Development and Validation of a Reading in Science Holistic Assessment (RISHA): a Rasch Measurement Study

Abstracts

Researchers in science education lacks valid and reliable instruments to assess students' *disciplinary* and *epistemic* reading of scientific texts. The main purpose of this study was to develop and validate a Reading in Science Holistic Assessment (RISHA) to assess students' holistic reading of scientific texts. RISHA measures students' content, procedural and epistemic domain of reading two texts, one history-of-science text and another socio-scientific text. The initial 24-item RISHA was administered to 161 Grade 9 students from 3 schools. The multidimensional Rasch partial credit model was used to analyse the reliability and validity of RISHA. All items demonstrated good fit and reliability. According to logit scores generated for each domain in Rasch analysis, students in our study performed better in content domain and less well in the epistemic domain. Students also performed significantly better in the epistemic domain of the socio-scientific text than in the history-of-science text. RISHA provides accurate measures in various domains of reading scientific texts and various contexts of scientific texts. We propose that RISHA could potentially be applied to studying the effect of reading-science intervention or predictors of students' performance in each domain of reading scientific texts.

Keywords History-of-science texts; Rasch analysis; Reading scientific texts; Science reading research; Socio-scientific texts

Introduction

Language is a fundamental component of scientific literacy (Fang, 2008; Norris & Phillips, 2003) According to the *Framework for K-12 Science Education* (NRC, 2012), reading scientific texts is a fundamental practice of science, as science is a way of knowing that is represented by text. Compared to generic texts, scientific texts comprise specific linguistic features such as technical vocabulary and passive voice (Fang, 2006). The contexts of scientific texts can be grounded in historical discovery of scientific knowledge (Chen, Chen, & Liu, 2022) and socio-scientific issues (Fazio, Gallagher, & DeKlerk, 2022; Stang Lund, Bråten, Brandmo, Brante, & Strømsø, 2019). History of science (HOS) in reading materials provides a room for addressing development of scientific knowledge (Mccomas, 2011; Monk & Osborne, 1997), while socio-scientific issues (SSI) in reading materials expose students to various values, assumptions and concepts underlying epistemic aspects of science (Matkins & Bell, 2007; Sadler, Chambers, & Zeidler, 2004).

Scientific texts inherit ontological and epistemic attributes, reasoning and arguments (Yore, Pimm, & Tuan, 2007). For reading scientific texts in K-12 curriculum, educators should reorient their view of reading from merely decoding linguistic features to a more holistic account of the disciplinary features of scientific texts (Tang, Lin, & Kaur, 2022; Yore & Tang, 2022). In this post-truth era, Tang et al. (2022) argued that not only should students equip the skills of identifying canonical content knowledge from scientific texts, they also need to be aware of epistemological source, evidence and accuracy of information presented in scientific texts. In the present study, we built a framework encompassing a set of competence in reading scientific texts, which comprises content, procedural and epistemic domains. *Content domain* includes detecting key scientific knowledge, reasoning scientific knowledge, generating inferences from scientific knowledge and understanding scientific vocabularies related to scientific knowledge in texts (Wang, Chen, Fang, & Chou, 2012); *procedural domain* involves identifying manipulative skills, understanding aims and purposes of scientific investigations, understanding evidence and metalanguages in science (Millar, Lubben, Got, & Duggan, 1994; Cheung and Sonkqayi, 2023); *epistemic domain* involves acquiring a sophisticated epistemic beliefs in sources, certainty, development and justification of scientific knowledge presented in scientific texts (Conley, Pintrich, Vekiri, & Harrison, 2004; Tsai, Jessie Ho, Liang, & Lin, 2011). Previous research studies only focused on content domain, while our newly developed framework addresses reading in science as a holistic fashion which included all three domains.

