JOURNAL OF BALTIC SCIENCE EDUCATION

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

INTERDISCIPLINARY THINKING AMONG SEVENTH-GRADE STUDENTS IN LOWER-SECONDARY SCIENCE EDUCATION

Shasha Chen,
Shaohui Chi,
Zuhao Wang

Abstract. *Interdisciplinary thinking is critical for equipping students to apply scientific knowledge and tackle societal challenges across various disciplines, which has been recognized as a key objective of twenty-first century science education. However, research on effective interdisciplinary assessment in secondary school science education is still limited. The purpose of this study was to develop and validate an instrument for evaluating seventh graders' interdisciplinary thinking within lower-secondary science contexts. A four-dimensional framework for evaluating interdisciplinary thinking was proposed, leading to the development of an assessment instrument. Participants were 316 seventh–grade students randomly selected from a lower-secondary school in Jiangsu, China. The multidimensional random coefficients multinomial logit (MRCML) model was employed to examine the reliability and validity of the instrument. The results indicated that the four–dimensional conceptual framework of interdisciplinary thinking was appropriate, and the multidimensional partial credit model (PCM) fitted the data reasonably well. The differential item functioning (DIF) analysis also revealed the absence of gender bias, confirming the measurement invariance statistics reached. In addition, disparities in student performance across the four dimensions were identified, highlighting the need for tailored instructional practices to foster their ability in lower-secondary school science education.*

Keywords: *educational assessment, interdisciplinary thinking, multidimensional item response model, Rasch measurement, science education*

Shasha Chen, Shaohui Chi, Zuhao Wang
East China Normal University, P. R. China

Introduction

As real-world challenges are inherently complex, ill-structured, and interdisciplinary in nature (Chin & Chia, 2006), the ability to think critically and integrate scientific ideas across multiple disciplines has become increasingly essential for effective scientific problem-solving (Alagona & Simon, 2010; Gardiner, 2020; Melton et al., 2022). In such societal challenges, interdisciplinary thinking provides students with the cognitive capabilities necessary to navigate causal uncertainty and devise effective solutions to pressing social issues, such as those related to climate change and sustainability (Tan & So, 2019; Thomas, 2009; Tripp & Shortlidge, 2020). However, research has shown that students frequently have difficulties recognizing contributions from multiple disciplines to complex topics (Richter & Paretti, 2009), and synthesizing insights across different disciplinary perspectives (Spelt et al., 2009; Zhan et al., 2017). As a result, fostering interdisciplinary thinking has been widely recognized as a critical goal at the K-12 levels (Bestelmeyer et al., 2015; National Research Council, 2012; Newell & Luckie, 2019; Scott, 2015). Interdisciplinary teaching and learning practices have been emphasized to engage students with societal issues (Asghar et al., 2012; Newell & Luckie, 2019), with the belief that such learning experiences could benefit students' cognitive learning outcomes and understanding of real-world problems (Lattuca et al., 2017; Markauskaite et al., 2024).

The assessment of students' interdisciplinary thinking is an important part for formulating tailored pedagogical strategies that effectively enhance its development in middle school science (Asghar et al., 2012). Despite the acknowledged significance of interdisciplinary thinking, developing a valid and reliable assessment for interdisciplinary learning remains challenging (Gao et al., 2020; Tripp et al., 2020). While previous studies have explored the theoretical elements of interdisciplinary thinking and emphasized its value for comprehensive scientific problem-solving, few studies have specifically developed valid and reliable assessment instruments designed to measure this complex construct among middle school students (Lattuca et al., 2017). Conventional assessment methods, including written assignments, surveys,

https://doi.org/10.33225/jbse/25.24.53

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

and think-aloud interviews, have often failed to capture the multifaceted nature of interdisciplinary thinking with empirical rigor (Claus & Wiese, 2019; Tripp & Shortlidge, 2020; Tripp et al., 2020; Wolfe & Haynes, 2003), resulting in a disparity between assessment practices and developmental objectives in interdisciplinary learning (Gao et al., 2020). Besides, these assessments did not report on the validation process for the quality of the instruments. Moreover, there is a scarcity of research concerning lower-secondary school students, which leads to a poor understanding of student competency development.

To bridge this gap, the current study aimed to develop and empirically validate an assessment instrument to evaluate seventh graders' proficiency in interdisciplinary thinking within lower-secondary school science education. The multidimensional Rasch analysis was employed to confirm the reliability and construct validity of interdisciplinary thinking while accounting for item difficulty and student ability. The findings from this study were intended to inform educational practices, thereby supporting the targeted development of interdisciplinary thinking throughout lower-secondary school science education.

Literature Review

Interdisciplinary Thinking

Interdisciplinary thinking are complex cognitive abilities that entail making connections among multiple ideas and integrating knowledge and approaches from various disciplinary perspectives within an interdisciplinary system (Lattuca et al., 2017; Mansilla & Duraising, 2007; Newell, 2001; Spelt et al., 2009; Spelt et al., 2017; Tripp & Shortlidge, 2020; Van den Besselaar & Heimeriks, 2001; van Merriënboer, 1997). Research suggests that interdisciplinary thinking encompasses multiple dimensions simultaneously (Spelt et al., 2009), including disciplinary grounding (Mansilla & Duraising, 2007; Tripp & Shortlidge, 2019), knowledge integration (Repko, 2008; Tripp & Shortlidge, 2020), critical awareness (Mansilla et al., 2009; Tripp et al., 2020), objectivity (Tripp & Shortlidge, 2020), system analysis (Wang & Song, 2021), reasoning and communication (Lattuca et al., 2013; Shen et al., 2015), and reflection (Claus & Wiese, 2019; Lattuca et al., 2013). For instance, Spelt et al. (2017) identified three key dimensions of interdisciplinary learning experiences as emotional, cognitive, and social. Tripp and Shortlidge (2019) categorized interdisciplinary thinking into four components: disciplinary grounding, diverse research methods, knowledge integration, and across-disciplinary collaboration. Interdisciplinary thinking is also exemplified in the Next Generation Science Standards (NGSS) by categorizing each standard into seven crosscutting concepts, such as stability and change; and scale, proportion, and quantity (NGSS Lead States, 2013).

In secondary science education, the cognitive complexity of interdisciplinary thinking is contextualized and actualized through students' engagement in scientific learning activities grounded in real-world environments (Tan & So, 2019; Tripp et al., 2020). Such authentic and contextualized tasks provide opportunities for students to advance their understanding by tackling interdisciplinary problem-solving (Tripp & Shortlidge, 2020). This means that students need to acknowledge the interdisciplinary nature of science to develop cognitive structures that incorporate disciplinary boundaries and approaches (Mansilla & Duraising, 2007), enabling reflective analysis of system components and constructing solid arguments to address interdisciplinary challenge, thus supporting the advancement of comprehensive decision-making.

Enhancing Students' Interdisciplinary Thinking through Science Education

Interdisciplinary thinking provides students with opportunities for cognitive advancement that would not be achieved through monodisciplinary learning (Borrego & Newswander, 2010; Mansilla & Duraising, 2007; Newell, 2001; Nikitina, 2005; Spelt et al., 2009). This cognitive flexibility enables them to analyze complex phenomena and create solutions in interdisciplinary ways, such as environmental sustainability, food security, and global health challenges (Bestelmeyer et al., 2015; Ivanitskaya et al., 2002; Knobloch et al., 2020; Valley et al., 2018). Lattuca et al. (2004) addressed interdisciplinarity through cognitive and sociocultural lenses. From a cognitive perspective, it promotes the integration of prior knowledge with learning activities, allowing students to tackle ill-structured problems and develop high-order thinking skills; while socioculturally, it enhances their epistemological growth through social and cultural interactions that shape learning (Illeris, 2003). These two perspectives contribute valuable insights into how and why students learn, therefore fostering interdisciplinary habits of mind and practices.

Interdisciplinary thinking emerges as students transition from fundamental disciplinary knowledge to more complex, interconnected understandings that transcend disciplinary boundaries (Cowden & Santiago, 2016; Kno-

https://doi.org/10.33225/jbse/24.23.53

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

bloch et al., 2020). This progress, accompanied by reflective behaviors (Brassler & Dettmers, 2017), underscores the need for tailored curriculum designed to support such cognitive growth. Numerous studies suggest that experiential and problem-based learning approaches, integrated into curricular frameworks within scientific practice, are instrumental in facilitating such core interdisciplinary competences (Brassler & Dettmers, 2017; Busta & Russo, 2020; Clemmons et al., 2020; Spelt et al., 2017; Stentoft, 2017). For example, Busta and Russo (2020) created an integrative plant chemistry module within laboratory settings. Their findings indicated that authentic hand–on scientific experiences promoted students' interdisciplinary thinking and led to innovative inferences. Nevertheless, limited assessment has been undertaken to assess the development of students' competence in this domain, hence posing challenges for the effective cultivation of interdisciplinary thinking within the lower-secondary school science context.

