JOURNAL OF APPLIED MEASUREMENT, *25*(1/2), 132-149

An Examination of an Epistemic Cognition Developmental Progression: A Rasch Analysis

Man Ching Esther Chan
The University of Melbourne

Mark Wilson
University of California, Berkeley
The University of Melbourne

In the era of post-truth and misinformation, there are continuing calls for an emphasis on critical thinking in school education. Epistemic cognition has been proposed as foundational to critical thinking inside and outside of the classroom. However, due to a lack of understanding of the construct and its development, teachers are not well equipped to foster effective epistemic cognition and thus critical thinking in the classroom. Drawing from previous literature, an assessment tool was piloted and subsequently administered to 168 Year 8 students (13- to 14-year-olds) in Melbourne, Australia, to examine the theorized development of epistemic cognition. Students' responses were examined qualitatively using think-aloud protocols and quantitatively using classical test theory and Rasch modelling. The instrument showed good person-separation reliability (.76) and Cronbach's alpha (.79). Based on the analysis, the student responses generally aligned with the theorized construct map, demonstrating strong construct validity. The findings offer empirical evidence for a developmental progression in epistemic cognition, which may be used to inform teaching of critical thinking.

Keywords: epistemic cognition, developmental progression, item response theory, test development, psychometrics

Dr. Man Ching Esther Chan https://orcid.org/0000-0001-6066-0738
Dr. Mark Wilson https://orcid.org/0000-0002-0425-5305
This research was funded by the University of Melbourne Early Career Researcher Grant (Grant ID 1757459). The authors wish to thank the participants of the study and the contribution of Professor David Clarke and Mrs. Jaya Masaldan for their assistance in conducting this study, and Dr. Amy Arneson for her technical support.
Requests for reprints should be sent to Man Ching Esther Chan, the University of Melbourne, Australia; e-mail: mc.chan@unimelb.edu.au

Professor Wen-Chung Wang is well recognized for his work to improve psychometric theories and methods (Wang & Wilson, 2005; Wang & Yeh, 2003; Wang et al., 2004) and the application of measurement models in various contexts (Hallinger et al., 2016; Hsueh et al., 2004; Lin et al., 2019). This paper contributes to the special issue in memory of Professor Wang by providing an account of an empirical study in Australia, which investigated the measurement of epistemic cognition in the school setting.

In the era of post-truth, misinformation, and the rise of generative artificial intelligence (genAI), there are continuing calls for an emphasis on critical thinking in school education (Kendeou et al., 2019; LaGarde & Hudgins, 2018; Mejia & Sargent, 2023). *Critical thinking* can be understood as a person's purposeful reflection and reasoning about what to believe in or act on when faced with complex issues (Ennis, 1987, 2018). It is seen as an important learning goal in various countries. In the United States, critical thinking is considered part of the scientific practices in the framework underpinning the Next Generation Science Standards (NGSS; National Research Council, 2012). In Australia, critical thinking, along with creative thinking, is one of the general capabilities in the national curriculum (Australian Curriculum Assessment and Reporting Authority, 2023). Due to the complexity of the construct, critical thinking has been seen as a "major unsolved problem of pedagogy" (Kuhn & Dean, 2004, p. 269). Even in recent times, critical thinking has been found to be not clearly understood and operationalized within school settings globally (Vincent-Lancrin et al., 2019), despite being one of the so-called 21st Century Skills (Griffin et al., 2012).

One first step to understand and therefore teach critical thinking is to identify the metacognition that critical thinking draws from (Greene et al., 2016; Kuhn & Dean, 2004). Critical thinking has been proposed to require *epistemic cognition*, which involves dispositions, beliefs, and skills about how individuals decide what they actually know, as opposed to what they believe, question, or trust (Greene & Yu, 2016). Epistemic cognition can thus be seen as foundational to critical thinking both inside and outside of the classroom.

This paper reports on the development and validation of a tool for assessing epistemic cognition for Year 8 students (13- to 14-year-olds) in a secondary school in Melbourne, Australia. The assessment tool was designed to examine a theorized developmental progression of epistemic cognition and the developmental levels of the students before they undertake a learning inquiry program in Year 9 on sustainability issues. This paper describes the development of the assessment tool for identifying the learning needs of the students for informing the learning inquiry program. As will be illustrated, the assessment tool development approach appeared useful for measuring the construct of interest and for informing teaching through assessment. Further background of the study is provided in the next section.

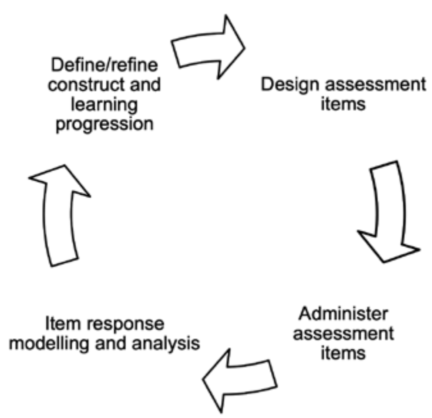
Study Background

This project arose from a learning inquiry program in Australia designed for Year 9 students (14- to 15-year-olds). The program aimed to promote 21st Century Skills (Binkley et al., 2012) and encourage students to make informed choices and positive actions in the context of sustainable living. During the program, students embarked on an intensive four-week period of research and investigation in order to answer the question: "How can we think globally and act locally for a sustainable future?" In teams and small groups, the students were expected to learn to generate relevant questions and seek information to answer those questions while exploring topics such as production and consumption, energy, transport, people and population, and biodiversity. At the end of the four-week inquiry cycle, students presented their understanding of the topics at an exhibition that was open to the public.

An assessment tool was developed to examine a theorized epistemic cognition developmental progression and assist the school with determining the developmental levels of the Year 8 students before they participated in the inquiry program in Year 9, so that the school can better tailor the program to the learning needs of the students. The assessment tool development approach drew upon the work of Wilson (2023) and comprised an iterative process of defining the construct of interest, differentiating the different levels of performance, creating items that target the different performance levels, collecting assessment responses, analyzing the data using a psychometric model, and comparing the results with the initially hypothesized performance levels (see Figure 1). In the context of this study, this involved defining the construct of epistemic cognition and differentiating different levels of performance that align with the construct definition, and designing assessment items and a scoring scheme that target and differentiate the different performance levels of epistemic cognition. The assessment tool was piloted, refined, and administered to Year 8 students. The student assessment responses were subsequently analyzed using classical test theory and item response theory to evaluate their correspondence with the theorized performance levels of epistemic cognition.