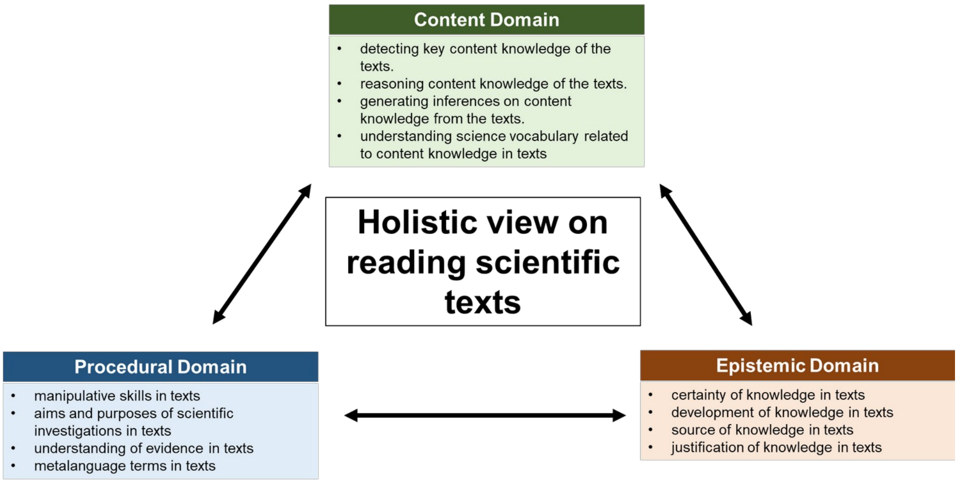
In science education literature, there is a lack of assessment that evaluates students' holistic canonical, procedural, and epistemic understanding of scientific texts, specifically in the contexts of HOS and SSI. As argued by Klopfer and Aikenhead (2022), both HOS and SSI contexts exhibited interplay between nature of science and scientists, as well as the interaction between science and society, affecting students' personal lives. Both types of text are conducive for eliciting students' epistemic understanding of science. HOS texts, which are narrative in nature, present processes of how scientific ideas are validated, modified and verified over temporal events (Li, Yu and Li, 2023); SSI, which is argumentative in nature, presents socio-institutional dimensions of science through evidence-based argumentation (Avsar Erumit and Yuksel, 2023). In other words, HOS texts can offer a participatory drama for how science progresses from the past to the present, while SSI texts engage students in collective decision-making on debatable issues (Klopfer and Aikenhead, 2022). Most of these research instruments measured students' topic-specific understanding of the reading texts and integration of scientific information (e.g., Bernholt, Härtig, & Retelsdorf, 2022; Härtig, Bernholt, Fraser, Cromley, & Retelsdorf, 2022; Ozuru, Dempsey, & McNamara, 2009; Wang & Chen, 2016), instead of providing optimal context for eliciting students' disciplinary and epistemic reading of texts related to their personal lives. In this regard, we reported on the procedures of developing and validating a Reading in Science Holistic Assessment (RISHA), with reference to latest research development in reading in science. RISHA encompasses students' understanding of content, procedural and epistemic domains of HOS and SSI text. Multidimensional Rasch partial-credit model (Masters, 1982) was used to assess the validity and reliability of RISHA items.

Theoretical Framework

In a recent special issue of *International Journal of Science and Mathematics Education (IJSME)*, Yore and Tang (2022) called for the need to move reading research in science education from simple view of reading to complete view of reading. The *simple* view of reading focuses on students' decoding linguistic devices from scientific texts, while the *complete* view of reading considers nature of discipline and promotes learning of science which focused on disciplinary literacy outcomes (Tang et al., 2022; Yore & Tang, 2022). Scientific texts are unique because they explain events and patterns, as well as presenting scientific language (Shymansky, Yore, & Good, 1991; Yore et al., 2004). Even good readers also found scientific texts confusing because of the complexity of scientific texts (Nigro & Trivelato, 2012). The term "complete" view of reading from Yore and Tang (2022) is modified to "holistic" reading of scientific texts in foregrounding the design of our reading instrument in science. This change of term better reflects the importance of drawing on various domains related to scientific literacy in reading scientific texts. Reading scientific texts is an integral part of scientific literacy which encompasses content knowledge, procedural knowledge and epistemic knowledge as stipulated in Programme for International Student Assessment (PISA) 2015/2018 framework (OECD, 2019): content knowledge refers to an understanding and explanation of scientific phenomena; procedural knowledge refers to characterization of procedures, methods and evidence to establish scientific knowledge; epistemic knowledge refers to understanding of rationales of scientific practices and the status of claims. Nevertheless, reading scientific texts does not only concern recalling knowledge but a set of skills and depositions involving interaction between prior knowledge and science texts, such as identifying evidence from texts for supporting arguments (Lammers, Goedhart, & Avraamidou, 2019; Symons, 2017). Thus, holistic reading of scientific texts comprises various *domains of competences and understanding in reading scientific texts* instead of merely *knowledge*, which are content, procedural and epistemic domains (***Figure 1***). These domains foreground the conceptual design of the instrument reported in this study.