Assessments Related to Interdisciplinary Learning Outcomes

Assessment is a crucial component of the learning environment that fosters interdisciplinary thinking among students. However, despite a significant emphasis on interdisciplinary learning and instruction, there exists a paucity of empirical research regarding effective interdisciplinary assessment methods within school science education (Gao et al., 2020; Mansilla, 2005; You et al., 2018). For example, Lan et al. (2021) conducted a pencil–and–paper assessment for upper–secondary students regarding environmental issues based on Wilson's Construct Modeling framework. Wang and Song (2021) evaluated the interdisciplinary competences of middle school students with short answers, extended–response, and hands-on assignments.

Several studies have performed interdisciplinary assessments utilizing college students as subjects. Shen et al. (2014) designed an interdisciplinary assessment within the Web–based inquiry science environment to evaluate college students' interdisciplinary understanding of osmosis from a knowledge integration perspective. Similarly, You et al. (2018) validated the Interdisciplinary Scientific Assessment of the Carbon Cycle for senior high school and college students, combining both multiple-choice and constructed-response tasks across the fields of physics, chemistry, biology, and earth science. Tripp and Shortlidge (2020) employed the written assignment to assess undergraduates' interdisciplinary science thinking in relation to real-world issues, and validate the construct validity of the Interdisciplinary Science Rubric (IDSR) using semi-structured and think-aloud interviews (Tripp et al., 2020). Jayathilaka (2021) further constructed and validated the Interdisciplinary Problem-Solving Quiz (IPSQ) to evaluate undergraduates' comprehension and application of concepts towards real-world issues at the junction of subdisciplines in chemistry, including Organic, Inorganic, Physical, Analytical, and Biochemistry.

To summarize, there is a deficiency of empirical research on the assessment of interdisciplinary thinking, particularly in a lower-secondary science education context. Moreover, the existing assessments often fail to capture the multifaceted features of students' interdisciplinary thinking in authentic societal contexts. This gap may stem from a lack of available instruments to assess interdisciplinary thinking throughout secondary school science, leaving the field remaining in an exploratory stage with limited empirical evidence concerning the dimensional structure of interdisciplinary thinking. Consequently, there is a pressing need for tasks linked to real–world social issues to thoroughly assess students' interdisciplinary thinking. The findings intend to inform instructional practices to develop and solidify students' such competence in an interdisciplinary science curriculum at the lower-secondary level.

Research Aim and Research Questions

Despite its acknowledgement as a complex and multifaceted cognitive competence, interdisciplinary thinking has not been thoroughly validated through empirical research. The purpose of this study was to develop and validate an assessment instrument for evaluating seventh graders' interdisciplinary thinking. The research sought first to enhance the current knowledge base by empirically validating a conceptual framework of interdisciplinary thinking using multidimensional Rasch analysis. And then, an assessment of seventh–grade students' proficiency in interdisciplinary thinking was performed within the context of lower-secondary science education. Specifically, the research was directed by the subsequent questions:

1. What evidence confirms the dimension structure, reliability, and construct validity of the instrument developed to evaluate seventh grade students' interdisciplinary thinking?
2. What are the performance levels of seventh-grade students in interdisciplinary thinking across its subdimensions?

https://doi.org/10.33225/jbse/25.24.53

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Measurement Framework

According to the literature review, the current study defined interdisciplinary thinking as a multifaceted cognitive ability that enables students to integrate knowledge and methods from multiple disciplines to dynamically analyze, articulate, and reflect on real-world phenomena from diverse perspectives (Borrego & Newswander, 2010; Claus & Wiese, 2019; Mansilla & Duraising, 2007; Spelt et al., 2009; Tripp & Shortlidge, 2020). The framework for interdisciplinary thinking consists of four related competency dimensions: comprehensive analysis (CA), interpretive argumentation (IA), systematic integration (SI), and critical reflection (CR) (Claus & Wiese, 2019; Lattuca et al., 2017; Wang & Song, 2021; Wolfe, 2011; Wolfe & Haynes, 2003). Together, these four dimensions foster deeper cognitive engagement by assisting students in constructing meanings across disciplinary boundaries, deepening their extensive understanding and cognitive flexibility in science learning (see Appendix A for complete framework and competency levels of interdisciplinary thinking).

Comprehensive analysis involves the ability to systematically recognize, identify, and analyze cause-and-effect relationships in a problem using insights from multiple disciplines (van Merriënboer, 1997; Wang & Song, 2021). The performance levels vary by the complexity of the relationships between causes and effects (Grotzer & Solis, 2015). Students at the basic level can recognize linear cause–and–effect chains, exhibiting the capacity to contemplate fragmented scientific concepts without considering multiple variables. Intermediate–level students are capable of identifying multiple relevant factors from two or more essential disciplinary viewpoints, yet tend to perceive them in isolation. Moreover, advanced students can account for interactions across several disciplines through a multivariate causal model, analyzing how various variables interact within a complex system (Hmelo-Silver & Pfeffer, 2004; Perkins & Grotzer, 2005).

Interpretive Argumentation denotes the ability to articulate and justify claims or explanations by thoroughly examining problems based on disciplinary knowledge (Shen et al., 2015; Wolfe & Haynes, 2003). The hierarchy of capability in this dimension progresses from basic assertions to composite arguments, illustrating how students at different proficiency levels engage with different scientific ideas, correlate evidence with these multidimensional perspectives, and present cohesive and reasoned arguments (Osborne et al., 2016; Sampson & Clark, 2008; Voss et al., 1993). At the advanced level, students construct well–reasoned and logical arguments supported by relevant and accurate scientific concepts across disciplines (Zohar & Nemet, 2002). Intermediate–level students, however, articulate claims utilizing various evidence sources and alternative disciplinary viewpoints, yet they may demonstrate flaws in the interactions and logic between claims and evidence (Lee et al., 2014). Furthermore, students at the elementary level can only justify explanations using isolated pieces of scientific knowledge, showing a lack of depth and coherence in their argumentation (Sandoval & Millwood, 2005).

Systematic integration signifies being capable to connect and apply varied disciplinary insights to produce novel ideas or attain a new understanding from integrative perspectives (Mansilla & Duraising, 2007; Tripp & Shortlidge, 2019). Disciplinary coordination places significant cognitive demands on students (Shen et al., 2014). The development of this learning continuum requires moving from the acquisition of isolated, single–discipline knowledge to recognizing interaction among disciplines, ultimately forming a holistic cognitive framework of the interdisciplinary phenomena (Ivanitskaya et al., 2002; Newell, 2001). At the basic levels, students create solutions from several disciplines but treat each independently, reflecting a compartmentalized disciplinary understanding. As students progress to the intermediate level, they can grasp and make connections between disciplinary insights to address specific target problems, although the integration may remain surface–level (Gouvea et al., 2013). Finally, students at advanced level demonstrate deep integration, applying facts and principles from multiple disciplines to address new complex interdisciplinary issues with an internalization of integrated perspective (Nikitina, 2005).

Critical reflection involves comparing and evaluating the means and limitations of disciplinary knowledge integration at a metadisciplinary level, which is also the key for transferring interdisciplinary knowledge (Howlett et al., 2016; Mansilla & Duraising, 2007; Mansilla, 2005; Schijf et al., 2023). The developmental trajectory in this domain emphasizes the shift from perceiving knowledge as absolute to growing a critical awareness that acknowledges the uncertainty and complexity among a range of disciplines, progressing towards evaluative epistemology (Kuhn, 1999; Mezirow, 1990). Students become aware of the inherent limitations within disciplinary assertions at the fundamental level. Intermediate-level students, as they advance, compare and reflect on intersections and distinctions among disciplines while exploring alternative solutions. In addition, the advanced–level students can critically reevaluate the processes and constraints associated with the integration of diverse disciplinary perspectives, thereby driving a more profound cognitive enhancement (Lai, 2011; Tripp & Shortlidge, 2020).

https://doi.org/10.33225/jbse/24.23.53

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Research Methodology

Instrument Development

Following Wilson's construct modeling approach (Wilson, 2005), a group of seven members (comprising two professors in science education, two experienced science educators, and three doctoral candidates in science education) assembled to develop interdisciplinary thinking items for the lower-secondary level. Previous studies emphasize that the sense of place generated through authentic real-world tasks is essential for promoting interdisciplinary learning outcomes and establishing connections among scientific concepts (Alagona & Simon, 2010; Rahm, 2002; Tan & So, 2019). Accordingly, four interdisciplinary assessment contexts pertaining to real–world issues in interdisciplinary science were identified, including global climate change, ecosystems and biodiversity, agriculture and food science, as well as science and technology. Within these contexts, a series of constructed–response items were created, as such tasks encourage open-mindedness and are more efficient for examining cross–disciplinary relevance than multiple-choice formats (Gouvea et al., 2013).

