi.org/10.33828/sei.v35.i4.11

Tian, P., Fan, Y., Sun, D., & Li, Y. (2024). Evaluating students' computation skills in learning amount of substance based on SOLO taxonomy in secondary schools. *International Journal of Science Education, 46*(15), 1578–1600. https://doi.org/10.1080/09500693.2023.2291691

Toh, K., Boo, H., & Woon, T. (1999). Students' perspectives in understanding light and vision. *Educational Research, 41*(2), 155–162. https://doi.org/10.1080/0013188990410203

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., . . . Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. https://doi.org/10.7326/m18-0850