Figure 1

Conceptual framework developing RISHA



Content domain in reading in scientific texts describes how students understand and reason information in scientific texts. The process of reading scientific texts is seen as a form of scientific inquiry (Norris & Phillips, 2008). When students read scientific texts, students detect scientific content and link their meaning to relevant prior scientific knowledge (Quellmalz & Hoskyn, 1996). Through the process, reasoning takes place because students need to analyze relationships between scientific elements, identifying similarities and differences between scientific phenomena (Shepardson & Gummer, 2001). Moreover, in detecting and reasoning scientific texts, there is a number of scientific vocabularies which learners have to understand their meaning (Perfetti, Landi, & Oakhill, 2005; Wellington & Osborne, 2001). The four components of content domain from Wang et al. (2012) were drawn on: (a) detecting the key idea from scientific texts, which includes identifying languages in texts, such as titles, subheadings, and topic sentences; (b) reasoning scientific knowledge from texts, which comprises analysing and comparing two scientific theories, objects, and processes; (c) inferring from scientific texts, which involves extracting information from the texts to clarify meaning of scientific theories, ideas, and processes as well as formulating explanations; (d) reasoning individual scientific vocabulary, which consists of defining the key labels of scientific theories, ideas, and processes in the texts.

Procedural domain in reading scientific texts describes how students understand diverse methods, evidence and procedures that derive knowledge in scientific texts. The four components of procedural domain from Millar et al. (1994) were drawn on: (a) manipulative skills in the texts, which involves identifying instruments and procedures of investigations from the texts; (b) aims and purposes of investigations, which comprises inferring the nature and purpose of investigation in relation to the main idea of the text; (c) understanding of evidence, which consists of explaining the criteria for assessing the quality of evidence. We also added one more component, (d) understanding of metalanguage terms in scientific investigation, which refers to recognizing meaning of terms in relation to scientific genres. In the process of scientific investigation, it is inevitable to mention terms for talking about scientific language, such as laws, models and hypotheses (Abrahams, Reiss, & Sharpe, 2013; K.-S. Tang, 2021; K.-S. Tang & Rappa, 2021). These metalanguages in scientific texts inform actions in day-to-day scientific investigation in general (Millar et al., 1994), hence being considered as a part of procedural domain.

Epistemic domain in reading scientific texts refers to a set of epistemic understanding and skills in justifying the status of claims and rationales of scientific practices described in texts. In this domain, students become aware that the idea is the one among multiple alternative, as well as evaluating this scientific idea based on evidence from the texts and using it to explain natural phenomena (Ford & Wargo, 2012). In other words, epistemic understanding constitutes knowledge of features of scientific processes that derive a claim (Lederman, 2013;

Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002). Empirical studies have argued for a significant relationship between epistemic understanding and reading texts (Strømsø, Bråten, & Samuelstuen, 2008; Yang, Chang, Chen, & Chen, 2016). For instance, students who held simplicity of knowledge tended to paraphrase short, factual information from the texts, while students who understood the complexity of knowledge were able to provide within-text elaborations (Strømsø & Bråten, 2009). Nonetheless, measurement of these epistemic understandings remained separate from the context of scientific texts in these studies (Strømsø & Bråten, 2009). In the study by Yang et al. (2016), Scientific Epistemological Questionnaire (SEQ) (Conley et al., 2004) was used to examine the relationship between students' epistemic beliefs and content domain of reading scientific texts. SEQ is a measure of students' epistemic belief in science and not targeting a particular scientific context. Using SEQ, Chen et al. (2022) examined if using history of science as homework could improve students' epistemic understanding. Nonetheless, epistemic understandings were specific to certain theories or topics within a domain (Strømsø et al., 2008). Leung (2020) suggested the concept of adherence to epistemic understanding, which refers to the extent of which students activate epistemic resources to evaluate scientific claims in particular contexts (e.g., new articles). In the context of reading comprehension in science, students need to develop a sophisticated epistemic understanding of the scientific claims presented by the texts and justify their sophisticated epistemic understanding using evidence from the texts. The epistemic domain of reading scientific texts is more than a set of understanding. This domain involves students' competence in explicating their reasons why they understand the status of scientific claim in the text in a particular way, which is based on information from the text.