Several guiding principles informed the development of constructed–response items: (1) Although interdisciplinary thinking involves integrating knowledge across multiple domains, each item was designed to align with students' existing disciplinary knowledge (You et al., 2018). This alignment was achieved through a thorough analysis of the Compulsory Science Curriculum Standards in mainland China to ensure that the content of each question matched what students had already learned (e.g., Biology, Chemistry, Geography) (Ministry of Education of China, 2022); (2) The items were crafted to motivate students to synthesize and utilize their disciplinary expertise in tackling real–world problems instead of merely reproducing factual information (Tan & So, 2019); (3) Each item was structured to facilitate connection among different scientific concepts, promoting students to move beyond isolated facts and consider broader relationships among disciplines; (4) Items were tailored to address specific performance levels within the interdisciplinary thinking construct map, with a range of complexity from fundamental knowledge integration to advanced problem-solving and evaluation. This allowed students to demonstrate their abilities across a spectrum of difficulty levels. A sample task of the instrument is presented in Figure 1. The "Evolutionary History of Giant Pandas" task presented a contextualized ecological problem, which was designed to target the dimension of systematic integration in interdisciplinary thinking. This competency dimension emphasized the ability to connect and integrate knowledge across different disciplines to attain new insights. Specifically, the task required students to integrate concepts from biological evolution, ecological adaptability, climate change and the impact of human activities on ecosystems. They were challenged to synthesize viewpoints from these disciplinary perspectives to comprehensively comprehend the evolution of panda from a carnivorous diet to primarily bamboo-based.

These principles accounted for the cognitive development and prior knowledge of seventh graders to ensure the items were accessible, while simultaneously challenging them to think beyond individual disciplines rather than simply memorizing facts. To ensure content validity, the instrument was examined through expert reviews, with feedback from science education experts and experienced science teachers. The initial version of the assessment tool included seven authentic scenarios with a total of 23 items, each intended to examine seventh graders' interdisciplinary thinking.

Figure 1
A Sample Item T2-2 within the "The Evolutionary History of Giant Pandas" Task

T2-2. Giant pandas, known as "living fossils", are classified within the order Carnivora in systematic taxonomy. Nevertheless, their nutrition has evolved to be highly specialized, primarily relying on bamboo.

What aspects may have contributed to the evolution of dietary shift of giant pandas from carnivorous to predominantly bamboo-based?

https://doi.org/10.33225/jbse/25.24.53

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Scoring Rubrics

Given the open-ended nature of the tasks, partial credit scoring was employed to assign points, enabling recognition of responses that demonstrated incomplete yet accurate elements. Items were rated using 0-2 and 0-3 scales, with detailed descriptors outlining each performance level. M was designated for an empty response. The rubric was developed to assess student responses regarding specified levels, differentiating answers based on their accuracy and completeness. An initial draft of the rubric was evaluated by a group of five experts in educational assessment. Their feedback was incorporated to refine the clarity of score descriptors and ensure that the rubric effectively represented the essential components of interdisciplinary thinking.

Multidimensional Rasch Analysis

The Rasch measurement allows for testing the structure validity of empirical response data and can convert raw scores into interval measures, simultaneously measuring latent traits by accounting for item difficulty and student ability (Boone, 2016; Bond & Fox, 2015). The interdisciplinary thinking was characterized as a multifaceted competence comprising four dimensions, with the calibration of these dimensions occurring concurrently. Specifically, the multidimensional random coefficients multinomial logit (MRCML) model was employed to evaluate the reliability and construct validity of the measurement instrument and to estimate students' interdisciplinary thinking (Adams et al., 1997). Therefore, this approach facilitates the estimation of item and ability parameters across dimensional clusters and calculates correlations among the four dimensions of interdisciplinary thinking (Ackerman et al., 2003; Briggs & Wilson, 2003). The raw data were converted into logits, enabling item difficulty and student competence to be measured on a unified scale across subdimensions, thus enhancing comparability (Wang et al., 2004).

To evaluate the suitability of the conceptual model of interdisciplinary thinking, the overall fit of unidimensional and four–dimensional models were compared. The partial credit model (PCM) was utilized to fit polytomous scored items in both models (Masters, 1982). Model fit was examined using deviance, Akaike's information criterion (AIC) (Akaike, 1973; 2011), and Bayesian information criterion (BIC) (Schwarz, 1978), where smaller values signify a superior fit (Rijmen, 2010). Moreover, the likelihood ratio test was employed to evaluate the adequacy of two models (Bock & Aitkin, 1981; Kang & Cohen, 2007), with the change in deviance approximately following chi–square distribution, and degrees of freedom reflecting the disparity in the quantity of model parameters estimated (Briggs & Wilson, 2003). The fit of items was examined using mean squares (MNSQs) and T–values. The MNSQs for Outfit and Infit values ranging from 0.7 to 1.3 indicate a good fit (Aryadoust et al., 2021; Wang et al., 2006). The T values, representing standardized MNSQs, were considered acceptable within the range of –2 to +2 (Smith et al., 2008; Wei et al., 2014). In Rasch measurement, the average item difficulty is set at 0. The item–person map, or Wright map, was subsequently created to analyze the relationship between item difficulty and student abilities, with item parameters and student estimates calibrated on the same logit scale in the MRCML model (Liu et al., 2008).

To evaluate measurement invariance by gender, the DIF analysis was performed to examine construct equivalence with respect to twenty-two items (Holland & Wainer, 1993; Lord, 1980). As suggested by previous research, a variance of 0.5 logits or greater in estimations between gender groups was considered indicative of significant DIF (Wang, 2000; Wang et al., 2006). Finally, the expected posterior (EAP) methods were used to estimate student abilities across dimensions (Bock & Aitkin, 1981; Hambleton et al., 1991; Mislevy, 1984). The multidimensional Rasch analysis was conducted using ConQuest version 2.0 software (Wu et al., 2007).

Pilot Test

A convenient sampling method was employed to recruit 120 seventh graders from a lower-secondary school in Jiangsu Province, China. This approach was selected for its cost-effectiveness, ease of access to data, and applicability to both quantitative and qualitative research settings (Emerson, 2015; Etikan et al., 2016). All participants volunteered their time. Since interdisciplinary thinking encompasses a broad framework of expected competencies, students in this study had already completed relevant science knowledge (e.g., Biology, Chemistry, Geography) and possessed a certain level of scientific proficiency. To provide ample time for task completion, no time limit was set during test administration. Subsequent to the pilot test, semistructured interviews were administered to collect

https://doi.org/10.33225/jbse/24.23.53

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

student feedback regarding item difficulty and to identify any misunderstanding arising from ambiguities in the phrasing or sentences of the tasks. Moreover, think-aloud protocols were utilized with a select group of students to elucidate their cognitive processes during task completion. This feedback guided modifications to the task design and rubrics, ensuring coherence with the intended learning outcomes.

Item fit statistics were examined by multidimensional Rasch analysis for the student responses. According to the results, one item with inadequate discrimination was eliminated, and two items were restructured to improve clarity, drawing on feedback from both students and experts. The final instrument comprised 22 extended-response items. Table 1 presents an overview of interdisciplinary contexts, task design, along with items.

Table 1
The Contexts and Tasks of Instrument Items in Field Test

Interdisciplinary context	Design of tasks	Items and level
Global Climate Change	Task1: Desert Locust Plague	T1–1 (L3), T1–2 (L1), T1–3 (L2), T1–4(L1), T1–5 (L1)
	Task7: Carbon Neutral Action	T7–1 (L2), T7–2 (L2), T7–3 (L3)
Ecosystems and Biodiversity	Task2: The Evolutionary History of Giant Pandas	T2–1 (L1), T2–2 (L3), T2–3 (L2)
	Task6: Making Ecological Bottle	T6–1 (L1), T6–2 (L2), T6–3 (L3)
Agriculture and Food Science	Task3: Grape Fermentation	T3–1 (L3), T3–2 (L2)
	Task4: Planting of Astragalus Herb	T4–1 (L2), T4–2 (L2)
Science and Technology	Task5: Research Station in Antarctica	T5–1 (L1), T5–2 (L3), T5–3 (L3), T5–4 (L1)

Note: L1 = Elementary level; L2 = Intermediate level; L3 = Advanced level.

Participants

The field test included 316 seventh graders randomly selected from the same school as in the pilot study, consisting of 159 males and 157 females. The sample size was established based on prior research suggesting that a minimum of 200 would be sufficient (Liu, 2020; MacCallum et al., 1999). This sample size was deemed appropriate for the statistical power required for valid conclusions, taking into account the expected effect size and the complexity of the research design. These students, distinct from those in the pilot test, volunteered for the field test. Each student was assigned 60 minutes to independently complete the assessment in a classroom setting, a duration determined by the average completion time recorded during the pilot phase. All procedures were executed in accordance with the approved ethical guidelines. No risks were associated with students' academic achievement, and anonymity was guaranteed for all participants. Confidentiality and privacy were rigorously maintained throughout the study.

