International Journal of Instruction
e-ISSN: 1308-1470 ● www.e-iji.net

January 2027 ● Vol.20, No.1
p-ISSN: 1694-609X
pp. 521-540

Article submission code: 20260515132738

Received: 15/05/2026
Revision: 28/07/2026

Accepted: 07/08/2026
OnlineFirst: 01/10/2026

Analytical Framework for Cognitive Demands in Physics Assessment Items

Hui Dong
School of Physics and Electronic Engineering, Zhengzhou Normal University, China, *215311638@qq.com*

Ruoshi Wang
Donostia Physics Education Research Group and Applied Physics Department, School of Engineering Gipuzkoa (UPV/EHU), Spain，*Shishi19920106@gmail.com*

Yanli Ding
School of Physics and Electronic Engineering, Zhengzhou Normal University, China, *donghui_2019@qq.com*

Assessment items are increasingly expected to serve not only as tools for evaluating students' achievement, but also as instruments that can actively support cognitive development and learning processes. However, existing research often adopts a fragmented perspective, with some studies focusing primarily on students' cognitive processes during problem solving, while others emphasize the structural characteristics of assessment tasks such as context, representation, or item format. The lack of integration between these two perspectives limits a comprehensive understanding of how assessment items may exert developmental functions.To address this gap, this study develops an analytical framework for examining the cognitive demands and developmental affordances of physics assessment items. Drawing on cognitive psychology, the sociological structure–function paradigm, and reasoning theory, the framework establishes connections between item structure and students' cognitive activities. It consists of three analytical dimensions and a structured three-step analytical procedure designed to support systematic and replicable analysis. The feasibility and practical applicability of the framework are examined through analyses of representative assessment items, inter-rater reliability testing, and illustrative applications using PISA, TIMSS, and AP physics tasks. The results suggest that the proposed framework offers a theoretically grounded and operationally explicit tool for analyzing assessment items across diverse educational and assessment contexts, with potential implications for both assessment design and instructional improvement.

Keywords: physics test items, cognitive processing demands, individual developmental functions, assessment for learning, analytical framework

Citation: Dong, H., Wang, R., & Ding, Y. (2027). Analytical framework for cognitive demands in physics assessment items. *international Journal of Instruction*, *20*(1), 521-540.

INTRODUCTION

At present, fostering students' problem-solving abilities and higher-order thinking has become a central goal of contemporary science education reform (Tamayo, 2023). In this context, assessment is increasingly viewed not merely as an endpoint of instruction, but as a mechanism that supports learning and promotes cognitive development. Accordingly, science assessment items are no longer considered solely tools for measuring achievement; they are also expected to provide opportunities for cognitive engagement, reflecting the orientation of "assessment for learning" (Ministry of Education of the People's Republic of China, 2023).

Existing research indicates that assessment items can influence learning in multiple ways. Features such as contextual settings, information organization, and reasoning demands may shape students' conceptual understanding and cognitive processing during problem solving (Stott et al., 2023; Posner et al., 1982). Studies based on conceptual change and cognitive perspectives further suggest that task structure and presentation affect how students construct and reorganize knowledge (Posner et al., 1982), highlighting the developmental potential of assessment items beyond evaluation.

However, how such developmental affordances emerge from the interaction between item structures and students' cognitive activities remains unclear. Two main limitations persist. First, perspectives are fragmented: cognitive psychology focuses on students' reasoning processes (e.g., Cheng, 2011), while curriculum- or textbook-based studies emphasize task contexts and structures (e.g., Zhang et al., 2023; Cai, 2007). These approaches are rarely integrated into a unified framework linking cognition and item design. Second, methods remain limited, often relying on descriptive or intuitive analysis, with few tools that are theoretically grounded, operationally explicit, and replicable, making comparison and accumulation of findings difficult.

To address these gaps, this study develops an analytical framework for examining the cognitive demands and developmental affordances of physics assessment items. Drawing on cognitive psychology, sociology, and reasoning theory, it connects item structures with students' cognitive engagement. The framework includes three analytical dimensions and a structured procedure for systematic analysis. Its feasibility is illustrated through example analyses, inter-rater reliability testing, and preliminary applications across assessment contexts (Owolade et al., 2022, Aminatun et al., 2022).

Development of an Integrative Analytical Framework

Theoretical Foundations: Integrating Cognitive, Structural, and Reasoning

Perspectives

To examine how assessment item structures relate to students' cognitive engagement an d developmental affordances, this study integrates perspectives from cognitive psycholo gy, structure–function theory, and reasoning approaches. From a cognitive perspective, Cognitive Load Theory (Sweller, 2010; Sweller et al., 2019) explains how characteristic s of assessment items shape cognitive demands by influencing intrinsic, extraneous, and germane cognitive load, while Internal–External Representation Theory describes how

students process scientific problems through interactions between external representatio ns and internal mental representations (Mayer, 1992; Deng, 2002). Together, these pers pectives provide a basis for understanding both the cognitive demands imposed by asses sment items and the processes through which students engage with them. In addition, th e structure–function perspective (Parsons, 1951; Merton, 1968/1990) suggests that the f unctions of a system emerge not only from its individual components but also from relat ionships among those components. Applied to assessment contexts, structural features s uch as contexts, prompts, and visual representations may jointly shape students' cogniti ve engagement and developmental opportunities. To support more systematic analysis, t his study further adopts a structured reasoning approach in which explicit conditional re lationships are used to connect item characteristics with potential cognitive and develop mental implications. Integrating these perspectives, the framework addresses three quest ions: how specific structural features are processed during problem solving, what forms of cognitive engagement assessment items support, and what developmental affordance s may emerge through students' interactions with assessment tasks.