We theorized the four components in epistemic domain in reading scientific texts, which includes: (a) (un)certainty of knowledge claims in the text, which involves recognizing the uncertainty of scientific theories, ideas, and processes in the texts; (b) development of knowledge claims in texts, which encompasses students’ realization of the tentative nature of scientific theories, ideas. and processes in the text; (c) source of knowledge claims in the texts, which comprises a sophisticated understanding that scientific theories, ideas and processes mentioned in the texts can be challenged; (d) justification of knowledge claims in texts, which consists of acknowledging that scientific theories, ideas, and processes mentioned in texts require supports by experimental investigation and scientific evidence (Conley et al., 2004; Tsai et al., 2011). This domain of reading scientific texts in our developed instrument shows a clear grounding of nature of science, hence addressing the argument made by Yore and Tang (2022) that epistemic attributes is an integral component of reading scientific texts. We expected that students did not merely develop a “declarative” epistemic understanding on claims in scientific texts, while they needed to draw information from the text to justify these sorts of epistemic understanding.

Framing the Study

Reading History of Science and Socio-scientific Texts

Integrating history of science and socio-scientific issues into reading materials is advocated by science education researchers (Chen et al., 2022; Fazio et al., 2022). Understanding history of science (HOS) is a part of the scientific literacy, as it is a viable context for students’ acquisition of subject content knowledge (Matthews, 2014) and epistemic understanding (Clough, 2006; Dass, 2005). HOS contextualises science through narrating the events of scientific discovery by scientists in the past (Dagenais, 2010; Lin, Cheng, & Chang, 2010). It provides historical thinking concurrent to students’ thinking and highlight contrast between thinking at present and that in the past (Monk & Osborne, 1997). Therefore, it was argued that HOS serves a reading material for characterising students’ conceptual, procedural and epistemic understanding of science (Kim & Irving, 2010). Chen et al. (2022) tested if using history of science reading materials can promote middle school students’ epistemic beliefs. The instrument used by Chen et al. (2022) used measures generic epistemic belief only (see Conley et al., 2004). Various domains of reading comprehension, such as understanding and reasoning information in scientific texts as well as understanding diverse methods, evidence and procedures that derive scientific knowledge in the texts, were not measured in the study. This called for the need to develop an instrument that measure students' reading comprehension of HOS-based texts.

Apart from HOS, students at very young age are exposed to socio-scientific issues by reading (Boggs, Wilson, Ackland, Danna, & Grant, 2016; Zeidler, 2014). Socio-scientific issues (SSI) refers to use of scientific topics that engage students in dialogue, debate and discussion (Zeidler & Nichols, 2009). SSI are often controversial as they often require moral or ethical reasoning in the process of arriving decisions on solutions (Zeidler & Nichols, 2009). An example of SSI is climate change, as there are two-sided arguments on whether human causes climate change (Farber, 2007; Gardiner, 2010). Although SSI-based reading is perceived as a part of scientific literacy (Fazio & Gallagher, 2019; Fazio et al., 2022; Leu, Kinzer, Coiro, Castek, & Henry, 2017), different domains of reading literacy in SSI texts is seldom considered in measures. Stang Lund et al. (2019) investigated effects of textual and individual factors on upper secondary school students' integration of source and context. The outcome measure they used focused on how students corresponded textual elements with their sources (Stang Lund, Bråten, Brante, & Strømsø, 2017). Instrument used in these studies do not consider how students understand content, procedural and epistemic domains of SSI-based texts, as there is a lack of instrument measuring such students' reading comprehension of SSI-based texts.