Establishment of Inter-rater Reliability

To ensure consistent and reliable scoring of the constructed-response items, the first author, a graduate student in science education, and an experienced lower-secondary school science educator independently evaluated the items. Prior to coding, the first author conducted an extensive training session that included a detailed presentation of the measurement framework, an explicit explanation of the scoring criteria, and illustrative examples of student responses. Any discrepancies in the rating process were resolved through thorough discussions until an agreement was achieved. The inter-rater reliability was evaluated with Kappa statistics (Kappa = 0.94, $p < .01$), indicating a high level of agreement beyond chance (Landis & Koch, 1977).

https://doi.org/10.33225/jbse/25.24.53

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Research Results

Dimensionality Examination with Rasch model

A Comparison of Unidimensional and Four-Dimensional Models

The data were analyzed using the unidimensional and the between-item four-dimensional PCM, respectively. The overall fit quality between the two models was compared with the final deviance, AIC, and BIC. The model exhibiting lower deviance would be supposed to align more closely with the true model. The four-dimensional model attained lower parameter values compared to the unidimensional model, culminating in a final deviation of 12339.51, an AIC of 12463.51, and a BIC of 12696.37. The difference in deviance between the two models was 432.27 with 9 degrees of freedom. The degrees of freedom present the disparity in the quantity of estimated parameters in the unidimensional and multidimensional models. The likelihood ratio test was performed to compare how well the two models fit together. The degrees of freedom showed how different the models' estimated parameters were, and the change in deviance approximately followed a chi-square distribution. For the chi-square distribution with 9 degrees of freedom, the critical value is 16.92 ($p = .05$). The findings present empirical evidence that the four-dimensional model showed a better fit than the unidimensional model in depicting the dimension structure of interdisciplinary thinking.

Dimension Structure

The covariance and correlation revealed the interrelationship among subdimensions in multidimensional Rasch measurement. Table 2 displays the covariance/correlation matrix for the four dimensions of interdisciplinary thinking, as produced by ConQuest. The values beneath the diagonal represent correlation coefficients, while those above the diagonal are covariances. The correlation values indicated the extent to which the four dimensions of interdisciplinary thinking varied together, whereas covariance values quantified the degree to which changes in one dimension aligned with changes in another. The correlation values varied from 0.738 to 0.871, indicating highly positive relationships. This suggested that the four dimensions are distinct but interrelated components of interdisciplinary thinking.

Table 2
Covariance/Correlation between Dimensions

Dimension	Dimension			
	CA	IA	SI	CR
CA		.369	.465	.286
IA	.744*		.625	.385
SI	.800*	.871*		.459
CR	.738*	.804*	.817*	

*$p < .05$

Reliability

Reliability reflects the consistency of estimates for individual ability and item difficulty in Rasch analysis, indicating the stability of measurement across equivalent evaluations (Bond & Fox, 2015). The expected a posteriori/plausible values (EAP/PV) reliability coefficient, equivalent to Cronbach's α, functions as a crucial indicator of dimensional stability. The EAP/PV reliability values for this study across four domains of interdisciplinary thinking were .74, .75, .77, and .71, respectively. The values demonstrated a moderate level of reliability (Cortina, 1993; Kline, 2000), supporting the robustness of the instrument in assessing seventh graders' interdisciplinary thinking.

https://doi.org/10.33225/jbse/24.23.53

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Construct Validity: Model-data Fit and Wright Map

Model-data Fit

Fit statistics are critical for identifying discrepancies between the empirical data and model assumptions. Robust item fit statistics can demonstrate the construct validity of the assessment. Table 3 provides the fit statistics for all items in the interdisciplinary thinking assessment, including Infit, Outfit MNSQ, and gender DIF. The MNSQs for all items, both unweighted (outfit) and weighted (infit), varied from 0.7 to 1.3. T–values ranged from –2.0 to +2.0 for each item, excluding T4–1. The item difficulty reflects the position of each item on the subjects' ability scale. The findings revealed that the difficulty for all items varied from –1.30 to +1.06. T5–1 was the least challenging item, while T7–3 was the most difficult. The standard error of the item difficulty parameter should approach 0 as closely as possible. The majority of items exhibited errors below 0.09, with the exceptions of T6–3 (0.12), T7–1 (0.13), T7–2 (0.14), and T7–3 (0.14), all of which remained within acceptable thresholds.

The DIF analysis investigated the performance discrepancies between male and female students on identical items, while accounting for variations in their latent trait levels. A difference of 0.5 logits or greater was considered evidence of significant DIF (Wang et al., 2006). As shown in Table 3, no significant DIF were identified. This indicates that male and female groups with comparable coping trait levels would exhibit analogous responses for each item. In summary, these results suggest a satisfactory fit with the four-dimensional Rasch model, thereby affirming the construct validity of the interdisciplinary thinking assessment instrument.

Table 3
Fit Statistics of Items from Multidimensional Rasch Analysis

Item	Estimate	Error	Unweighted fit		Weighted fit		Gender DIF **(M-F)**
			MNSQ	T	MNSQ	T	
T1–1	0.38	0.06	0.99	-0.1	0.98	-0.1	0.28
T2–3	0.14	0.07	0.97	-0.3	0.97	-0.3	0.06
T5–4	-0.21	0.07	0.96	-0.4	0.96	-0.4	-0.06
T6–1	-0.44	0.05	1.15	1.8	1.11	1.6	0.22
T7–1	0.13	0.13	1.00	0.0	1.00	0.0	-0.49
T1–3	0.49	0.07	1.02	0.3	1.02	0.2	0.29
T2–1	-1.18	0.07	0.91	-1.1	0.91	-1.3	0.27
T3–1	0.76	0.06	1.04	0.5	1.03	0.5	-0.38
T5–1	-1.30	0.07	0.95	-0.6	0.95	-0.8	-0.17
T6–2	0.17	0.05	1.03	0.4	1.02	0.4	-0.02
T7–3	1.06	0.14	1.07	0.8	1.05	0.7	-0.31
T1–2	-0.46	0.06	1.08	1.0	1.09	1.2	0.12
T1–5	-0.35	0.07	0.93	-0.8	0.94	-0.6	0.03
T2–2	0.57	0.07	0.96	-0.5	0.96	-0.5	0.04
T4–2	-0.25	0.07	1.12	1.4	1.12	1.5	0.15
T5–3	0.72	0.06	0.95	-0.6	0.94	-0.7	-0.12
T7–2	-0.23	0.14	0.95	-0.6	0.95	-0.6	-0.22
T1–4	-0.55	0.07	0.88	-1.5	0.88	-1.8	0.15
T3–2	-0.28	0.07	0.97	-0.4	0.97	-0.4	-0.05
T4–1	-0.16	0.05	1.29	3.3	1.22	2.9	-0.15
T5–2	0.60	0.07	0.96	-0.4	0.96	-0.5	-0.16
T6–3	0.38	0.12	1.00	0.1	1.03	0.4	0.19

https://doi.org/10.33225/jbse/25.24.53

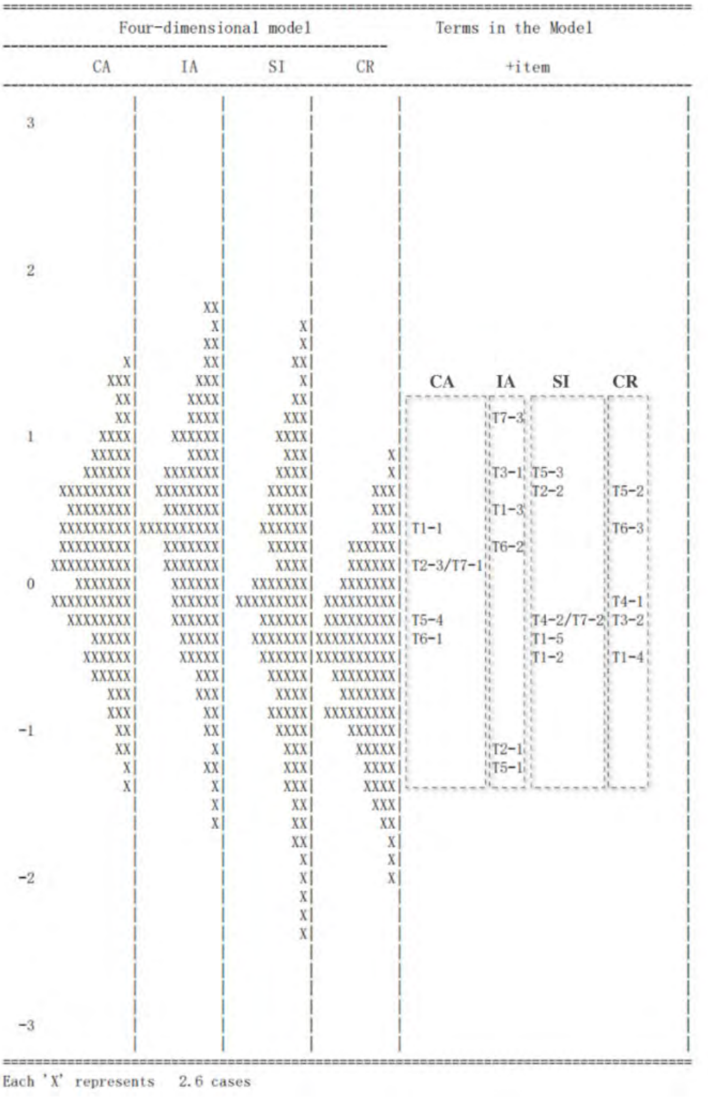
ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

The Wright Map: Distribution of Item Difficulty and Person Ability

The Wright map displays the distribution between estimated item difficulties and student competencies. As shown in Figure 2, the four vertical lines on the left symbolize the four dimensions that constitute interdisciplinary thinking. The left panel depicts the distribution of students' abilities along a continuum from low (bottom) to high (top), with students exhibiting greater proficiency positioned at the top and those with lower–level abilities at the bottom. The right side of the map illustrates item difficulties ranging from easiest (bottom) to most difficult (top) endorsement. In our multidimensional Rasch model analysis, each item is allocated to a singular dimension. The spectrum of students' abilities across four dimensions exhibited a normal distribution, spanning from −2.51 to +1.80, suggesting that the assessment accurately captures a spectrum of competencies. The most challenging item (T7–3) and least difficult item (T5–1) both came from the interpretive argumentation dimension.