Three-Dimensional Analytical Content

Logical Generation of the Three-Dimensional Analytical Content

The development of the three-dimensional analytical content is guided by Cognitive Lo ad Theory as the central framework, while integrating Internal–External Representation Theory and structural functionalism. It proceeds through three stages: "content focus → division of cognitive stages and structural levels → determination of analytical dimensi ons."

(1) Stage One: Content Focus

The three types of cognitive load distinguished by Cognitive Load Theory correspond to three ways in which the structural features of test items influence problem-solving perf ormance. Based on this, three preliminary focal points for content analysis are identified : presentation features, degree of knowledge integration, and requirements of cognitive (thinking) levels (see Table 1).

Table 1
Three preliminary derived analytical content areas

Type of Cogniti ve Load	Core Mechanism (How It Affects Learning)	Derived Analytica l Focus
Extraneous Loa d	The presentation of information in the problem (e.g., contextual description, diagrams, clarity of questions) affects cognitive effi ciency	Presentation features of the item
Intrinsic Load	The knowledge required for problem solving and the degree of it s interactivity (conceptual complexity and interconnectedness) d etermine cognitive difficulty	Degree of knowledge integration
Germane Load	The thinking required for problem solving (e.g., understanding, analysis, integration, construction) guides deep cognitive engage ment	Requirements of cognitive (thinking) levels

(2) Stage Two: Division of Cognitive Stages and Structural Levels

To which cognitive stages of students' problem solving do these three analytical content areas correspond? And do they belong to the overall structure of the test item or to its c

onstituent elements? These two questions are addressed by Internal–External Represent ation Theory and structural functionalism.

According to Internal–External Representation Theory, the problem-solving process can be divided into two stages: external representation (understanding the problem) and int ernal representation (constructing a solution). The presentation features of a test item co rrespond to the external representation stage, while knowledge integration and cognitive (thinking) level requirements correspond to the internal representation stage.

At the same time, according to structural functionalism, the analysis of test item functio ns must consider both partial functions (i.e., constituent elements such as context, questi ons, and diagrams) and overall functions (i.e., the relationships among these elements). By intersecting these two perspectives, a contentanalysis matrix is formed (see Table 2).

Table 2
Analytical matrix based on theory and cognitive stages

Structural Level / Representati on Stage	External Representation (Unders tanding the Problem)	Internal Representation (Constructin g a Solution)
Partial Functions	Extracting key information	Applying specific skills
Overall Functions	Forming a holistic perception	Systemic thinking and integration of knowledge networks

Within external representation, holistic perception (e.g., contextual atmosphere, visual-textual style) and information extraction are cognitively separable. Students may first form an overall impression and then extract key information sentence by sentence; each process can independently exert influence. Therefore, these are retained as two distinct dimensions to examine the different functions of test items at the levels of affective attitudes and basic information-processing abilities.

In internal representation, however, the essence of cognitive activity lies in extracting, reorganizing, and integrating the knowledge, methods, and skills required for problem solving. Systemic thinking at the overall level and the application of specific skills at the local level are inseparable. Thus, the overall and partial functions within internal representation are combined into a single dimension—problem solving—which better reflects the integrative and holistic nature of higher-order thinking.

Accordingly, three structural levels are distilled:

1）External representation—overall function

2）External representation—partial function

3）Internal representation (integration of overall and partial functions)

(3) Stage Three: Determination of Analytical Dimensions

Based on the three analytical focal points initially derived from Cognitive Load Theory—presentation features, degree of knowledge integration, and cognitive (thinking) level requirements—three final analytical dimensions are established. Table 3 presents the theoretical foundations and structural levels of these three analytical dimensions.

Table 3
Formation of three-dimensional analytical content based on cognitive load theory

Type of Cognitive L oad	Focus of Analysis	Cognitive Stage and Struct ural Level	Formation of Analytical D imension
Extraneous Load	Presentation features of the item	External representation — overall function	Dimension 1: Holistic per ception
Extraneous Load	Presentation features of the item	External representation — partial function	Dimension 2: Information extraction and integration
Intrinsic Load + Ger mane Load	Knowledge integration + cognitive (thinking) l evel requirements	Internal representation (inte gration of overall and papar tial functions	Dimension 3: Problem sol ving

Operational Definitions of the Three Analytical Dimensions

The three dimensions correspond respectively to key stages in students' cognitive proce sses during problem solving and to different levels of the structural features of test items (see Table 4).

Table 4
Specific content of the three-dimensional analysis

Analytical Di mension	Corresponding Structural Focus	Core Functional Orientation	Examples
Dimension 1: Holistic Perception	Overall style, contextual theme, and mode of presentation of the item	Stimulating interest, shaping attitudes, conveying values	Does the context of the item stimulate interest? Does the overall presentation reflect scientific rigor?
Dimension 2: Information Extraction and Integration	Various components of the text (context, conditions, questions)	Basic abilities such as reading, extracting, and transforming information	How do students extract key information? Is transformation of information required (e.g., converting between text and diagrams)?
Dimension 3: Problem Solving	Logical structure between problem goals and given conditions	Knowledge application, thinking methods, and comprehensive abilities	What knowledge is required? What cognitive processes are involved (e.g., reasoning, modeling, argumentation)?

Dimension 1 focuses on how the overall presentation of an item shapes students' first impressions and emotional responses. Based on Cognitive Load Theory, overall presentation influences extraneous cognitive load, which may further affect learning motivation and affective attitudes.

Dimension 2 concerns the information-processing required for understanding the problem. Drawing on Internal–External Representation Theory, external representation involves two subprocesses: extraction of key information from text and diagrams, and integration to form a coherent mental representation.