Instruments Measuring Students' Reading Comprehension of Scientific Texts

Research in science education documented some instruments that measure students' reading comprehension of scientific texts. The texts used in these instruments were not related to history of science or socio-scientific issues, while the questions placed an exclusive focus on content domain. Härtig et al. (2022) developed a topic-specific reading test on the topic on atoms, coupled with questions on extracting information, drawing inferences and integrating information from the texts; Jian (2018) developed a reading instruments to test fourth-grade students' comprehension of text-pictorial materials, with questions testing participants' understanding of biological knowledge, extracting information and integrating text-picture information; Bernholt et al. (2022) constructed introductory reading texts for grade eight students to test their ability to extract information from single sentence, drawing inferences from two sentence and integrating prior knowledge into information presented in the texts. As urged by Yore and Tang (2022), there should be more research efforts in developing students' understanding of disciplinary and epistemic nature of scientific texts.

HOS and SSI reading materials also provide viable contexts for students' understanding of processes of science (Kolstø, 2008; Sadler & Dawson, 2012). As many researchers argued, scientific knowledge communicated in modern texts should not be regarded as an ended product, instead it should engage readers in the processes of science (Secko, Amend, & Friday, 2013; Cheung, Chan and Erduran, 2023; Chan, Cheung and Erduran, 2023). As stipulated in *Framework for K-12 Science Education*, reading scientific texts is one of scientific inquiry practices (NRC, 2012), as scientists read texts to gain a deeper understanding of science rather than only applying prior knowledge to understanding of scientific texts. More importantly, the questions in an instrument that measure students' reading comprehension in science need to include different disciplinary and epistemic domain of reading scientific texts (Yore & Tang, 2022). Thus, the section below presents the procedures of developing and validating RISHA which encompasses various domains of reading comprehension in science.

Research Questions

The current study was to develop and validate RISHA instrument that measures students' holistic reading comprehension of scientific texts. Two types of scientific texts were included in this study, HOS text on the discovery of watertight-bulkhead technology and SSI text on the debate on climate change. The following research questions underpin the present study:

RQ1. Is RISHA best characterised as one dimension or multiple dimensions (content, procedural and epistemic domains)?

RQ2. Are RISHA items effectively and appropriately aligned with the target population?

RQ3. Are there any differences between students' performance in various domains of reading scientific texts?

RQ4. Is there any substantial differential item functioning between subgroups of students with different genders and language statuses?

Methodology

We adopts a construct driven approach (Wilson, 2023) to align developed construct of holistic reading of scientific texts, collection of students' responses to the RISHA instrument, and interpretation of students' performance on reading scientific texts established from Rasch analysis (Rasch, 1966). Such quantitative methodology is grounded in positivist paradigm. The contexts of procedures of developing the RISHA are explained below.

Participants and Contexts

The focus of this study was a large-scale funded research project that aimed to improve junior secondary school students' reading in science in Hong Kong. Hong Kong is a multilingual and multicultural society where the official languages are Chinese and English. The local curriculum requires students to develop their competence in reading English, so they are considered as multilingual learners (Cheung and Pun, 2023). According to the local curriculum (CDC and HKEAA, 2017), reading across curriculum was one of the key foci, with promotion of reading in content subjects. Nonetheless, the curriculum did not explicitly mention disciplinary and epistemic nature of reading scientific texts. The second author sent invitation to secondary schools to invite students to participate in a program which improves students’ reading scientific texts in English. The teacher-in-charged promoted the programme and students voluntarily participated in the programme, together with their parents’ consents. The participants in this study were 168 junior secondary science students (grade 9) from three public schools. Ninth graders were the target group because they developed a comprehensive understanding of scientific knowledge on the contexts of the texts. In Hong Kong, students start to receive science instruction as an individual subject at seventh grade, while ninth graders already spent two years in getting used to reading scientific texts in science lessons. These public schools taught science in English. Seven students did not complete more than 50% of the items of the reading test, so their responses were removed. A total of 161 students’ responses (male = 82, female = 79, mean age= 14.14) were reported in this study (**Table 1**). Within this sample of participants, English was the second language of 51.6% of students, while English was the first language of 36.6% of students. The remaining students did not provide a valid response for their language status. Completion of RISHA takes place at lesson time which takes one hour at the beginning of the program, to avoid the influence of reading programme on their responses.