Figure 1

The Assessment Tool Development Process



Note. Adapted from "Constructing Measures: An Item Response Modeling Approach (2nd ed.)," by M. Wilson, 2023, *Routledge*.

The approach allowed the construct of interest to be operationalized and translated as differentiated levels of student performance, and also structured the investigation of validity and reliability evidence for the instrument (American Educational Research Association et al., 2014). The approach connects instruction with assessment and student learning by using a criterion-referenced framework to articulate student performance (Black et al., 2011; Lehrer et al., 2014). The next section explains the theory underpinning the approach.

Connecting Learning, Teaching, and Assessment

According to the National Research Council, assessment is "a process of reasoning from evidence" (Pellegrino et al., 2001, p. 43). Assessment involves making inferences about the internal functions of people based on observations of their behaviors. The understanding that any observation can be theory laden (Kuhn, 1996; Overton, 2003) suggests differences in the beliefs about the nature of internal functions can lead to divergence in the evidence that people seek and the ways in which evidence is gathered and interpreted. The assessment framework, which relates to the assessment development process proposed by the U.S. National Research Council as depicted in Figure 1, comprises three interconnected elements: 1) construct, 2) observation, and 3) interpretation. A construct is "a theoretical model of a person's cognition, such as their understanding of a certain set of concepts" (Wilson, 2023, p. 9). In the context of this study, this key element is the theory or set of beliefs about what epistemic cognition is and how it develops or changes over time. Observation represents the set of beliefs about the kinds of tasks or situations that would elicit responses from students that would provide evidence of varying levels of epistemic cognition. Interpretation refers to the methods and tools that are used to interpret the observations, such as a particular quantitative or qualitative analytical technique. As described by the National Research Council

(Pellegrino et al., 2001), these three elements are interconnected and are fundamental to the design and implementation of any assessment. The underlying theory about the nature of epistemic cognition determines the kinds of tasks, events, or settings that are selected or specified for observing the display of epistemic cognition behaviors. It also underpins the reasoning process involved in the consideration of what constitutes evidence of epistemic cognition based on the observations.

Construct Definition

In light of the focus of the Year 9 Inquiry Program, which involved students making decisions and judgments based on information that they have gathered to address issues of sustainability, the assessment tool developed for this project particularly focused on assessing the kinds of connections that students are trying to make when processing and making sense of information. The construct is captured by the notion of epistemic cognition, which is "the thinking that people do about what and how they know" (Sandoval et al., 2016, p. 457). Epistemic cognition is seen as foundational to critical thinking because it concerns the cognitive process regarding how a person evaluates and decides among different claims (Greene & Yu, 2016). It relates to a person's theory of knowledge (epistemological beliefs) and the processes involved in the definition, acquisition, validation, and use of knowledge (Greene et al., 2008; Grix, 2002). The construct has been referred to in different ways including "responses to intellectual and moral relativism" (Perry, 1968/1999), "ways of knowing" (Belenky et al., 1986), "epistemological reflection" (Baxter Magolda, 2004), and "reflective judgement" (King & Kitchener, 2004).

According to Chinn et al. (2011), the earliest psychological research on epistemic cognition was Perry's (1968/1999) study of epistemic development. Based on survey and interview questions of Harvard College students in the U.S. during the 1950s to 1960s and drawing from the work of Piaget, Perry formulated a model of epistemic growth with nine positions ranging from basic duality (right *vs*. wrong, we *vs*. others, and good *vs*. bad) to more complex relativist view of the world. Perry's work has been influential in subsequent research on epistemic cognition development. Felder and Brent (2004) compared four different developmental models of epistemic

Figure 2

A Comparison of Different Models of Intellectual Development

Models of intellectual development	Sophistication/Complexity →							
Baxter Magolda (2000)	Absolute knowing		Transitional knowing		Independent knowing		Contextual knowing	
	Mastery Pattern	Receiving Pattern	Impersonal Pattern	Inter-personal Pattern	Individual Pattern	Inter-individual Pattern		
Perry (1968)	2 Late Dualism		3 Multiplicity Subordinate		4 Multiplicity		5-7 Contextual Relativism Preliminary Commitment	
Belenky (1986)		Received Knowledge		Subjective Knowledge	Procedural Knowledge: Separate Pattern	Procedural Knowledge: Connected Pattern	Constructed Knowledge	
King-Kitchener (2001)	Early Reflective Thinking		Late Pre-reflective Thinking		Quasi-Reflective Thinking		Reflective Thinking	

Note. From "The intellectual development of science and engineering students. Part 1: Models and challenges," by R. M. Felder and R. Brent, 2004, *Journal of Engineering Education*, *93*(4), p. 272. Figure reproduced with permission.

beliefs (called "intellectual development" in the paper) by Perry (1968/1999), Belenky et al. (1986), Baxter Magolda (2004), and King and Kitchener (2004), and found general correspondence between the levels identified by these four models (see Figure 2). As can be seen in the figure, the levels in the four models ranged from less to greater sophistication/complexity, with some levels with one-to-one or one-to-many correspondence between the models.

Summarized by Felder and Brent (2004) and using Baxter Magolda's terminology, epistemic belief development can be generally distinguished by four levels:

- Absolute Knowing: There is certainty in knowledge claims that reflect reality. Positions are either right or wrong. Authorities have the responsibility to communicate The Truth, which is then learnt and repeated by students.
- Transitional Knowing: There is certainty in some but not all knowledge claims. Authorities have the responsibility to communicate the certainties, but students are responsible for making their own judgements about the uncertainties.
- Independent Knowing: There is uncertainty in most knowledge. Rather than relying on authorities or personal feelings, students need to collect and use evidence to support their judgment and take responsibility for their own learning.
- Contextual Knowing: All knowledge claims are contextual and constructed. There are many sources of evidence, some objective and some intuitive, some from students' own thoughts and feelings, and some from recognized experts. Students need to take responsibility for making judgments in the face of uncertainty and ambiguity. They also need to accept changing their decisions in the face of new evidence.