Dimension 3 integrates the functions of internal representation and serves as the core dimension for realizing item functions. According to Internal–External Representation Theory, internal representation involves constructing a problem model and executing solutions. From a Cognitive Load Theory perspective, intrinsic and germane loads jointly shape cognitive demands. Thus, this dimension examines how items support knowledge integration, reasoning, and scientific argumentation, focusing on the level of thinking required (e.g., recalling formulas versus designing experiments or constructing arguments).

Syllogistic Reasoning Method

The syllogistic reasoning method is embedded in the analytical framework through a two-step operational procedure, providing a clear and replicable way to infer individual developmental functions.

Step 1: Analyzing the problem-solving process (minor premise). This step examines students' actual cognitive operations based on the three-dimensional framework, focusing on authentic cognitive processes rather than surface-level knowledge.

Step 2: Inferring item functions (conclusion construction). The cognitive operations identified in Step 1 are combined with established learning science principles (major premise). Through syllogistic reasoning, the potential developmental functions of test items are inferred, with an example provided in Table 5.

Table 5
Example of the Syllogistic Reasoning Process for Inferring the Functions of a Physics P roblem

Step	Stage	Specific Content and Operational Description
Analyze the Problem-Sol ving Process	Problem Content	Use your physics knowledge to explain why one can smell food as soon as the door is opened.
	Required Cognitive Op erations	Students need to identify the relevant physical domain (kinetic mo lecular theory), and apply concepts such as "thermal motion of mo lecules" and "diffusion" to construct a causal explanatory model c onnecting these concepts with a real-life phenomenon.
Infer Item Fu nctions	Minor Premise (Factua l Judgment)	To solve this problem, students must perform the cognitive task of applying the concept of "molecular thermal motion" to explain a real-life phenomenon.
	Major Premise (Theoretical Basis)	Contextualized application and explanation of knowledge are important pathways for promoting conceptual understanding and retention (cognitive constructivist theory).
	Conclusion (Preliminary Functional Inference)	By solving this problem, students can strengthen their understanding of the concept of "molecular thermal motion" and enhance their ability to apply this concept to explain phenomena.

Three-Step Systematic Operational Procedure

Step 1: Analyze students' problem-solving process (cognitive deconstruction). Based on the three-dimensional framework, this step systematically examines the cognitive operations required at different stages of problem solving (e.g., understanding the problem, processing information, and identifying strategies), rather than listing isolated knowledge points.

Step 2: Infer individual developmental functions. Each key cognitive operation identified in Step 1 (minor premise) is combined with learning science principles (major premise). Through syllogistic reasoning, the potential developmental function of the item is inferred within each dimension (conclusion).

Step 3: Integration and hierarchical articulation. The functions derived across the three dimensions are integrated, categorized, and hierarchically organized according to their nature or corresponding core competency dimensions, forming a systematic representation of the item's overall developmental functions.

Summary and Research Questions

This framework is designed to address the questions of "what to analyze" and "how to analyze."

Regarding what to analyze, the structure–function paradigm provides a logical framework for examining the structure of test items (i.e., at both the overall and component levels), while Cognitive Load Theory and Internal–External Representation Theory offer specific analytical dimensions and content.

Regarding how to analyze, the deductive syllogism (major premise – minor premise – conclusion) serves as the methodological tool to ensure the logical necessity of functional inference. On this basis, a three-step operational procedure is constructed to clarify the analytical process of the framework.

The relationships between the components of the framework and their theoretical foundations are presented in Table 7.

Table 7
Correspondence between the analytical framework and its theoretical foundations

Analytical Framework		Theoretical Foundations
Analytical Content	Logical framework	Structure–function paradigm (structure determines function; overall and partial levels)
	Specific dimensions	Internal–External Representation Theory (internal and external representations); Cognitive Load Theory (intrinsic, extraneous, and germane cognitive load)
Analytical Method		Deductive syllogism
Analytical Procedure		Three-step operational process constructed based on analytical content and meth ods

This framework translates theory into replicable operational procedures, enabling the an alysis of test item functions to move beyond empiricism and subjective arbitrariness tow ard systematization and scientific rigor. Its operability, reliability, and validity are furthe r examined through three progressive research questions:

(1) How can the framework's operational procedures and explanatory power be clearly demonstrated through the analysis of typical example problems?

(2) Can different raters achieve a high level of consistency when applying this framework to analyze the same set of test items?

(3) Is this analytical framework applicable to science assessment items across different cultural contexts (e.g., PISA, TIMSS, U.S. AP)?

Relationship Between the Present Framework and International Assessment Frameworks

(1) Differences in functional positioning

PISA focuses on assessing students' ability to apply scientific knowledge to identify issues, explain phenomena, and draw evidence-based conclusions. TIMSS, in contrast, evaluates students' mastery of curricular content across three cognitive domains: knowing, applying, and reasoning. AP Physics places greater emphasis on higher-order

scientific practices, including independent modeling, quantitative reasoning, and experimental argumentation. All three frameworks are primarily oriented toward summative assessment and seek to answer the question, *"What competencies have students acquired?"* In contrast, the present framework does not measure competencies that students have already developed. Rather, it analyzes the developmental potential embedded within assessment items themselves. By examining the structural characteristics of items, it infers the cognitive development and competency enhancement that students may potentially achieve through the problem-solving process, thereby addressing a different question: *"What kinds of development can an assessment item facilitate?"*

(2) Selective adoption of conceptual elements from international frameworks

Despite its distinct functional positioning, the present framework selectively draws on conceptual elements from the aforementioned assessment frameworks. Dimension 1 (holistic perception) draws on the integrated cognitive, affective, and behavioral conception underlying the assessment of *science identity* in PISA 2025. Accordingly, contextual themes and modes of presentation are established as independent analytical dimensions to examine how assessment items may stimulate students' interest in learning and shape their attitudes toward science. Dimension 2 (information extraction and integration) draws on the cognitive processes of *locating information* and *understanding* in the PISA 2018 Reading Framework, systematically analyzing how multimodal materials, including textual information, diagrams, and tables, influence students' information-processing processes. Dimension 3 (problem solving) integrates the higher-order cognitive processes of evidence analysis, synthesis, and generalization emphasized within the *reasoning* domain of TIMSS 2019, together with the emphasis on scientific practices in AP Physics. It also incorporates the analytical logic of intrinsic and germane cognitive load from Cognitive Load Theory, thereby deepening the analysis of the cognitive demands embedded in assessment items.