Table 1

Demographics of participants with valid responses in each school

Table 1. Demographics of participants with valid responses in each school

School	Gender		English as the first language		
	Female	Male	Yes	No	Not provided
School 1	60	56	38	66	12
School 2	19	6	13	9	3
School 3	0	20	8	8	4

Development of RISHA reading passages

The research team critically examined a range of reading materials from local suggested science curriculum resources, extra-curricular reading books, science education practitioner and research journals, websites on scientific advancements. Passages from textbooks, or content like those covered in local science textbooks were not targeted as students would not just recall what they read in responding to questions in RISHA. The reading materials of RISHA selected consists two passages in English: a HOS passage on the history of watertight-bulkhead technology (passage A) (Cheung, 2017) and another SSI passage on the debate of climate change (passage B) (Khishfe, Alshaya, BouJaoude, Mansour, & Alrudiyan, 2017; Lindsey & Dahlman, 2023). The HOS passage of history of wartertight-bulkhead technology describes how a famous Chinese navigator, Zhang He, uses watertight-bulkhead technology and the concept of density, to reduce the sinking of ships in long-mile navigation. Another SSI passage on the debate of climate change discusses contrasting viewpoints on whether human behaviour is its major cause. The identical topic for both passages was avoided because students' responses in one passage might depend on reading of another passage. These two passages were purposefully selected as the RISHA instrument can see whether students' performance on reading scientific text vary across students' familiarity with contexts. The contexts of these two passages, watertight-bulkhead technology and global warming were specifically selected owing to three reasons. Firstly, the respective topics, watertight-bulkhead technology and global warming, were selected because they were covered in grade 7-8 local science curriculum (CDC and HKEAA, 2017) such that students had some prior knowledge in order to respond to both passages. Secondly, the passages should consist of a key scientific idea, two competing theories related to scientific ideas, information that clarify the key scientific idea, and disciplinary scientific vocabularies. These elements provided optimal conditions for selecting and eliciting students' reasoning of information from scientific texts. Thirdly, both passages covered evolution of scientific knowledge: in the passage on watertight-bulkhead technology, it explained how ship-building technology evolved from the 13th century to the Ming Dynasty; in the passage on global warming, it first addressed global temperature since 1880 and then addressed the Kyoto

Conference in 1997, as well as modern techniques that study climate change. Such evolution of scientific knowledge was apt for eliciting students' responses to questions on epistemic domain of reading, which students can reason their epistemic understanding by drawing inferences on the information from the passages. Fourthly, we also followed Klopfer and Aikenhead (2022)'s criteria to select HOS and SSI texts: the selected HOS text presents a story supported by investigations and investigations which describes people and their assumptions and society's industrial-economic complex; the selected SSI text identifies a critical science concept, enculturing students' values to judge which conclusions can be trusted.

The two passages were rewritten by a group of expert teachers and researchers in the field of disciplinary literacy in science, which matches with the subject matter knowledge of grade 9 students and controls the difficulty in language between two passages. For example, the original phrase "Chinese junks" were replaced with "Chinese sailing ships" in the passage that describes history of watertight-bulkhead technology in inventing ships for sea navigation. The word counts of both passages are between 300 and 400 words, consisting of 27-28 lines of words. The structure of two passages follows introduction, elucidating the theme and transition to a science concept, scientific investigation of the phenomenon, and conclusion. Scientific investigation of the phenomenon was added to both passages because such information could elicit a higher level of students' performance in procedural domain. In the procedural domain, a more competent student could identify method and purpose of investigation, as well as identifying contextualized and decontextualized meaning of metalanguage and explaining how scientific evidence support a claim logically. Automated reading index (ARI) (Senter & Smith, 1967) and Flesch-Kincaid Grade Level (FKG) (Flesch, 2007) indicate that their readability suits the levels between ninth graders which are the target group of administration of RISHA. Both passages were reviewed by three experienced science teachers for content validity.

Development of RISHA assessment items

The development of RISHA questions follows Wilson (2023)'s approach, an iterative procedure with a pilot study and a main study. Before data collection, the items in RISHA were designed according to the *Items design* process (Wilson, 2023) which consists of three stages: construct map, item design and outcome spaces.