For example, when being asked how someone can tell whether the information given on a website is true or not, a child at the Absolute Knowing level may defer to their trusted authority (e.g., parents or teacher). A child at the Transitional Knowing level may start to acknowledge possible untruthfulness in websites and refer to authorities (e.g., governments or established organizations) for trustworthiness. At the Independent Knowing level, a child may go beyond the official sources of information and start to seek corresponding or counter information to verify the truthfulness of the information on the website. A child at the Contextual Knowing level may indicate the impossibility of fully knowing, stating the ever-changing nature of knowledge and what kind of evidence can be collected to inform decisions. These levels follow other similar Piagetian-style developmental progressions like the Structure of Observed Learning Outcomes (SOLO) taxonomy, which orders a student's understanding in a content area based on increasing structural complexity (Biggs & Tang, 2007; Teaching and Educational Development Institute, n.d.).

King and Kitchener (2004) reported on a study that reviewed 25 studies involving more than 1,500 respondents across different educational levels in the United States high school, first year of college, senior year of college, early graduate study, and advanced doctoral study. Their review suggests that, on average, students enter college at the Prereflective Thinking level (equivalent to Absolute Knowing according to Baxter and Magolda's [2004] level) and leave college at the Quasi-Reflective Thinking level (equivalent to Independent Knowing). Very few students leave college at the highest epistemic cognition level of Reflective Thinking (equivalent to Contextual Knowing). Advanced doctoral students were found to be more consistently reason at this highest level. The review suggests a possible connection between educational level and development of epistemic cognition.

In summary, epistemic cognition is a set of interconnected dispositions, beliefs, and cognitive skills about knowledge (Greene &

Yu, 2016). For the purpose of assessment, the construct is operationalized in this study through examining the underlying evaluative criteria that a person implies when trying to make sense of and act on information. While the construct can be seen as overlapping with comprehension (understanding of information), epistemic cognition goes beyond that to involve reasoning, evaluative judgment, and decision-making.

Construct Map Formulation

Beyond defining the construct, the idea of a construct map describes the form of the construct from a measurement perspective, like a "road map" (Wilson, 2023, p. 9). A construct map is essentially a theorized, simple developmental progression modelling how students might learn and develop in a particular topic area over an extended period of time (Lehrer & Schauble, 2015). The construct map differentiates different levels of performance based on the construct (Wilson, 2023). The differentiation in performance guides the overall design of the assessment tool as well as the scoring of individual items. For the purpose of this study, five levels (called "waypoints" hereafter) of epistemic cognition ordered in increasing sophistication (as denoted by the arrow in Figure 3) were theorized and represented as a construct map in Figure 3. The right column in the figure provides example responses collected from the study for one of the items to illustrate the different waypoints. To provide a clear focus for the assessment tool, the construct map was designed to be unidimensional, representing a single latent variable (Wilson, 2023, p. 8).

The formulation of this construct map drew ideas from existing models of epistemic cognition (Felder & Brent, 2004; Greene & Yu, 2016), the SOLO taxonomy (Biggs & Tang, 2007; Teaching and Educational Development Institute, n.d.), other developmental progressions (Wilson, 2023), and the Australian national (Australian Curriculum Assessment and Reporting Authority, 2023) and Victorian state curriculum standards (Victorian Curriculum and Assessment Authority, 2023). Connecting these different models and standards is a Piagetian constructivist view that conceptualizes development in terms of a

Figure 3

A Proposed Construct Map for Epistemic Cognition

Skill level ("Waypoints")	**Example response for the question "How can you tell whether the information given on a website is true or not?"**
Extended abstract (analytical) Queries underlying assumptions and makes connections beyond the materials given	"I don't think you can ever know whether something is true or not on a website, unless you have witnessed this or if there is clear evidence on the website."
Abstract (relational) Draws links between ideas and integrates information into a coherent whole	"Check the author of the webpage. Also see if (the website) is trying to advertise itself."
Abstract (structural) Identified certain elements relevant to the information	"If there are multiple websites that says the same thing."
Concrete (personal) Relates information to personal interests and experiences	"Ask my teacher/parents"; "I just know."
Pre-concrete (mimicry) Engages with the information/task in a superficial manner	"It's always true/wrong."

progression with increasing sophistication. For example, the SOLO taxonomy (Biggs & Tang, 2007; Teaching and Educational Development Institute, n.d.) has five levels: Prestructural, Unistructural, Multistructural, Relational, and Extended abstract. The levels are described in terms of both quantitative and qualitative changes, where quantitative changes refer to an increase in the amount of detail in a student's response, and where qualitative changes refer to the integration between different details. In this study, the epistemic cognition construct map begins with the lowest waypoint where a person engages with the information/task in a superficial manner and progresses to more complex connections a person might be making when processing information (e.g., making comparisons or examining assumptions). The five waypoints in the construct map, Pre-concrete, Concrete, Abstract (structural), Abstract (relational), and Extended abstract (analytical) waypoints roughly correspond to the five levels in the SOLO taxonomy.

Study Design and Methods

The assessment tool underwent multiple iterations following the Constructing Measures approach of Wilson (2023). The main considerations for the item design included correspondence with and coverage of the progression (construct map), time required for test completion and scoring, suitability of the item context for the prior knowledge of the Year 8 students and their upcoming Year 9 learning inquiry program. For the last two considerations, care was taken to design items to relate to more general knowledge rather than relying on the students' understanding of issues related to the topic of sustainability (which was the topic they were going to study). In light of the limitation of the multiple choice response format (Ku, 2009), a range of response types (multiple choice questions, short answer questions, and extended response questions) were trialled and included in the assessment tool in order to examine the epistemic cognition involved in inquiry activities. Table 1 shows some sample items for each response format.

An initial version of the assessment tool was provided to one of the coordinators of the learning inquiry program. The feedback received was that "the instrument seems to reflect many of the key thinking and research skills needed to participate effectively in the inquiry, especially at Year 9. Students complete a learning journal at Year 9 which has many of these ideas build in to it such as understanding reliable sources and looking at facts and opinions." The feedback suggests good face validity of the assessment tool.

Table 2 shows the item format and the construct map waypoints targeted by each item in the final version of the epistemic cognition assessment tool. The final version has a total of 30 items (note that Question 25 is split into two parts) in three formats: multiple choice questions (16 questions), short answer questions (four questions), and extended response questions (10 questions). Most items require students to evaluate information and provide a response. All of the extended response questions

Table 1

Example Items

Item type	Sample items
Multiple choice question	Q8 "Recycling creates the mindset that it is all right to be wasteful in other areas because we can make up for it through recycling." Which of the following does the statement suggest? A. A problem with recycling. B. A benefit of recycling. C. A solution to recycling. D. A cause of recycling.
Short answer question	Q18 Give three reasons why we should conserve water.
Extended response question	Q23 When someone says "these are the facts", what could the person mean?

require students to justify their responses or explain their reasoning. Example extended questions include "When someone says 'these are the facts,' what could the person mean?" and "Imagine that you are a scientist being consulted by the government about whether kangaroos should be included in the list of protected animals in Australia or not. What recommendations would you give? Explain your reasoning."