(3) Complementarity and further development: A novel process-oriented analytical perspective

While international assessment frameworks can reveal *"how students perform,"* they are less equipped to systematically explain *"why and how assessment items can—or cannot—promote learning."* This is precisely the focus of the present framework. Rather than replacing existing international assessment frameworks, it complements and extends them by introducing a process-oriented analytical perspective. Specifically, the framework clarifies the interaction pathways between the structural characteristics of assessment items—including context, representations, and internal logical relationships—and students' cognitive processing. It thus shifts the analytical focus from *"how students perform"* to *"how assessment items may facilitate such performance,"* namely, how items interact with students' cognitive processes and, consequently, influence learning outcomes.

On this basis, the framework can be applied to the in-depth analysis of international assessment items from PISA, TIMSS, AP Physics, and other large-scale assessment programs. By interpreting student performance from the perspective of item design, it

provides actionable guidance for improving assessment design and optimizing instruction. In this way, the framework establishes a complementary relationship with existing large-scale assessment frameworks and contributes to a more comprehensive assessment system through a differentiated yet mutually reinforcing analytical perspective.

METHOD

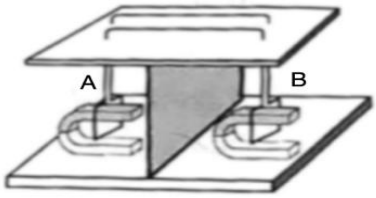
Application and Validation of the Framework

This section aims to demonstrate and validate the operability and explanatory power of the analytical framework constructed above through specific case analysis, inter-rater reliability testing, and preliminary cross-cultural exploration.

Case Demonstration: Application to a Chinese Physics Item

To verify the framework's operability, we analyzed a comprehensive item from the 2020 Henan High School Entrance Examination. The item involves "pushing one coil to cause another to swing," covering electromagnetic induction and energy transformation.

Figure 1
Example schematic figure



Step One: Multidimensional Deconstruction of the Cognitive Process

We systematically decomposed the cognitive operations required to solve the item across three dimensions:

Dimension 1 (Holistic Perception): The "technology club" context and text-diagram integration foster a sense of scientific inquiry and cooperative attitudes upon first reading.

Dimension 2 (Information Extraction): Students must identify the closed circuit and mechanical motion from multimodal inputs, transforming them into logical parameters like energy flow and physical roles.

Dimension 3 (Problem Solving): Solving requires applying induction principles and motor laws to analyze coil oscillations and energy conservation, involving high-order reasoning and knowledge transfer.

Step Two: Inferring Individual Developmental Functions

Using syllogistic reasoning ("if P, then Q"), we inferred the potential developmental affordances

Table 8
Syllogistic inference of the developmental functions of example item 1

Analytical Dimension	Major Premise (Theoretical Basis)	Minor Premise (Item Characteristics and Cognitive Demands)	Conclusion (Inferred Potential Functions)
Dimension 1: Holistic Perception	Scientific tasks situated in authentic contexts and involving collaborative features are conducive to stimulating students' learning motivation, shaping scientific attitudes, and fostering an initial understanding of the relationship between science and society.	This item is situated in a "technology club project" context and presents a collaborative inquiry scenario through the integration of text and diagrams. Students are required to form an overall understanding of the task context before engaging in problem solving.	At the level of holistic perception, the item may promote students' interest in learning physics, enhance their awareness of collaborative inquiry, and guide them toward an initial understanding of science–technology–society (STS) relationships.
Dimension 2: Information Extraction	Cognitive activities involving the extraction and integration of key information from multimodal texts (combining text and graphics) contribute to the development of students' abilities in information selection, comprehension, and integration.	To solve this item, students must extract key information from both the textual description and the schematic diagram, such as the fact that coils A and B form a closed circuit and their respective motion states. Students must also consider questions such as the basic components of an electric circuit and the causes of the coils' motion, and establish correspondences among these pieces of information.	At the level of information extraction, the item may promote students' reading comprehension, multimodal information integration skills, and their ability to identify the structural features of complex problems.
Dimension 3: Problem Solving	Systematic reasoning that integrates multiple physics principles promotes the development of knowledge integration and higher-order thinking.	Solving this item requires the coherent application of concepts related to electromagnetic induction, electric motor principles, circuit knowledge, and energy transformation to analyze why coil B oscillates after coil A is pushed. Students must reason about the functional roles of coils A and B within the circuit and the energy transformations involved, engaging in cognitive operations such as analysis, synthesis, and inference.	By solving this item, students may strengthen their understanding of electromagnetic induction, electric motor principles, circuit composition and functions, and energy transformation, while also developing logical thinking, scientific reasoning, and integrated problem-solving abilities involving analysis and judgment.

Step Three: Integration and Hierarchical Articulation

Finally, the fragmented functions are integrated into three levels of systematic expression:

Knowledge Understanding: Deepens comprehension of core concepts (induction, motors, energy) through knowledge structuring.