In general, the item–person map demonstrated a wide and appropriate spread of both students' coping trait levels and item difficulties, validating that items across all subdimensions were well-targeted to students' coping capabilities. This distribution confirmed the three–level construct for the four dimensions of interdisciplinary thinking, with items covering student ability distributions from elementary to advanced levels. However, students with low proficiency require relatively simple tasks for dimensions of systematic integration and critical reflection.

Figure 2
Wright Map



https://doi.org/10.33225/jbse/24.23.53

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Seventh-grade Students' Performance in Interdisciplinary Thinking

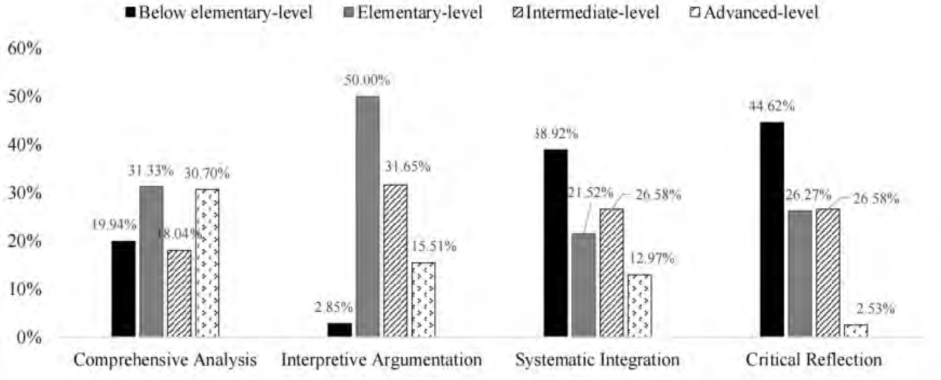
To comprehensively examine students' proficiency in interdisciplinary thinking, the mean item difficulty estimates were computed within each dimension to determine the threshold value regarding each proficiency level, as illustrated in Table 4. Students were deemed to have reached a specific proficiency level when their ability estimates met or exceeded the threshold for that level within a dimension.

Table 4
Cutoff Values for Performance Levels across Dimensions of Interdisciplinary Thinking

Dimensions	Proficiency level	Items and estimates	Cutoff values
Comprehensive Analysis (CA)	Elementary	T6–1(-0.44), T5–4(-0.21)	–0.33
	Intermediate	T2–3(0.14), T7–1(0.13)	0.14
	Advanced	T1–1(0.38)	0.38
Interpretive Argumentation (IA)	Elementary	T2–1(-1.18), T5–1(-1.30)	–1.24
	Intermediate	T1–3(0.49), T6–2(0.17)	0.33
	Advanced	T7–3(1.06), T3–1(0.76)	0.91
Systematic Integration (SI)	Elementary	T1–2(-0.46), T1–5(-0.35)	–0.41
	Intermediate	T4–2(-0.25), T7–2(-0.23)	–0.24
	Advanced	T2–2(0.57), T5–3(0.72)	0.65
Critical Reflection (CR)	Elementary	T1–4(-0.55)	–0.55
	Intermediate	T3–2(-0.28), T4–1(-0.16)	–0.22
	Advanced	T5–2(0.60), T6–3(0.38)	0.49

As presented in Figure 3, student performance across the four dimensions of interdisciplinary thinking showed considerable variation. In the CA dimension, approximately one-third of students reached the elementary level (31.33%) and advanced level (30.70%), while 18.04% achieved intermediate level. Nonetheless, 19.94% of students exhibited proficiency below the elementary level. Conversely, results for the IA dimension revealed that half of the students could perform at the elementary level. The proportions of students achieving the intermediate level and advanced level decreased to 31.65% and 15.51%, respectively. For the SI dimension, merely 12.97% of students attained the advanced level, while a substantial percentage (38.92%) failed to reach the elementary level. Similarly, the findings in the CR dimension indicated that around one-quarter of students achieved the elementary level (26.27%) and intermediate level (26.58%), with the predominant group (44.62%) falling below the elementary level. These results suggest that the four dimensions of interdisciplinary thinking present distinct challenges for students, as evidenced by their varying performance levels.

Figure 3
Distribution of Student Performance Levels across Four Dimensions



https://doi.org/10.33225/jbse/25.24.53

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Discussion

The research aimed to develop and validate an assessment instrument for evaluating interdisciplinary thinking in seventh-grade students. A construct map was structured to delineate the interdisciplinary thinking that encompassed four distinctive dimensions and corresponding performance levels. The multidimensional Rasch analysis demonstrated that the instrument achieved satisfactory reliability and validity according to the PCM within a four-dimensional framework. The findings of the item quality evaluation demonstrated that all items met acceptable model fit. Furthermore, we evaluated students' performance across four subdimensions of interdisciplinary thinking. The subsequent sections discuss these findings regarding the instrument's quality and students' performance in interdisciplinary thinking.

Reliability and Validity of Interdisciplinary Thinking Assessment

The current study concurrently examined dimension structure, reliability, and construct validity of interdisciplinary thinking, which includes the four aspects of comprehensive analysis, interpretive argumentation, systematic integration, and critical reflection. After removing one poor-fitting item, the field study results demonstrated that the remaining 22 items fit well with the multidimensional Rasch model. A better model fit was observed when comparing the unidimensional and four-dimensional models based on PCM, hence supporting the dimension structure of the measurement framework.

The application of the multidimensional PCM accounted for correlations among the domains, revealing that the four dimensions of interdisciplinary thinking are both distinct and interrelated. This multidimensional approach provided more precise estimates of the associations among the four latent traits and significantly improved the utility of the instrument (Adams et al., 1997). The moderate reliability values suggested that the instrument was appropriate for assessing student performance, confirming its ability to differentiate among levels of interdisciplinary thinking, despite the reported values not attaining the higher threshold (Cortina, 1993). Item fit statistics indicated that all items adhered to the accepted thresholds for evaluating fit in Rasch models. The MNSQ values associated with both weighted and unweighted statistics varied between 0.7 and 1.3, with T–values for all items within the range of –2.0 and +2.0, supporting the acceptable degree of fit for item responses. Although item T4–1 exhibited slightly elevated T–values, this deviation was not substantial enough to compromise the overall integrity of the instrument (Wright & Linacre, 1994). Moreover, the DIF analysis confirmed the instrument's measurement invariance across genders, indicating good psychometric validity and no bias issues. The lack of significant DIF is essential for preserving the fairness and dependability of assessments, particularly within increasingly diverse educational environments (Liu et al., 2008). The Wright map illustrated a well-targeted alignment between item difficulty and student aptitude levels, with the distribution of coping abilities and item difficulties across a broad spectrum of the latent trait scale. Nonetheless, the slight gap in the systematic integration and critical reflection dimensions suggested that students with lower ability levels were not being comprehensively evaluated. Therefore, incorporating easier tasks could be organized to provide a more thorough assessment of student competencies across the full spectrum of abilities. In conclusion, these findings improved knowledge of the multifaceted nature of interdisciplinary thinking, while the proof of reliability and validity substantiated that the instrument functioned as a reliable tool for assessing seventh graders' interdisciplinary thinking.