Pilot and Data Collection

The initial version of the assessment tool was given to a single class of Year 8 students in a school in Melbourne, Australia. The students completed the assessment tool administered by the class teacher within a single class period (45 minutes). Responses from 17 students were collected, and based on their responses, some items were replaced or modified.

The final version of the assessment tool was given to 164 Year 8 students in the same school. Among the students, 73 were females (44.5%) and 87 were males (53.0%), with four students with missing responses to the gender question. The age of the students ranged between 13 and 16 years, with the majority of students (71.3%) being 14 years old at the time of assessment. Less than half of the students (39.0%) were born in Australia. For those born overseas (62 students), their time in Australia ranged from 4 months to 14 years, with an average of 5.8 years. More than half of the students (64.0%) spoke mostly English at home.

After the initial examination of the student responses, two students (Student I and Student V) who performed at different waypoints (relatively high and medium) on the assessment were chosen for an exit interview to discuss their thoughts when completing the assessment tool. From the student exit interviews, when asked what Student I thought the tool was assessing, she thought it might be "to see the way students think," and it requires students to be "logical and creative at the same time" when completing the responses. Her impression of the assessment tool appears to align with the intention of the assessment tool design.

From the interviews, it was found that the students have completed activities in the past that relate to the evaluation of the trustworthiness of internet sources (Q21) or understanding the steps in inquiry process (Q11 and Q12). This suggests that the students' responses to some of the questions reflect what they have learned in school, consistent with the

Table 2

Target Waypoint(s) and Item Format on the Construct Map (i.e., Score Range)

Item no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25a	25b	26	27	28	29
Target waypoint																														
Extended abstract (analytical)																	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Abstract (relational)																	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Abstract (structural)							X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Concrete (personal)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Pre-concrete (mimicry)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Note. Item format:
Items 1 to 16: Multiple choice questions
Items 17 to 20: Short answer questions
Items 21 to 29: Extended response questions

remark of the program coordinator.

Scoring and Data Processing

An initial scoring scheme, which included potential responses on each waypoint of the construct map, was developed when designing the items. After the data collection, all 14 open-ended questions for approximately 10% of the students were first scored using the initial scoring scheme. The scoring scheme was then given to a colleague to comment on its clarity and coherence. The feedback received mainly pertained to the presentation of the scoring scheme (e.g., bolding key phrases and expanding the level description to include more example responses). After the feedback, which resulted in minor adjustments to the scoring scheme, all of the students' responses were scored based on the revised scoring scheme. The responses were coded as follows:

0: Missing response

1: Insufficient information: Misinterprets the task or expresses inability to complete the task (e.g., due to lack of knowledge or incomplete information given)

2: Pre-concrete (mimicry)

3: Concrete (personal)

4: Abstract (structural)

5: Abstract (relational)

6: Extended abstract (analytical)

The iterative process for developing the scoring scheme is important for improving the clarity of the scheme and capturing the content of the student responses. Score level 1 (Insufficient Information) was not initially part of the construct map. However, some of the student responses (e.g., "I don't know") appear to be qualitatively distinct from the Pre-concrete (e.g., repeating the question) or Concrete (e.g., relating to personal experience) waypoint responses. Such responses were therefore coded as a separate score level to see if such responses might form a distinct category.

After all the responses were scored, the scores were recorded on a scoring sheet and then entered into Excel. Data accuracy (e.g., plausible maximum and minimum values) was checked using simple Excel formulas.

Inter-Rater Agreement

Two raters scored 60% and 40% of the 164 students' responses, respectively. After initial coding training for the raters to become familiar with the scoring scheme, the short answers and extended responses of 16 randomly selected cases (approximately 10% of cases) were double-scored to check for inter-rater agreement between the two raters. Since the scoring scheme to the individual questions of the assessment tool forms an ordinal scale, inter-rater agreement was evaluated using non-parametric Spearman's ρ (Spearman, 1904) and calculated using SPSS (IBM Corporation, 2019). The Rank Order Correlation between the two raters was .86 (n = 224 response count), indicating a strong level of rater agreement.

Data Analysis

The partial credit item response model (PCM; Masters & Wright, 1997) was estimated by the computer software R (R Core Team, 2014) with supporting packages (TAM, WrightMap, and SIRT; Irribarra & Freund, 2014; Kiefer et al., 2015; Robitzsch, 2015) using Expected A Priori (EAP) estimates. The PCM allows items that have more than two item steps (i.e., score levels) to be analyzed, and assumes the difficulty of the item steps within an item to be of different intervals (Wu et al., 2007). Mathematically, with θ representing person ability, δ representing item difficulty, for item i and the item category number k, if item i is a partial credit item with item steps 0, 1, 2, …, m_i (where the subscript i allows for the items to have different numbers of categories), the probability of person j scoring x on item i can be expressed as:

$$\mathrm{P}_{ijx} = \frac{\exp \sum_{k=0}^{x}\left(\theta_j - \delta_{ik}\right)}{\sum_{h=0}^{m_i} \exp \sum_{k=0}^{h}\left(\theta_j - \delta_{ik}\right)}, \quad x = 0, 1, \ldots, m_i, \qquad (1)$$

where

$$\sum_{k=0}^{0}\left(\theta_n - \delta_{ik}\right) = 0 \text{ and}$$

$$\sum_{k=0}^{h}\left(\theta_n - \delta_{ik}\right) \equiv \sum_{k=1}^{h}\left(\theta_j - \delta_{ik}\right).$$

Since the PCM allows the item step parameters ($\delta_{i1}, \delta_{i2}, \delta_{i3}, \ldots, \delta_{im_i}$) for each item i to be estimated, substituting these estimates into Equation 1 provides the estimated probabilities of scoring 0, 1, … m_i on item i for any specified ability θ (Masters & Wright, 1997).

Results

The PCM converged successfully. Multiple classical test theory and item response theory statistics were inspected to examine the psychometric properties of the epistemic cognition assessment tool, including reliability, model-data fit, and item difficulty estimation in relation to the construct map.