Ability Development: Trains higher-order skills in information extraction and logical judgment within authentic contexts.

Affective Attitudes: Stimulates inquiry interest and cultivates scientific responsibility via the "technology project" theme.

Inter-Rater Reliability Testing

To examine the stability and operability of the framework, five physics test items from the 2025 Beijing High School Entrance Examination were selected (covering mechanics, electromagnetism, optics, and electricity; item types included multiple-choice, fill-in-the-blank, experimental inquiry, short answer, and calculation). Four independent raters were invited (Table 9),they completed Step 1 (cognitive deconstruction) and Step 2 (functional inference).

Table 9
Background information of the raters

Rater	Research Area	Relevant Experience
A: University Faculty Member (PhD)	Physics Curriculum and Instruction	Systematic training in educational and psychological measurement; doctoral dissertation focused on cognitive load analysis of physics assessment items; familiar with the development and application of the PISA scientific literacy assessment framework
B: University Faculty Member (PhD)	Science Education Assessment	Systematic training in educational and psychological measurement; familiar with the cognitive dimensions and operationalization methods of international assessment frameworks, including PISA and TIMSS
C: Senior Secondary School Physics Teacher (Master's Degree)	Secondary School Physics Education	12 years of frontline teaching experience; multiple experiences in developing and reviewing items for local academic quality monitoring assessments; specialized training in international assessment frameworks, including PISA and TIMSS; extensive experience studying released PISA assessment items
D: Secondary School Physics Teacher (Master's Degree)	Secondary School Physics Education	10 years of frontline teaching experience; multiple experiences in developing and reviewing items for local academic quality monitoring assessments; extensive experience studying released PISA assessment items; participated in the localization, pilot testing, and research of a physics literacy assessment instrument based on the PISA framework

All raters first received standardized training on the analytical framework and coding rules. They then independently completed two tasks:

(1) coding the three types of cognitive processing involved in students' problem solving (Step 1);

(2) inferring the potential individual developmental functions of the items (Step 2).

During the coding process, raters worked independently without any form of discussion.

Consistency of Coding in "Step One"

(1) Coding Framework for Step One

According to the analytical framework, students' problem solving involves external representation (holistic perception and information extraction) and internal representation (problem solving). Among these, "holistic perception" mainly corresponds to the context of the item, "information extraction" is reflected through its presentation, and "problem solving" involves the required knowledge content and cognitive operations.

Therefore, analyzing the individual developmental functions of test items requires addressing two key questions:

(1) What effects do the context and presentation of the item have on students?

(2) What knowledge content and cognitive operations are required for students to solve the problem?

Based on this analysis, a coding framework for Step One was developed (see Table 10) for inter-rater consistency testing. The coding follows three principles:

1. Each item is single-coded (except where multiple selections are explicitly allowed);

2. Coding is based on the objective characteristics of the item, rather than the rater's personal problem-solving process;

3. For "complexity" ratings (e.g., M3, M4, M5), a relative comparison method is used—items are ranked by complexity among the set and assigned scores from 1 to 5 accordingly.

Table 10
Coding framework for "step 1"

First-Level Indicator	Second-Level Indicator	Guiding Question	Scoring Criteria and Examples
Context	Presence of context	1. Is the item contextualized? A. Yes B. No	"Context" refers to whether the item is situated in a real or simulated scenario. Purely abstract items without any background are coded as "No context."
	Type of context	2. What type of context does the item belong to? (Applicable only if Question 1 = A) A. Daily life practice B. Engineering or production C. Experimental inquiry D. Interdisciplinary (e.g., language, mathematics) E. Moral and civic education F. Other (specify) G. None (select if Question 1 = B)	Daily life practice (e.g., transportation, sports); engineering or production (e.g., machinery, circuits); experimental inquiry (e.g., experimental data, inquiry processes); interdisciplinary (background drawn from other subjects); moral and civic education (e.g., values education, national identity). If F is selected, the specific content must be specified.
Presentation	Contextual description	3. How complex is the contextual description? (e.g., amount of information, level of detail, background description) 1 (very simple) – 5 (very complex)	Rated based on the amount of detail, background information, and coherence of the contextual description. Brief and minimal descriptions are rated 1; rich and multidimensional descriptions are rated 5.
	Problem statement	4. How complex is the problem statement? (e.g., implicit conditions, distracting information) 1 (clear and direct) – 5 (very complex)	Rated based on the presence of implicit conditions, distracting information, and ambiguity in wording. Direct and explicit statements are rated 1; statements requiring repeated interpretation are rated 5.
	Representational difficulty	5. What is the number and transformation difficulty of representational forms required to understand the item? (e.g., text, diagrams, formulas, tables) 1 (single) – 5 (multiple and complex)	Count the representational forms involved (e.g., text, images, formulas, tables, schematic diagrams) and evaluate the difficulty of transforming among them. A single representational form is rated 1; multiple forms requiring complex integration and transformation are rated 5.
Knowledge Content	Number of knowledge elements	6. How many physics knowledge elements are required to solve the item? 1 (one element) – 5 (five or more elements)	Only core physics concepts, laws, and formulas that are essential for solving the item are counted. Repeated or redundant counting is avoided.
Cognitive Operations	Type of cognitive operation	7. What are the primary cognitive operations required to solve the item? (Multiple selections allowed) A. Analysis and synthesis B. Reasoning and judgment C. Modeling and model construction D. Induction and deduction E. Other (specify)	Multiple selections are allowed. For each option, raters indicate "Yes" or "No" based on whether the cognitive operation is a major component of the problem-solving process. If E is selected, the specific operation must be specified.