Characteristics of Students' Interdisciplinary Thinking Performance

The findings revealed that most students attained only a fundamental level of cognitive competence in interdisciplinary thinking, which is consistent with prior studies (Shen et al., 2014; Wang & Song, 2021). Moreover, the assessment results identified variations and weaknesses in student performance across the four constructs of interdisciplinary thinking, highlighting the distinct cognitive demands associated with each subcompetency. This corresponds with prior research that emphasized the multifaceted nature of interdisciplinary thinking, wherein students tended to excel in certain skills while encountering difficulties in others (Claus & Wiese, 2019; Spelt et al., 2009). It also suggested that certain dimensions are intrinsically more complex than others in interdisciplinary thinking (Tripp et al., 2020).

https://doi.org/10.33225/jbse/24.23.53

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

While the majority of students demonstrated intermediate or advanced capabilities in comprehensive analysis and interpretive argumentation, they encountered difficulties in meeting the expectations for systematic integration and critical reflection, as indicated by a smaller proportion of individuals achieving basic proficiency or higher in these two constructs compared to comprehensive analysis and interpretive argumentation. In accordance with Richter and Paretti (2009), students had difficulties in integrating interdisciplinary subjects with their own expertise and recognizing contributions of various fields to address complicated problems. According to Tripp et al. (2020), students exhibited superior proficiency in understanding disciplinary grounding compared to knowledge integration or critical awareness, which might be attributed to the latter's inherent ambiguity in the natural sciences (Borrego et al., 2009). Overall, these findings are in line with extensive literature suggesting that students should be instructed to integrate scientific evidence and engage in critical thinking during scientific argumentation when confronting interdisciplinary complexities (Balgopal et al., 2017; Gouvea et al., 2013; Mansilla et al., 2009; Tripp & Shortlidge, 2019; You et al., 2018). For instance, educators could exemplify interdisciplinary learning experience in their teaching to assist students in internalizing interdisciplinary problem-solving mindsets that span multiple disciplines, such as critical thinking (Rowe et al., 2015).

The deficiency in students' interdisciplinary thinking cannot be exclusively ascribed to the constraints of their disciplinary knowledge. As it is shaped by a confluence of factors, including their self–confidence (Zhan et al., 2017), attitudes (Song & Wang, 2021), motivation (Xu et al., 2022), interdisciplinary learning experiences (Busta & Russo, 2020; Knobloch et al., 2020; Shen et al., 2014), learning environment (Spelt et al., 2009; Stentoft, 2017), peer support (Claus & Wiese, 2019), and scientific proficiency (Wang & Song, 2021). Future research ought to further examine the mechanisms underlying students' development of interdisciplinary thinking and investigate the root causes that affect their performance. The combination of qualitative and quantitative data could help elucidate specific barriers and disclose the developmental pathways for competency acquisition. For instance, implementing dynamic assessments by incorporating longitudinal and interview–based methodologies may illuminate how students' such competence evolves over time (Carr et al., 2002; Shen et al., 2014), offering critical insights into instructional strategies aimed at improving interdisciplinary thinking and addressing specific challenges within interdisciplinary contexts.

Conclusions and Implications

This study developed and validated an instrument for assessing interdisciplinary thinking among seventh-grade students within a lower-secondary school science context. The interdisciplinary thinking was conceptualized as a four–dimensional theoretical framework, encompassing a broader scope of competencies, including the elements of comprehensive analysis, interpretive argumentation, systematic integration, and critical reflection. The multidimensional Rasch analysis demonstrated strong psychometric properties across 22 items, confirming the instrument's reliability and construct validity. The findings highlighted specific aspects in which students faced challenges with interdisciplinary cognitive competencies, thereby furnishing educators with quantitative data to better understand and support lower-secondary school students' competency development in school science education. The assessment results yield insights into the developmental characteristics of interdisciplinary thinking among lower-secondary students. It is recommended that faculty adopt this competency model in science classrooms to cultivate students' capacity to think interdisciplinarily by exposing them to in–depth disciplinary knowledge and challenging them to broaden their mindset across various scientific domains. Future research could benefit from longitudinal studies or mixed-method techniques to examine the developmental trajectories of interdisciplinary thinking and identify barriers or factors that influence this competency. Furthermore, the sample in this study may not be broad enough to represent the diversity of all student populations. To boost the statistical robustness and generalizability, it is essential to expand the sample size through a more random selection among a wider range of socioeconomic backgrounds and educational institutions. Consequently, the statistical results will be more persuasive.

https://doi.org/10.33225/jbse/25.24.53

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Declaration of Interest

The authors declare no conflict of interest.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B.N., Csaki, F. (Eds), In *Selected papers of hirotugu akaike* (pp. 199–213). Springer. https://doi.org/10.1007/978-1-4612-1694-0_15

Akaike, H. (2011). Akaike's information criterion. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science (*pp. 25–25). Springer. https://doi.org/10.1007/978-3-642-04898-2_110

Adams, R. J., Wilson, M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1–23. https://doi.org/10.1177/0146621697211001

Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice, 22*(3), 37–51. https://doi.org/10.1111/j.1745-3992.2003.tb00136.x

Alagona, P. S., & Simon, G. L. (2010). The role of field study in humanistic and interdisciplinary environmental education. *Journal of Experiential Education, 32*(3), 191–206. https://doi.org/10.1177/105382590903200302

Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing, 38*(1), 6–40. https://doi.org/10.1177/0265532220927487

Asghar, A., Ellington, R., Rice, E., Johnson, F., & Prime, G. M. (2012). Supporting STEM education in secondary science contexts. *Interdisciplinary Journal of Problem–Based Learning, 6*(2), 85–125. https://doi.org/10.7771/1541-5015.1349

Bock, R., & Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*(4), 443–459. https://doi.org/10.1007/BF02293801

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE—Life Sciences Education*, *15*(4), Article rm4. https://doi.org/10.1187/cbe.16-04-0148

Balgopal, M. M., Wallace, A. M., & Dahlberg, S. (2017). Writing from different cultural contexts: How college students frame an environmental SSI through written arguments. *Journal of Research in Science Teaching, 54*(2), 195–218. https://doi.org/10.1002/tea.21342

Bestelmeyer, S. V., Elser, M. M., Spellman, K. V., Sparrow, E. B., Haan-Amato, S. S., & Keener, A. (2015). Collaboration, interdisciplinary thinking, and communication: new approaches to K–12 ecology education. *Frontiers in Ecology and the Environment, 13*(1), 37–43. https://doi.org/10.1890/140130

Borrego, M., Newswander, C. B., McNair, L. D., McGinnis, S., & Paretti, M. C. (2009). Using concept maps to assess interdisciplinary integration of green engineering knowledge. *Advances in Engineering Education*, *1*(3), 1–26.

Borrego, M., & Newswander, C. B. (2010). Definitions of interdisciplinary research: Toward graduate-level interdisciplinary learning outcomes. *Review of Higher Education, 34*(1), 61–84. https://doi.org/10.1353/rhe.2010.0006

Brassler, M., & Dettmers, J. (2017). How to enhance interdisciplinary competence—interdisciplinary problem-based learning versus interdisciplinary project-based learning. *Interdisciplinary Journal of Problem–Based Learning, 11*(2). https://doi.org/10.7771/1541-5015.1686

Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement, 4*(1), 87–100.

Busta, L., & Russo, S. E. (2020). Enhancing interdisciplinary and systems thinking with an integrative plant chemistry module applied in diverse undergraduate course settings. *Journal of Chemical Education, 97*(12), 4406–4413. https://doi.org/10.1021/acs.jchemed.0c00395

Carr, M., & Claxton, G. (2002). Tracking the development of learning dispositions. *Assessment in Education: Principles, Policy & Practice, 9*(1), 9–37. https://doi.org/10.1080/09695940220119148

Chin, C., & Chia, L. G. (2006). Problem-based learning: Using ill–structured problems in biology project work. *Science Education, 90*(1), 44–67. https://doi.org/10.1002/sce.20097

Claus, A. M., & Wiese, B. S. (2019). Development and test of a model of interdisciplinary competencies. *European Journal of Work and Organizational Psychology, 28*(2), 191–205. https://doi.org/10.1080/1359432x.2019.1567491

Clemmons, A. W., Timbrook, J., Herron, J. C., & Crowe, A. J. (2020). BioSkills guide: Development and national validation of a tool for interpreting the vision and change core competencies. *CBE–Life Sciences Education, 19*(4), Article ar53. https://doi.org/10.1187/cbe.19-11-0259