Model–Data Fit

Item fit statistics were inspected to examine the model–data fit. According to item response theory, items that fit a unidimensional model are generally assumed to measure the same construct or latent trait (Wu & Adams, 2007). Items that do not fit the model well may be measuring a different construct, although the misfitting could also be due to other reasons, such as poor item design or systematic errors associated with items or population sampling (Embretson & Reise, 2000; Wilson, 2023).

Model–data fit can be assessed using statistical means by examining the *mean square fit statistics* for individual items. Wright and Masters (1982) proposed the weighted fit mean square (infit) statistic for examining item fit:

$$\text{Weighted mean square} = \frac{\sum_n z_{ni}^2 Var\left(x_{ni}\right)}{\sum_n Var\left(x_{ni}\right)} = \frac{\sum_n \left[x_{ni} - E\left(x_{ni}\right)\right]^2}{\sum_n Var\left(x_{ni}\right)} \qquad (2)$$

where x_{ni} is the observed score for person n on item i, and z_{ni} the standardized residual.

Table 3 presents the weighted fit statistics for the 30 epistemic cognition items based on the unidimensional PCM. The values in the table are sorted from low to high according to the weighted t-statistics.

Table 3

Weighted Mean Square Fit Statistics

Item	Difficulty estimate	Standard error	Weighted mean square	Weighted t
Q26	0.02	0.13	0.83	−1.10
Q08	−1.21	0.13	0.87	−1.07
Q05	−2.09	0.24	0.89	−0.63
Q24	−0.67	0.13	0.92	−0.63
Q13	−1.63	0.16	0.93	−0.36
Q10	−1.55	0.15	0.94	−0.24
Q06	−4.13	0.58	0.95	0.09
Q28	−1.56	0.16	0.96	−0.27
Q23	−1.55	0.12	0.96	−0.19
Q04	−3.61	0.46	0.96	0.05
Q12	−1.63	0.15	0.97	−0.15
Q09	−1.77	0.15	0.98	−0.20
Q15	−2.00	0.24	0.98	−0.06
Q03	−4.55	0.71	0.99	0.21
Q17	1.09	0.13	0.99	−0.10
Q21	0.39	0.12	0.99	−0.04
Q07	−0.90	0.11	0.99	−0.05
Q01	−3.84	0.51	0.99	0.14
Q11	−0.78	0.11	1.01	0.11
Q02	−4.55	0.71	1.01	0.25
Q14	−0.69	0.17	1.02	0.35
Q25b	0.16	0.16	1.03	0.24
Q25a	−0.95	0.13	1.04	0.30
Q29	−0.24	0.11	1.04	0.31
Q18	1.13	0.13	1.06	0.40
Q19	−0.48	0.11	1.06	0.44
Q27	0.10	0.10	1.09	0.52
Q20	−0.34	0.11	1.09	0.85
Q22	−0.49	0.09	1.19	1.40
Q16	−0.79	0.11	1.20	1.77

The weighted t-statistics were first examined to identify mean square values that fell outside of the confidence interval of 95% (two-tailed test), and items with weighted mean square values that exceeded the recommended range indicate item misfitting (weighted mean square 0.5 to 1.5 and *t*-statistics ±1.96; Wright & Linacre, 1994; Wu et al., 2007). As can be seen from the table, none of the 30 items has a weighted mean square outside the 0.5 to 1.5 range, and none of the *t* values are less than −1.96 or greater than 1.96. An examination of the item step statistics also found that none of the item steps exceeded the recommended thresholds (weighted mean square ranged from 0.88 to 1.09 and *t*-statistics ranged from −1.37 to 1.34). The fit statistics suggest that, on the whole, the epistemic cognition items were measuring the same construct and fit the unidimensional PCM.

Response Count

The distribution of responses according to item step (i.e., Pre-Concrete, Concrete, Structural, etc.) was examined (see Figure 4). An examination of the total response count for each item step found that out of 4,711 scored responses, only 48 (1.0%) were difficult to interpret and did not fit into any of the other five response categories. Out of all the scored responses, most responses were at the Concrete (45.3%) or Structural (44.5%) waypoints. There were a low number of responses at the Relational (2.8%) or the Analytical (0.4%) waypoint. The item step and person ability estimates, therefore, need to be interpreted in light of the response count as a larger count provides more confidence in the accuracy of the item step difficulty and person ability estimates at those waypoints.

Reliability Indices

Classical Test Theory Items Results

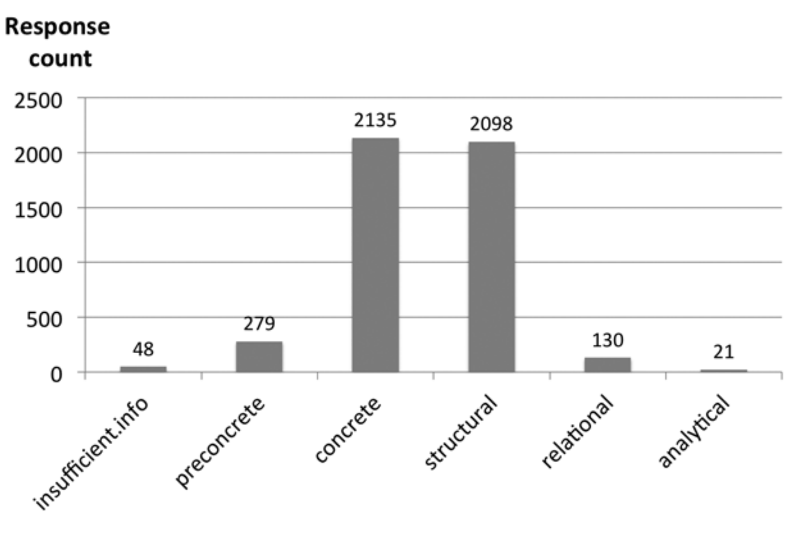
The internal consistency of the epistemic cognition assessment tool was examined based on classical test theory statistics using the alpha coefficient, which provides an indication of the agreement of scores between the items in a test (Allen & Yen, 1979). The alpha coefficient was .81, indicating good reliability.

Rasch Modelling Results

Person-separation reliability indicates how well the test items in a particular test separate the ability levels of students in the sample based on PCM estimates (Wright & Masters, 1982). The index ranges from 0 to 1, where a higher value indicates better separation and more precision in the measurement (Wright & Stone, 1999). The person-separation reliability of the epistemic cognition assessment tool was .74, indicating an acceptable level of separation between the ability levels of people in the sample.