(2) Data Processing for "Step One"

The coding results were compiled into a summary record sheet (see Table 11). Indicator s M3, M4, M5, and M6 were treated as scored variables and entered into SPSS for analy

sis. The intra-class correlation coefficients (ICC) were calculated as 0.88, 0.82, 0.88, an d 0.83, respectively, all exceeding 0.80, indicating good inter-rater reliability.

Indicators M2 and M7 were categorical or multiple-choice variables, for which item-lev el percentage agreement was calculated. The results were 100% and 90%, respectively (with M1 assumed to be consistent by default), also reaching an acceptable level of agree ment for academic research.

Overall, the results indicate that the coding in this study demonstrates strong reliability.

Table 11
Coding record sheet for "step 1"

Item	Rater	Scoring Indicator（M）										
								Cognitive Operations (Multiple selections allowed)（M7）				
		Presence of Context (M1)	Type of Context (M2)	Contextual Description (M3)	Problem Statement (M4)	Representational Difficulty (M5)	Number of Knowledge Elements (M6)	Analysis & Synthesis	Reasoning & Judgment	Model Construction	Induction & Deduction	Other
Q1	R1	A	A	1	2	2	3	No	Yes	Yes	No	No
	R2	A	A	1	2	2	3	No	Yes	Yes	No	No
	R3	A	A	1	2	2	3	No	Yes	Yes	No	No
	R4	A	A	2	2	2	2	No	Yes	No	No	No
Q2	R1	A	C	1	1	2	2	Yes	Yes	No	No	No
	R2	A	C	2	2	2	2	Yes	Yes	No	No	No
	R3	A	C	2	2	2	2	Yes	Yes	No	No	No
	R4	A	C	2	2	2	2	Yes	Yes	No	No	No
Q3	R1	A	C	2	2	2	3	Yes	Yes	No	Yes	No
	R2	A	C	2	2	2	2	Yes	Yes	No	No	No
	R3	A	C	2	2	2	2	Yes	Yes	No	Yes	No
	R4	A	C	1	2	2	2	Yes	Yes	No	No	No
Q4	R1	A	C	3	2	3	2	Yes	Yes	Yes	No	No
	R2	A	C	2	2	3	2	Yes	Yes	Yes	No	No
	R3	A	C	2	2	3	2	Yes	Yes	Yes	No	No
	R4	A	C	3	2	3	2	Yes	Yes	Yes	No	No
Q5	R1	A	A	4	3	4	3	Yes	Yes	No	Yes	No
	R2	A	A	4	3	3	3	Yes	Yes	No	No	No
	R3	A	A	4	3	3	3	Yes	Yes	No	Yes	No
	R4	A	A	4	3	4	3	Yes	Yes	No	No	No

Inter-Rater Consistency in Functional Inference for "Step 2"

In Step 2, the two raters independently inferred the potential individual developmental f unctions of the test items based on the coding results obtained in Step 1. A comparison of the results indicates that the raters demonstrated a high level of consistency in terms of functional level classification (i.e., knowledge understanding, competency developm ent, and affective attitudes) as well as in the identification of core functional orientation s. Minor discrepancies were observed only at the level of wording in the articulation of specific functions. These findings suggest that functional inference based on the propos ed framework exhibits good interpretability and stability, enabling different researchers to arrive at highly consistent inferential conclusions when working from the same analyt ical evidence.

Summary

The results of the inter-rater reliability testing indicate that the proposed analytical fram ework supports stable coding of the cognitive processing demands of test items by diffe rent researchers and enables consistent inference of potential individual developmental f unctions based on those codings. These findings provide methodological support for ap plying the framework to larger samples in subsequent systematic analyses.

FINDINGS

Cross-Cultural Applicability: PISA, TIMSS, and AP Analysis

To test the framework's versatility, we analyzed ten science items from PISA, TIMSS, and U.S. AP assessments. This analysis focuses on the framework's consistency across different curricula rather than comparing student performance.

Coding of Cognitive Deconstruction for Cross-Cultural Items

Following the Step One procedure, items were coded for context complexity (M1–M3), representational difficulty (M5), and cognitive operations (M7). The results are recorded in the *Coding Record Table for Cross-Cultural Science Sample Items* (see Table 12).

Coding Results: Most international items (e.g., PISA Q5, AP Q3) featured authentic contexts and multimodal representations.

Complexity Trends: Items like PISA Q5 (Heat Dissipation) showed high contextual complexity (M3=4) and multivariable analysis requirements , while others like AP Q1 focused on logical reasoning within simpler contexts (M3=1).

Table 12
Coding table for cross-cultural science sample items

Item	Presence of Context （M1）	Type of Context （M2）	Contextual Description （M3）	Problem Statement （M4）	Representational Difficulty （M5）	Number of Knowledge Elements （M6）	Cognitive Operations (Multiple selections allowed) （M7）
Q1	A	A	1	2	1	2	B C
Q2	A	C	2	1	3	2	A B C
Q3	A	F（theoretical physics）	2	3	2	2	A B C
Q4	A	D	3	1	2	2	A B
Q5	A	A	4	3	4	2	A B C D
Q6	A	A	1	2	2	2	B C
Q7	A	A	2	3	3	2	A B C
Q8	A	B	3	2	3	2	A B C
Q9	A	C	2	2	3	3	A B C
Q10	A	C	2	3	2	2	A B

Inference and Articulation of Individual Developmental Functions

We applied Step Two and Three to infer developmental functions across three levels: knowledge, competency, and attitudes.(see Table 13).