Cortina, J. M. (1993). What is coefficient alpha - an examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Cowden, C. D., & Santiago, M. F. (2016). Interdisciplinary explorations: Promoting critical thinking via problem-based learning in an advanced biochemistry class. *Journal of Chemical Education, 93*(3), 464–469. https://doi.org/10.1021/acs.jchemed.5b00378

https://doi.org/10.33225/jbse/24.23.53

Emerson, R. W. (2015). Convenience sampling, random sampling, and snowball sampling: How does sampling affect the validity of research? *Journal of Visual Impairment & Blindness*, *109*(2), 164–168. https://doi.org/10.1177/0145482x1510900215
Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11
Gao, X., Li, P., Shen, J., & Sun, H. (2020). Reviewing assessment of student learning in interdisciplinary STEM education. *International Journal of Stem Education, 7*(1), 1–14. https://doi.org/10.1186/s40594-020-00225-4
Gardiner, P. (2020). Learning to think together: Creativity, interdisciplinary collaboration and epistemic control. *Thinking Skills and Creativity, 38,* Article 100749. https://doi.org/10.1016/j.tsc.2020.100749
Gouvea, J. S., Sawtelle, V., Geller, B. D., & Turpen, C. (2013). A framework for analyzing interdisciplinary tasks: Implications for student learning and curricular design. *CBE–Life Sciences Education, 12*(2), 187–205. https://doi.org/10.1187/cbe.12-08-0135
Grotzer, T. A., & Solis, S. L. (2015). Action at an attentional distance: A study of children's reasoning about causes and effects involving spatial and attentional discontinuity. *Journal of Research in Science Teaching, 52*(7), 1003–1030. https://doi.org/10.1002/tea.21233
Holland, P. W. & Wainer, H. (1993). *Differential item functioning*. Erlbaum.
Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Sage.
Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science, 28*(1), 127–138. https://doi.org/10.1207/s15516709cog2801_7
Howlett, C., Ferreira, J.-A., & Blomfield, J. (2016). Teaching sustainable development in higher education Building critical, reflective thinkers through an interdisciplinary approach. *International Journal of Sustainability in Higher Education, 17*(3), 305–321. https://doi.org/10.1108/ijshe-07-2014-0102
Illeris, K. (2003). Towards a contemporary and comprehensive theory of learning. *International Journal of Lifelong Education, 22*(4), 396–406. https://doi.org/10.1080/02601370304837
Ivanitskaya, L., Clark, D., Montgomery, G., & Primeau, R. (2002). Interdisciplinary learning: Process and outcomes. *Innovative Higher Education, 27*(2), 95–111. https://doi.org/10.1023/A:1021105309984
Jayathilaka, A. S. (2021). *Development of a novel instrument to evaluate interdisciplinary problem-solving skills of Chemistry and Biochemistry Majors.* Master›s thesis, University of Minnesota Duluth.
Kline, P. (2000). *Handbook of psychological testing*. Routledge.
Kang, T., & Cohen, A. S. (2007). IRT model selection methods for dichotomous items. *Applied Psychological Measurement, 31*(4), 331–358. https://doi.org/10.1177/0146621606292213
Knobloch, N. A., Charoenmuang, M., Cooperstone, J. L., & Patil, B. S. (2020). Developing interdisciplinary thinking in a food and nutritional security, hunger, and sustainability graduate course. *Journal of Agricultural Education & Extension, 26*(1), 113–127. https://doi.org/10.1080/1389224x.2019.1690014
Kuhn, D. (1999). A developmental model of critical thinking. *Educational researcher, 28*(2), 16–46. https://www.jstor.org/stable/1177186
Lord, F. M. (1980). *Applications of item response to theory to practical testing problems*. Routledge.
Lai, E. R. (2011). Critical thinking: A literature review. *Pearson's Research Reports, 6*(1), 40–41.
Lan, Y., Chi, S., & Wang, Z. (2021). Assessing upper-secondary school students' interdisciplinary understanding of environmental issues. *Journal of Baltic Science Education, 20*(5), 811–826. https://doi.org/10.33225/jbse/21.20.811
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310
Lattuca, L. R., Knight, D., & Bergom, I. (2013). Developing a measure of interdisciplinary competence. *International Journal of Engineering Education, 29*(3), 726–739. https://www.ijee.ie/articles/Vol29-3/19_ijee2724ns.pdf
Lattuca, L. R., Knight, D. B., Ro, H. K., & Novoselich, B. J. (2017). Supporting the development of engineers' interdisciplinary competence. *Journal of Engineering Education, 106*(1), 71–97. https://doi.org/10.1002/jee.20155
Lattuca, L. R., Voight, L. J., & Fath, K. Q. (2004). Does interdisciplinarity promote learning? Theoretical support and researchable questions. *Review of Higher Education, 28*(1), 23–48. https://doi.org/10.1353/rhe.2004.0028
Lee, H.-S., Liu, O. L., Pallant, A., Roohr, K. C., Pryputniewicz, S., & Buck, Z. E. (2014). Assessment of uncertainty-infused scientific argumentation. *Journal of Research in Science Teaching, 51*(5), 581–605. https://doi.org/10.1002/tea.21147
Liu, O. L., Wilson, M., & Paek, I. (2008). A multidimensional Rasch analysis of gender differences in PISA mathematics. *Journal of Applied Measurement, 9*(1), 18–35.
Liu, X. F. (2020). *Using and developing measurement instruments in science education: A Rasch modeling approach*. Information Age Publishing.
MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99. https://doi.org/10.1037//1082-989x.4.1.84
Mansilla, V. B., & Duraising, E. D. (2007). Targeted assessment of students' interdisciplinary work: An empirically grounded framework proposed. *Journal of Higher Education, 78*(2), 215–237. https://doi.org/10.1080/00221546.2007.11780874
Mansilla, V. B., Duraisingh, E. D., Wolfe, C. R., & Haynes, C. (2009). Targeted assessment rubric: An Empirically grounded rubric for interdisciplinary writing. *Journal of Higher Education, 80*(3), 334–353. https://doi.org/10.1080/00221546.2009.11779016

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Mansilla, V. B. (2005). Assessing student work at disciplinary crossroads. *Change: The Magazine of Higher Learning, 37*(1), 14–21. https://doi.org/10.3200/chng.37.1.14-21

Markauskaite, L., Schwarz, B., Damsa, C., & Muukkonen, H. (2024). Beyond disciplinary engagement: Researching the ecologies of interdisciplinary learning. *Journal of the Learning Sciences, 33*(2), 213–241. https://doi.org/10.1080/10508406.2024.2354151

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174. https://doi.org/10.1007/bf02296272

Melton, J. W., Saiful, J. A., & Shein, P. P. (2022). Interdisciplinary STEM program on authentic aerosol science research and students' systems thinking approach in problem–solving. *International Journal of Science Education, 44*(9), 1419–1439. https://doi.org/10.1080/09500693.2022.2080886

Mezirow, J. (1990). *Fostering critical reflection in adulthood: A guide to transformative and emancipatory learning*. Jossey-Bass.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49*(3), 359–381. https://doi.org/10.1007/bf02306026

Ministry of Education of the People's Republic of China, (2022). *Compulsory Education Science Curriculum Standards*. People's Education Press.

Newell, W. (2001). A theory of interdisciplinary studies. *Issues in Interdisciplinary Studies*. *19*, 1–25.

National Research Council, (2012), *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. National Academies Press.

Newell, W. H., & Luckie, D. B. (2019). Pedagogy for interdisciplinary habits of mind. *Journal of Interdisciplinary Studies in Education*, *8*(1), 6–20. https://doi.org/10.32674/jise.v8i1.584

Nikitina, S. (2005). Pathways of interdisciplinary cognition. *Cognition and Instruction, 23*(3), 389–425. https://doi.org/10.1207/s1532690xci2303_3

Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S.-Y. (2016). The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching, 53*(6), 821–846. https://doi.org/10.1002/tea.21316

Perkins, D. N., & Grotzer, T. A. (2005). Dimensions of causal understanding: The role of complex causal models in students' understanding of science. *Studies in Science Education*, *41*(1), 117–165. https://doi.org/10.1080/03057260508560216

Rahm, J. (2002). Emergent learning opportunities in an inner-city youth gardening program. *Journal of Research in Science Teaching, 39*(2), 164–184. https://doi.org/10.1002/tea.10015

Repko, A. (2008). *Interdisciplinary research: Process and theory.* Sage.

Richter, D. M., & Paretti, M. C. (2009). Identifying barriers to and outcomes of interdisciplinarity in the engineering classroom. *European Journal of Engineering Education, 34*(1), 29–45. https://doi.org/10.1080/03043790802710185

Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement, 47*(3), 361–372. https://doi.org/10.1111/j.1745-3984.2010.00118.x

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education, 92*(3), 447–472. https://doi.org/10.1002/sce.20276

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23–55. https://doi.org/10.1207/s1532690xci2301_2

Schijf, J. E., van der Werf, G. P. C., & Jansen, E. (2023). Measuring interdisciplinary understanding in higher education. *European Journal of Higher Education, 13*(4), 429–447. https://doi.org/10.1080/21568235.2022.2058045

Scott, C. L. (2015). The futures of learning 2: What kind of learning for the 21st century? In education research and foresight working papers. *UNESCO,* 1–9. https://unesdoc.unesco.org/ark:/48223/pf0000242996

Shen, J., Liu, O. L., & Sung, S. (2014). Designing interdisciplinary assessments in sciences for college students: An example on osmosis. *International Journal of Science Education, 36*(11), 1773–1793. https://doi.org/10.1080/09500693.2013.879224

Shen, J., Sung, S., & Zhang, D. (2015). Toward an analytic framework of interdisciplinary reasoning and communication (IRC) processes in science. *International Journal of Science Education, 37*(17), 2809–2835. https://doi.org/10.1080/09500693.2015.1106026

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology, 8*, 33. https://doi.org/10.1186/1471-2288-8-33

Song, G., & Wang, Z. (2021). Factors influencing middle school students' interdisciplinary competence in science education. *Journal of Research in Science Teaching, 58*(7), 1041–1072. https://doi.org/10.1002/tea.21692