Figure 4

Response Count Across Thurstonian Thresholds for the Full Assessment Tool (30 Items)



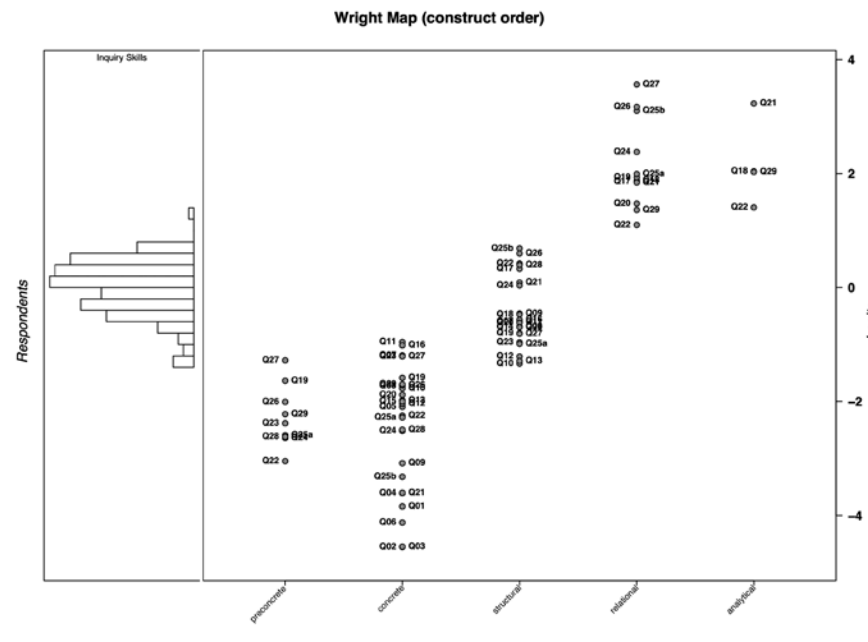
Item Difficulty

Figure 5 shows the Wright map from the PCM analysis, which is a graphical representation of the item threshold estimates based on the Thurstonian thresholds. Thurstonian thresholds for a score category is defined as "the ability at which the probability of achieving *that score or higher* reaches 0.50" (Wu et al., 2016, p. 170). In the Wright map, the respondents' ability estimates are charted on the left side of the figure with *y*-axis as the logits scale and the *x*-axis the frequency count. On the right side of the respondent chart, the difficulty of each Thurstonian threshold is mapped onto the same logit scale for the respondents grouped according to the construct map waypoints (Pre-concrete, Concrete, Structural, etc.). The location of the respondents and the thresholds along the scale can be interpreted in the same way, where the higher up the respondent is on the scale, the more able the person is, and the higher up the threshold is on the scale, the more difficult the threshold is. Thresholds that are at the same level horizontally are of similar difficulty.

The Wright map provides an empirical and graphical way to examine whether the respondents' responses align with the ordered response waypoints theorized in the construct map (Figure 3). As can be seen on the Wright map, the difficulty of the five item steps at the Pre-concrete waypoint overlaps (along the *y*-axis) with those at the Concrete waypoint, and there is some overlap in difficulty between the harder Concrete responses and the easier Structural responses. This is not surprising, as even though the items may not be perfect, the three categories do increase in difficulty and do not overlap a great deal, suggesting internal structure validity is adequate. The difficulty of the thresholds at the Analytical waypoint overlaps completely with those at the Relational waypoint. This is likely due to the number of students at this waypoint, where

Figure 5

Wright Map With Items Ordered According to Thurstonian Thresholds



not many students provided responses at the Analytical waypoint. Further work would involve collecting a sample of students at the Analytical waypoint in order to better map the item thresholds of that waypoint. Based on the Wright map, the epistemic cognition assessment tool is best able to distinguish responses between the Concrete, Structural, and Relational waypoints. Pre-concrete waypoint responses are not distinguishable from Concrete waypoint responses. This may be because there were relatively few responses at the Pre-concrete waypoint for this sample.

Nonetheless, from the Wright map, the thresholds more than cover the range of respondent ability in the sample, and therefore appear to be appropriate for the ability levels of the student participants. In light of the response count where the students' responses concentrate at the Concrete and Structural waypoints (see Figure 4), collecting responses at the lower (Pre-concrete) and higher (Relational and Analytical) ends of the scale may help to provide a better estimation of the item step difficulty and person ability at the extreme ends of the scale.

Discussion

The epistemic cognition assessment tool was designed to examine a theorized developmental progression to assist a school with determining what developmental levels their Year 8 students possess before they participate in the inquiry program in Year 9, so the school can better tailor the program to the learning needs of the students. The assessment tool's face validity was supported by the feedback from the program coordinator and the exit interview with students. An examination of the student responses found most of the students assessed were at the Concrete (45.3%) and Structural (44.5%) waypoints, with very few Relational (2.8%) and Analytical (0.4%) waypoint responses. This is consistent with the study reported by King and Kitchener (2004) that high school students consistently displayed Prereflective Thinking level (equivalent to Concrete and Structural waypoints). From a developmental perspective, this suggests most students in the current study were at the beginning stage of epistemic cognition and may require support to promote their thinking to the Relational and Analytical waypoints in order to grapple with issues related to sustainability. For example, an examination of the students' qualitative responses to questions regarding internet sources (Q20 and Q29) found that most students appear to be critical of information from internet sources (e.g., Student B: "Wikipedia is not a reliable source because anyone can edit therefore anyone can spread lies and fake information"; Student H: "[Wikipedia] is not a reliable source because unknown people can easily manipulate the information and feed you false facts"). However, many of the students tend to apply very basic rules (e.g., check if the URL has the suffix of .gov or .org; or the recency of the article) to determine the trustworthiness or the accuracy of the information (Concrete waypoint), or suggest making comparison between multiple online sources (Structural waypoint), rather than evaluating the content itself. Having the opportunity to discuss or think about the author's intention and sources of bias may help the students engage with information at a deeper level.