Table 13
Individual developmental functions of cross-cultural science sample items

Item	Individual Developmental Functions		
	Knowledge understanding	Competency development	Affective attitudes
Q1	Independence of mass variation; universality of Newton's second law	Eliminating irrelevant information; model identification; logical reasoning	Interest in space exploration; rigorous and truth-seeking scientific attitude
Q2	Relationship between velocity–time (v–t) graphs, direction of motion, and acceleration	Transformation between graphical and textual representations; analysis of dynamic processes; mathematical modeling	Confidence in using graphical tools to analyze motion problems
Q3	Scalar superposition of electric potential; electric potential formula for point charges; conceptual differentiation	Information extraction; construction of superposition models; logical reasoning and symbolic computation	Interest in exploration; rigorous attitude; appreciation of the integration of mathematics and physics
Q4	Relationship between universal gravitation and acceleration in celestial motion	Information extraction; elimination of interference; scientific argumentation; logical judgment	A cognitive orientation toward explaining natural phenomena scientifically
Q5	Combined effects of humidity and temperature on heat dissipation efficiency; interdisciplinary connections	Data extraction; multivariable analysis; scientific argumentation; inductive reasoning	Awareness of the value of scientific knowledge in health and everyday decision-making
Q6	Causes of object falling; conservation of mechanical energy; energy dissipation	Abstracting physical models from real-life phenomena and providing explanations	Interest in everyday inquiry; habits of observation, thinking, and verification; a truth-seeking attitude
Q7	Law of conservation of mass and its applications; essence of chemical reactions	Integration of textual and graphical information; scientific reasoning; scientific argumentation; verbal expression	Disciplinary identity in using scientific laws to explain real-world phenomena
Q8	Conditions for lever equilibrium and their applications; force analysis	Extraction of graphical and textual information; formula-based calculation and reasoning; modeling and solving practical problems	Recognition of the value of scientific knowledge in improving labor efficiency; strengthened sense of collaboration
Q9	Relationship between force and motion; conditions for equilibrium under multiple forces	Graphical–textual transformation; model construction; logical judgment	Awareness and confidence in making inferences based on physical laws
Q10	Methods for extracting period and frequency information from oscilloscope waveforms	Data processing; calculation; systematic reasoning	Positive attitudes toward scientific measurement and experimental inquiry

Summary

Taken together, the results reported above directly address the three research questions guiding this study. With respect to RQ1, the case demonstration (see Case Demonstration: Application to a Chinese Physics Item) shows that the three-step operational procedure can transparently and systematically uncover the cognitive

functions embedded in a typical physics item, confirming the framework's operability and explanatory power. With respect to RQ2, the inter-rater reliability testing yielded consistently high agreement across raters (ICC = 0.82–0.88 for scored indicators; percentage agreement = 90%–100% for categorical indicators), demonstrating that the framework can be applied with a high degree of consistency by different raters. With respect to RQ3, the successful coding and functional inference of PISA, TIMSS, and AP items confirm that the framework is applicable to physics assessment items across different cultural and assessment contexts, providing preliminary evidence of its cross-cultural robustness. Together, these findings confirm that the framework can systematically characterize cognitive demands and potential functions across diverse educational contexts, offering preliminary methodologicalsupport for its use as a universal tool for international science assessment research.

DISCUSSION

Bridging the Gap: Linking Item Structure to Cognitive Development

The core theoretical contribution of this study lies in the transition from descriptive item analysis to inferential functional analysis. Unlike traditional approaches that evaluate items based on difficulty or discrimination indices, our framework posits that the developmental value of a physics item is "embedded" within its structural design. Harrison et al.(2023) demonstrated a systematic relationship between characteristics of PISA science items, such as cognitive demand, and item difficulty; however, their analysis stops short of explaining “why students answer an item correctly or incorrectly.”However,By integrating Cognitive Load Theory with the structure-function paradigm, we have moved beyond observing what students know to explaining how the item's architecture (context, representation, and logic) constraints and guides cognitive growth. This precisely responds to and extends this line of inquiry.

The discussion below elaborates on these findings in turn: Bridging the Gap and Beyond Evaluation address the theoretical implications of RQ1, while Practical Utility and Cross-Cultural Robustness addresses the empirical implications of RQ2 and RQ3.

Beyond Evaluation: Assessment Items as Learning Scaffolds

Our findings suggest that assessment items can function as "micro-curricula". For instance, the case analysis of the electromagnetic induction item demonstrates that Dimension 1 (Holistic Perception) and Dimension 2 (Information Extraction) act as essential precursors to Dimension 3 (Problem Solving). Kusumaningtyas et al. (2024) developed a physics assessment instrument grounded in higher-order thinking skills that can accurately assess high school students’ understanding and application of concepts related to work and energy, while also effectively enhancing their problem-solving abilities. In contrast, the present framework provides an analytical tool for explaining“why it is effective” by revealing how assessment items promote students’cognitive development across the three dimensions.This hierarchical interaction implies that item designers can intentionally manipulate extraneous load (via context) and germane load (via reasoning demands) to stimulate specific competencies,

such as scientific argumentation or multimodal literacy. This shifts the role of the item from a passive measurement tool to an active mechanism for Assessment for Learning.

Practical Utility and Cross-Cultural Robustness

The high inter-rater reliability (ICC > 0.80) and successful application to PISA, TIMSS, and AP tasks provide strong evidence for the framework's systematic rigor and cross-cultural applicability. Compared with Letmon's (2023) approach of using Bloom's taxonomy alone to analyze the cognitive demand of assessment items, the present framework incorporates Cognitive Load Theory to identify how each of the three dimensions contributes to students' cognitive development, rather than simply assigning an overall "cognitive level."The syllogistic reasoning method ensures that functional inferences are not subjective "guesses" but logically necessary conclusions derived from observable task characteristics. This provides educators and test developers with a unified "analytical language" to diagnose and optimize the developmental potential of physics assessments globally (Sarjono et al., 2018, Istiyono et al., 2019).