Spelt, E. J. H., Biemans, H. J. A., Tobi, H., Luning, P. A., & Mulder, M. (2009). Teaching and learning in interdisciplinary higher education: A systematic review. *Educational Psychology Review, 21*(4), 365–378. https://doi.org/10.1007/s10648-009-9113-z

Spelt, E. J. H., Luning, P. A., van Boekel, M. A. J. S., & Mulder, M. (2017). A multidimensional approach to examine student interdisciplinary learning in science and engineering in higher education. *European Journal of Engineering Education, 42*(6), 761–774. https://doi.org/10.1080/03043797.2016.1224228

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Stentoft, D. (2017). From saying to doing interdisciplinary learning: Is problem-based learning the answer? *Active Learning in Higher Education, 18*(1), 51–61. https://doi.org/10.1177/1469787417693510

Tan, E., & So, H. J. (2019). Role of environmental interaction in interdisciplinary thinking: From knowledge resources perspectives. *Journal of Environmental Education, 50*(2), 113–130. https://doi.org/10.1080/00958964.2018.1531280

Thomas, I. (2009). Critical thinking, transformative learning, sustainable education, and problem-based learning in universities. *Journal of Transformative Education, 7*(3), 245–264. https://doi.org/10.1177/1541344610385753

Tripp, B., & Shortlidge, E. E. (2019). A framework to guide undergraduate education in interdisciplinary science. *CBE–Life Sciences Education, 18*(2), Article es3. https://doi.org/10.1187/cbe.18-11-0226

Tripp, B., & Shortlidge, E. E. (2020). From theory to practice: Gathering evidence for the validity of data collected with the interdisciplinary science rubric (IDSR). *CBE–Life Sciences Education, 19*(3), Article ar33. https://doi.org/10.1187/cbe.20-02-0035

Tripp, B., Voronoff, S. A., & Shortlidge, E. E. (2020). Crossing boundaries: Steps toward measuring undergraduates' interdisciplinary science understanding. *CBE–Life Sciences Education, 19*(1), Article ar8. https://doi.org/10.1187/cbe.19-09-0168

Valley, W., Wittman, H., Jordan, N., Ahmed, S., & Galt, R. (2018). An emerging signature pedagogy for sustainable food systems education. *Renewable Agriculture and Food Systems, 33*(5), 467–480. https://doi.org/10.1017/s1742170517000199

Van den Besselaar, P., & Heimeriks, G. (2001). Disciplinary, multidisciplinary, interdisciplinary: Concepts and indicators. In M. Davis & C.S. Wilson (Eds.), *Proceedings of the 8th International conference on scientometrics and informetrics—ISSI 2001* (pp. 705–716). University of New South Wales.

van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four–component instructional design model for technical training*. Educational Technology Publications.

Voss, J. F., Fincher-Kiefer, R., Wiley, J., & Silfies, L. N. (1993). On the processing of arguments. *Argumentation, 7*(2), 165–181. https://doi.org/10.1007/bf00710663

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates.

Wang, W.-C., Yao, G., Tsai, Y.-J., Wang, J.-D., & Hsieh, C.-L. (2006). Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-BREF using multidimensional Rasch analysis. *Quality of Life Research, 15*(4), 607–620. https://doi.org/10.1007/s11136-005-4365-7

Wang, W. C., Chen, P. H., & Cheng, Y. Y. (2004). Improving measurement precision of test batteries using multidimensional item response models. *Psychological Methods, 9*(1), 116–136. https://doi.org/10.1037/1082-989x.9.1.116

Wang, W. C. (2000). Modeling effects of differential item functioning in polytomous items. *Journal of Applied Measurement, 1*(1), 63–82.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest version 2.0: Generalised item response modelling software*. Australian Council for Educational Research.

Wang, Z., & Song, G. (2021). Towards an assessment of students' interdisciplinary competence in middle school science. *International Journal of Science Education, 43*(5), 693–716. https://doi.org/10.1080/09500693.2021.1877849

Wei, S., Liu, X., & Jia, Y. (2014). Using Rasch measurement to validate the instrument of students' understanding of models in science (sums). *International Journal of Science and Mathematics Education, 12*(5), 1067–1082. https://doi.org/10.1007/s10763-013-9459-z

Wolfe, C. R. (2011). Argumentation across the curriculum. *Written Communication, 28*(2), 193–219. https://doi.org/10.1177/0741088311399236

Wolfe, C. R., & Haynes, C. (2003). Interdisciplinary writing assessment profiles. *Issues in Integrative Studies, 21*, 126–170.

Xu, C., Wu, C.-F., Xu, D.-D., Lu, W.-Q., & Wang, K.-Y. (2022). Challenges to student interdisciplinary learning effectiveness: An empirical case study. *Journal of Intelligence, 10*(4), 1–32. https://doi.org/10.3390/jintelligence10040088

You, H. S., Marshall, J. A., & Delgado, C. (2018). Assessing students' disciplinary and interdisciplinary understanding of global carbon cycling. *Journal of Research in Science Teaching, 55*(3), 377–398. https://doi.org/10.1002/tea.21423

Zhan, Y., So, W. W. M., & Cheng, I. N. Y. (2017). Students' beliefs and experiences of interdisciplinary learning. *Asia Pacific Journal of Education, 37*(3), 375–388. https://doi.org/10.1080/02188791.2017.1301880

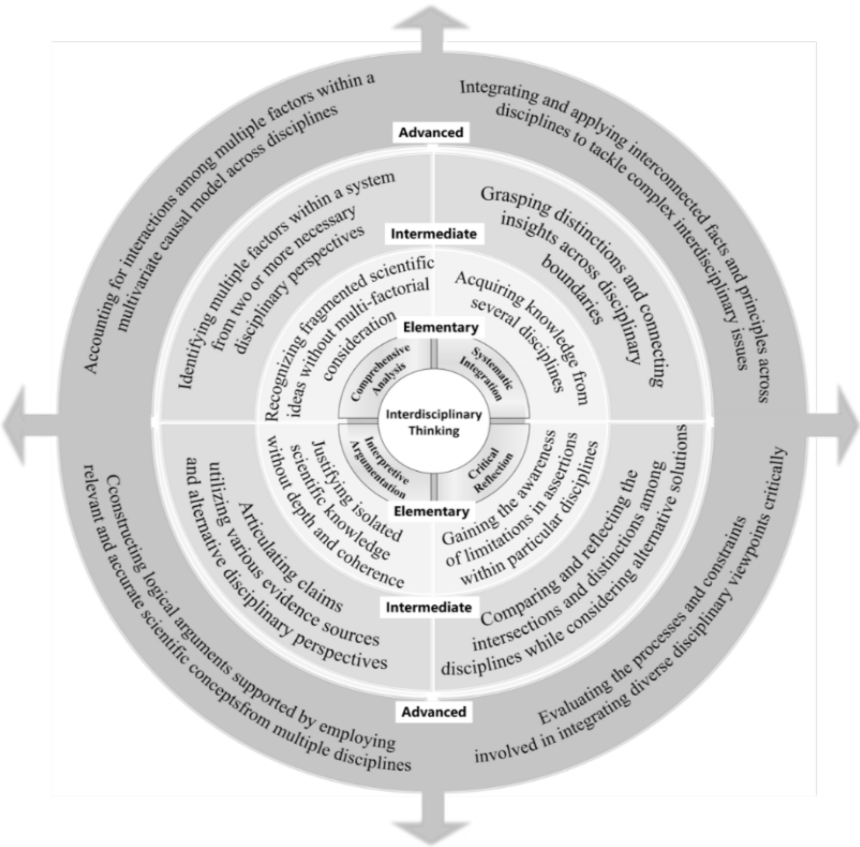
Rowe, M. P., Gillespie, B. M., Harris, K. R., Koether, S. D., Shannon, L.-J. Y., & Rose, L. A. (2015). Redesigning a general education science course to promote critical thinking. *CBE–Life Sciences Education, 14*(3), 1–12. https://doi.org/10.1187/cbe.15-02-0032

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39*(1), 35–62. https://doi.org/10.1002/tea.10008

Journal of Baltic Science Education, Vol. 24, No. 1, 2025
ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Appendix A

The Four-Dimensional Framework and Competency Levels of Interdisciplinary Thinking



Received: *November 13, 2024* Revised: *January 31, 2025* Accepted: *February 05, 2025*

Cite as: Chen, S., Chi, S., & Wang, Z. (2025). Interdisciplinary thinking among seventh-grade students in lower-secondary science education. *Journal of Baltic Science Education*, *24*(1), 53–70. https://doi.org/10.33225/jbse/25.24.53

Shasha Chen MEd, (Science Education), PhD Candidate, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai, P. R. China.
E-mail: sschen016@126.com
ORCID: https://orcid.org/0009-0004-5193-3772

Shaohui Chi *(Corresponding author)* PhD, Professor, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai, P. R. China.
E-mail: charlcy@163.com
ORCID: https://orcid.org/0000-0002-1261-7951

Zuhao Wang PhD, Professor, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai, P. R. China.
E-mail: wangzuhao@126.com
ORCID: https://orcid.org/0000-0002-2967-8232