Although the school's inquiry learning program needs can be met with an examination of the raw distribution of the student responses at the different waypoints according to the theorized construct map, psychometric methods can be used to empirically evaluate the validity of the epistemic cognition assessment tool and identify aspects of the tool that may require improvements. Psychometrically, the assessment tool appears to show good properties with reasonably high internal consistency and person-separation reliability, and good model-data fit. An examination of the thresholds in relation to person ability estimates suggests the assessment tool is at an appropriate difficulty level for the Year 8 students. There was a clear difference in thresholds between Structural and Relational waypoints, and between the Concrete and Structural waypoints, even though there was some overlap between the latter two waypoints. Although there were overlaps between the Pre-concrete and Analytical waypoints and other waypoints according to item difficulty, in combination, these responses account for only 3.2% of the total response count. Further development work on the assessment tool will involve targeting the extreme ends of the scale (e.g., younger children or adults). The current study only assessed the students before they started the learning inquiry program. Assessing the students before and after the program would help to determine the effectiveness of the program in promoting growth in epistemic cognition among the students.

Considering this as an example of measurement practice, the process of developing the construct map, creating items and scoring scheme that corresponds to the construct map, and applying item response modelling to examine the psychometric properties of the items, offers insights into the construct of interest and provides a structure for the teaching staff to reflect on the performance that they seek from students as a result of the inquiry learning program. Discussion with the program staff during the assessment tool development process found that the staff appeared to have mainly focused on providing students with the opportunity to inquire (i.e., carrying out investigative activities) and not necessarily trying to guide students towards more sophisticated levels of thinking skills. When the assessment results were presented to the program staff after the completion of the study, the staff reported finding the sample student responses ordered according to the construct map particularly useful for their teaching to distinguish different student performance levels. The responses from students that were coded at the higher levels of epistemic cognition (e.g., Abstract [relational] and Extended abstract [analytical]) provided the teachers with explicit examples of where they may want students to progress. This could help the teachers to devise learning objectives to guide their teaching and student learning (Biggs & Tang, 2007). The assessment development process, therefore, both offers an opportunity to operationalize epistemic cognition development and provides information about the students' epistemic cognition to inform teaching. Subsequent conversations with the school staff and other schools in Melbourne indicated that teachers generally find it difficult to conceptualize and teach critical thinking, highlighting the value of the assessment tool.

As suggested by Black et al. (2011), an assessment tool like the one designed for this study can offer a useful "roadmap" to inform teaching and student learning. At the broader level, the assessment tool distinguishes strategic aims for teachers to prioritize their teaching or the design of the students' inquiry learning activities in terms of focusing on the subject content (e.g., production and consumption, energy, transport, people and population, and biodiversity) and/or modelling reasoning skills in relation to a specific topic. The construct map and scoring scheme offer teachers a framework for assessing the sophistication of a student's epistemic cognition based on their responses to questions or discussion topics. The assessment items offer teachers example questions and discussion activities to elicit student responses for formative assessment. The assessment tool can also serve summative purposes to examine the progress or change in student thinking at the end of a learning sequence for individual students or for the whole class.

In terms of usage, the current assessment items have incorporated topics and contexts that are specific to the Year 9 inquiry learning program on sustainable living. Usage on other topics (e.g., ethics, political, or social issues) or with other year levels would require adapting the questions and language to suit the content knowledge and age of the students. With the rise of genAI, some of the items may also be adapted to examine how students make decisions regarding texts generated by large language models like ChatGPT. While any adaptation of an assessment tool will require re-examination of the psychometric properties of the modified tool, the current tool, including the construct map, assessment items, and scoring scheme, provides a starting point and process for the work. Such adaptation work may follow the process depicted in Figure 1, beginning with reviewing the construct map and learning progression, then revising the assessment items and scoring scheme, piloting the assessment items with the targeted student group, and analyzing the item response and revising the assessment tool before broader administration.

Conclusion

This paper contributes to Professor Wang's work in promoting the relevance of measurement in educational settings and specifically in relation to the teaching of critical thinking in schools. The development of the epistemic cognition assessment tool was found to be particularly useful for assisting the participating school with identifying the learning needs of their students for undertaking the learning inquiry program. The differentiation of levels of student responses (Pre-concrete, Concrete, Abstract, etc.) in terms of sophistication in their epistemic cognition was helpful to support teachers to target their teaching to meet the learning needs of the students. Considering the heavy emphasis on critical thinking and inquiry learning internationally (Chu et al., 2017), this paper highlights the importance of using the assessment development process reported in this paper to shape the pedagogical approach through defining the anticipated student performance levels to guide teaching.

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Brooks/Cole Publishing Company.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Australian Curriculum Assessment and Reporting Authority. (2023). *Critical and creative thinking* (Version 8.4). https://www.australiancurriculum.edu.au/f-10-curriculum/general-capabilities/critical-and-creative-thinking/

Baxter Magolda, M. B. (2004). Evolution of a constructivist conceptualization of epistemological reflection. *Educational Psychologist*, *39*(1), 31–42.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). *Women's ways of knowing: The development of self, voice, and mind*. Basic Books.

Biggs, J. B., & Tang, C. (2007). *Teaching for quality learning at university* (3rd ed.). Society for Research into Higher Education; Open University Press.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st Century skills* (pp. 17–66). Springer. https://doi.org/10.1007/978-94-007-2324-5_2

Black, P., Wilson, M., & Yao, S.-Y. (2011). Road maps for learning: A guide to the navigation of learning progressions. *Measurement: Interdisciplinary Research and Perspectives*, *9*(2–3), 71–123. https://doi.org/10.1080/15366367.2011.591654

Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist*, *46*(3), 141–167.

Chu, S., Reynolds, R. B., Tavares, N. J., Notari, M., & Lee, C. W. Y. (2017). *21st century skills development through inquiry-based learning: From theory to practice*. Springer. https://doi.org/10.1007/978-981-10-2481-8

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9–26). W H Freeman; Times Books; Henry Holt & Co.

Ennis, R. H. (2018). Critical thinking across the curriculum: A vision. *Topoi: An International Review of Philosophy*, *37*(1), 165–184. https://doi.org/10.1007/s11245-016-9401-4

Felder, R. M., & Brent, R. (2004). The intellectual development of science and engineering students. Part 1: Models and challenges. *Journal of Engineering Education*, *93*(4), 269–277.

Greene, J. A., Azevedo, R., & Torney-Purta, J. (2008). Modeling epistemic and ontological cognition: Philosophical perspectives and methodological directions. *Educational Psychologist*, *43*(3), 142–160. https://doi.org/10.1080/00461520802178458

Greene, J. A., Sandoval, W. A., & Bråten, I. (2016). *Handbook of epistemic cognition*. Routledge.