CONCLUSION

Grounded in the shift toward competency-oriented science education, this study developed an integrative framework to analyze the cognitive demands and developmental affordances of physics assessment items. By synthesizing cognitive psychology and the structure-function paradigm, we established a three-dimensional structure—holistic perception, information extraction, and problem solving—supported by a three-step operational procedure using syllogistic reasoning.

The application and validation results demonstrate that the framework provides a theoretically grounded and replicable tool for systematic item analysis. It successfully links the external structural features of items to internal cognitive processing, offering a transparent pathway for inferring how assessment tasks promote knowledge integration and higher-order thinking.

While this study focuses on theoretical potential rather than measured learning outcomes, it offers significant practical value for item design and quality evaluation. Furthermore, the validation of the present framework was based on expert raters who had extensive experience in physics education research and teaching practice and had been trained in the analytical procedures of the framework. However, none of the raters had been directly involved in the development of PISA or TIMSS assessment items. Therefore, the applicability and acceptability of the present framework among experts in international large-scale assessment warrant further validation in future research.

Future research should focus on developing standardized manuals for large-scale textbook analysis and integrating the framework with learning analytics to examine how these potential functions translate into actual student performance in diverse classroom settings.

ACKNOWLEDGEMENTS

This work was supported by the Graduate Education Reform Project of Henan Province (Grant No. 2025SJGLX102Y).

The corresponding author acknowledges the financial support provided by the China Scholarship Council (CSC) for doctoral studies in Spain.

REFERENCES

Aminatun, T., Subali, B., Yuningsih, Y., Dwiyani, A., Prihartina, I., & Meliana, D. (2022). Developing android-based mobile through local ecosystem materials to improve thinking skills of high school students. *Anatolian Journal of Education, 7*(1), 73–82.

Cai, J., & Nie, B. (2007). Problem solving in Chinese mathematics education: Research and practice. ZDM – *The International Journal on Mathematics Education, 39*(5–6), 459–473.

Cheng, S., & Siow, H.-L. (2011). Cognitive processes in problem solving: Bridging the intersection of cognition and management science. Journal of Management & Organization, 17(6), 792–806.

Deng, Z. (2002). Representational state theory and empirical research on problem solving (Doctoral dissertation). Nanjing Normal University.

Du, Y., & Huang, Q. (2020). Mechanisms and facilitation strategies of online deep learning. *China Higher Education Research*, (6), 58–63.

Harrison, S., Kroehne, U., Goldhammer, F., Lüdtke, O., & Robitzsch, A. (2023). Comparing the score interpretation across modes in PISA: An investigation of how item facets affect difficulty. *Large-scale Assessments in Education, 11*(1), 1–21.

Istiyono, E., Dwandaru, W. S. B., Lede, Y. A., Rahayu, F., & Nadapdap, A. (2019). *Developing IRT-based physics critical thinking skill test: A CAT to answer 21st century challenge. International Journal of Instruction, 12*(4), 267–280.

Jiang, Q., & Qin, C. (2022). Logic. Higher Education Press.

Kusumaningtyas, D. A., Manyunu, M., Kurniasari, E., Awalin, A. N., Rahmaniati, R., & Febriyanti, A. (2024). Enhancing learning outcomes: A study on the development of higher order thinking skills based evaluation instruments for work and energy in high school physics. *Indonesian Journal on Learning and Advanced Education (IJOLAE), 6*(1), 14–31.

Letmon, D. (2023). Examining the cognitive demand of high-stake physics and chemistry examinations in Ireland [Doctoral dissertation, Dublin City University]. *DORAS DCU Research Repository.*

Lin, J. (2025). Advancing science education through socioscientific issues learning. *Primary and Secondary School Science Education, 2*(4), 26–31.

Mayer, R. E. (1992). *Thinking, problem solving, cognition*. W.H. Freeman.

Merton, R. K. (1990). *On theoretical sociology* (F. He et al., Trans.). Huaxia Press. (Original work published 1968)

Ministry of Education of the People's Republic of China. (2023, May). *Action plan for deepening curriculum and instructional reform in basic education*. http://www.moe.gov.cn/srcsite/A26/jcj_kcjcgh/202306/t20230601_1062380.html

Owolade, A. O., Salami, M. O., Kareem, A. O., & Oladipupo, P. O. (2022). Effectiveness of guided inquiry and open inquiry instructional strategies in improving biology students' achievement. *Anatolian Journal of Education, 7*(2), 19–30.

Parsons, T. (2008). The structure of social action (M. Zhang, Y. Xia, & G. Peng, Trans.). Yilin Press. (Original work published 1951)

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Towards a theory of conceptual change. *Science Education, 66*(2), 211–227. https://doi.org/10.1002/sce.3730660207

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, 101860.

Sarjono, Mardapi, D., & Mundilarto (2018). Development of physics lab assessment instrument for senior high school level. *International Journal of Instruction, 11*(4), 17–28.

Stott, A. E. (2023). Epistemological lessons from inconsistencies in teachers' errors related to use of the mole ratio in stoichiometry calculations: A cue for professional development. *Journal of Chemical Education, 100*(7), 2548–2557. https://doi.org/10.1021/acs.jchemed.2c01252

Tamayo, O. E. (2023). *Critical thinking and epistemic sophistication in science education. Education Sciences, 13*(8), 803.

Yang, S. (1994). A re-examination of the definition of deductive reasoning. *Journal of South China Normal University (Social Science Edition)*, (3), 28–33.

Zhang, D. (2023). A comparative study of worked examples in junior secondary mathematics textbooks from a problem-context perspective: *A comparison of PEP and BNU editions. Journal of Mathematics Education, 32*(3), 45–50.