Greene, J. A., & Yu, S. B. (2016). Educating critical thinkers: The role of epistemic cognition. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 45–53. https://doi.org/10.1177/2372732215622223

Griffin, P. E., McGaw, B., & Care, E. (Eds.). (2012). *Assessment and teaching of 21st century skills*. Springer. https://doi.org/10.1007/978-94-007-2324-5

Grix, J. (2002). Introducing students to the generic terminology of social research. *Politics*, *22*(3), 175–186.

Hallinger, P., Dongyu, L., & Wang, W.-C. (2016). Gender differences in instructional leadership: A meta-analytic review of studies using the principal instructional management rating scale. *Educational Administration Quarterly*, *52*(4), 567–601. https://doi.org/10.1177/0013161x16638430

Hsueh, I.-P., Wang, W.-C., Sheu, C.-F., & Hsieh, C.-L. (2004). Rasch analysis of combining two indices to assess comprehensive ADL function in stroke patients. *Stroke*, *35*(3), 721–726. https://doi.org/10.1161/01.STR.0000117569.34232.76

IBM Corporation. (2019). *IBM SPSS Statistics for Windows* (Version 26.0) [Computer software].

Irribarra, D. T., & Freund, R. (2014). *Wright map: IRT item-person map with ConQuest integration R package* (Version 1.1). University of California. http://github.com/david-ti/wrightmap

Kendeou, P., Robinson, D. H., & McCrudden, M. T. (Eds.). (2019). *Misinformation and fake news in education*. Information Age Publishing.

Kiefer, T., Robitzsch, A., & Wu, M. (2015). *TAM: Test analysis modules R package* (Version 1.6-0). Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens. http://CRAN.R-project.org/package=TAM

King, P. M., & Kitchener, K. S. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist*, *39*(1), 5–18.

Ku, K. Y. L. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity*, *4*(1), 70–76. https://doi.org/10.1016/j.tsc.2009.02.001

Kuhn, D., & Dean, D., Jr. (2004). Metacognition: A bridge between cognitive psychology and educational practice. *Theory into Practice*, *43*(4), 268–273. https://doi.org/10.1207/s15430421tip4304_4

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). University of Chicago Press.

LaGarde, J., & Hudgins, D. (2018). *Fact vs. fiction: Teaching critical thinking skills in the age of fake news*. International Society for Technology in Education.

Lehrer, R., Kim, M.-J., Ayers, E., & Wilson,

M. (2014). Toward establishing a learning progression to support the development of statistical reasoning. In A. P. Maloney, J. Confrey, & K. H. Nguyen (Eds.), *Learning over time: Learning trajectories in mathematics education* (pp. 31–59). Information Age Publishing.

Lehrer, R., & Schauble, L. (2015). Learning progressions: The whole world is NOT a stage. *Science Education*, *99*(3), 432–437. https://doi.org/10.1002/sce.21168

Lin, C.-Y., Hwang, J.-S., Wang, W.-C., Lai, W.-W., Su, W.-C., Wu, T.-Y., Yao, G., & Wang, J.-D. (2019). Psychometric evaluation of the WHOQOL-BREF, Taiwan version, across five kinds of Taiwanese cancer survivors: Rasch analysis and confirmatory factor analysis. *Journal of the Formosan Medical Association*, *118*(1 Pt 2), 215–222. https://doi.org/10.1016/j.jfma.2018.03.018

Masters, G. N., & Wright, B. D. (1997). The partial credit model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 101–121). Springer.

Mejia, M., & Sargent, J. M. (2023). Leveraging technology to develop students' critical thinking skills. *Journal of Educational Technology Systems*, *51*(4), 393–418. https://doi.org/10.1177/00472395231166613

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. National Academy Press. http://www.nap.edu/books/0309072727/html/

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press. https://www.nap.edu/catalog/13165/a-framework-for-k-12-science-education-practices-crosscutting-concepts

Overton, W. F. (2003). Development across the life span. In R. M. Lerner, M. A. Easterbrooks, & J. Mistry (Eds.), *Handbook of psychology* (Vol. 6, pp. 13–42). John Wiley and Sons.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. National Academy Press.

Perry, W. G. (1999). *Forms of intellectual and ethical development in the college years: A scheme*. Jossey-Bass. (Original work published 1968)

R Core Team. (2014). *R: A language and environment for statistical computing* (Version 3.1.0 GUI 1.64). R Foundation for Statistical Computing. http://www.R-project.org

Robitzsch, A. (2015). *sirt: Supplementary item response theory models R package* (Version 1.5-0). Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens. http://CRAN.R-project.org/package=sirt

Sandoval, W. A., Greene, J. A., & Bråten, I. (2016). Understanding and promoting thinking about knowledge: Origins, issues, and future directions of research on epistemic cognition. *Review of Research in Education*, *40*(1), 457–496. https://doi.org/10.3102/0091732x16669319

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159

Teaching and Educational Development Institute. (n.d.). *Biggs' structure of the observed learning outcome (SOLO) taxonomy*. University of Queensland.

Victorian Curriculum and Assessment Authority. (2023). *Critical and creative thinking*. https://victoriancurriculum.vcaa.vic.edu.au

Vincent-Lancrin, S., González-Sancho, C., Bouckaert, M., de Luca, F., Fernández-Barrerra, M., Jacotin, G., Urgel, J., & Vidal, Q. (2019). *Fostering students' creativity and critical thinking: What it means in schools*. Organisation for Economic Co-operation and Development. https://doi.org/10.1787/62212c37-en

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., & Wilson, M. (2005). The Rasch testlet model. Applied Psychological Measurement, *29*(2), 126–149. https://doi.org/10.1177/0146621604271053

Wang, W.-C., & Yeh, Y.-L. (2003). Effects of anchor item methods on differential item functioning detection with the likelihood ratio test. *Applied Psychological Measurement*, *27*(6), 479–498. https://doi.org/10.1177/0146621603259902

Wilson, M. (2023). *Constructing measures: An item response modeling approach* (2nd ed.). Routledge.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(3), 370. http://www.rasch.org/rmt/rmt83b.htm

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Mesa Press.

Wright, B. D., & Stone, M. (1999). *Measurement essentials* (2nd ed.). Wide Range. https://www.rasch.org/measess/me-all.pdf

Wu, M. L., & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach*. Educational Measurement Solutions. https://www.edmeasurement.com.au/pdf/RaschMeasurement_Complete.pdf

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

Wu, M. L., Tam, H. P., & Jen, T.-H. (2016). *Educational measurement for applied researchers: Theory into practice*. Springer.