Stage 1: Construct Maps

Despite our conceptualisation of three domains in holistic reading comprehension of scientific texts, a refined construct map is needed to justify the scoring rubric of the reading test. Levels stipulated for each domain do not exhibit linear progression, which means one level in a domain does not correspond to the same level of another domain (Zhang & Browne, 2022). ***Appendix Table S1*** shows the construct map for each domain which accounts for disciplinary and epistemic nature of reading in science.

Stage 2: Item Design

With reference to the construct map, the operational definitions of each component in Figure 1 were drafted (***Appendix Table S2***). 12 questions were drafted for each passage (PA and PB), with each question tailoring a single component (content domain: CD1 to CD4; procedural domain: PD5 to PD8; epistemic domain: ED9 to ED12). In the first draft of RISHA, there is a combination of short-answered questions and multiple-choice questions. For CD and PD, the first draft of questions comprised questions that require short answers. The pilot results showed that students in junior science grade gave brief answers on these questions. One member of our expert teacher groups raised concerns on whether these questions were testing students' writing ability instead of reading comprehension of scientific texts. Therefore, to factor linguistic diversity and reading ability of grade level into our item design, most the questions in CD and PD were multiple choice questions in the final study, except for CD-3 and PD-7. CD-3 requires generation of inferences from the text, while PD-7 requires explanation of how scientific evidence in the text justify a claim.

For content domain, the range of score of CD-1 and CD-4 is between 0 and 1 because they require simple identification of key scientific concepts and meaning of scientific vocabularies respectively. Such performance corresponds to level 1 in content domain in the construct map. While for the procedural domain, the range of score of PD-5 and PD-6 is between 0 and 1 as both sub-components require cognitively less-demanding identification of method and purpose of scientific investigations. Such performance corresponds to level 1 in procedural domain in the construct map. The rest of components in content, procedural and epistemic domains adopted a scale of 0 to 2 as the performance levels of these components can be more cognitive-demanding and correspond to levels 0-2 in the construct map.

Stage 3: Outcome Space

Outcome space provides qualitative description to an assessment (Wilson, 2023). We developed a set of performance expectations (Zhai, 2022) based on two passages and pilot study results (***Appendix Table S3***). These performance expectations indicate what competence of reading scientific texts is to be assessed. In the process of developing the assessment, the performance expectations were modified according to the interview feedback of students and negotiation between raters. Coding rubrics on these performance expectations were further developed for each reading passage before data collection (***Appendix Table S4 and S5***).

Pilot Study

Before conducting the main study on 161 students, a pilot study was conducted to examine the clarity of questions and reading passages. 9 junior science seventh to ninth graders (six girls and three boys) were invited to complete the first draft of RISHA, while four students were invited for a follow-up semi-structured interview. In the first draft of RISHA, most questions were questions that required sentence-writing. The third author made a field note on students’ performance and interview responses to each question and this field note was reviewed by the first author. It was found that most students expressed familiarity with concepts of both passages, as the topics were covered in seventh-grade science curriculum. However, in the interview, students expressed linguistic challenges in answering the open-ended questions, specifically for epistemic domain. They also justified that their reading examination was mostly on multiple-choice questions. Therefore, we revised the RISHA questions such that the linguistic challenges of writing responses were reduced.

Data analysis

The quantitative data from main round students’ written responses to RISHA with students who participated in the main round study were reported in this paper. All responses to questions requiring students’ written texts (CD3 and PD7) were scored by the first author, the third author and two research assistants who have experienced in teaching scientific languages according to pre-designed coding rubric. The team rated few students’ responses and modified the scoring rubric according to students’ responses. For instance, the original scoring rubric for PA_PD7 requires students getting 2 marks to describe the role of multiple evidence in watertight-bulkhead technology. There was not any student giving such an explicit articulation of the role of multiple evidence. The scoring rubric was then modified as that student who mentioned both Zhang He and Macro Polo’s successes in using watertight-bulkhead technology for long-mile navigation can get 2 marks. This implied that students considered multiple evidence. Remaining responses were rated after several meetings of negotiations. Cohen’s *Kappa* was calculated for these questions. Cohen’s κ of PA_CD3 was 0.810 ($p < 0.001$); that of PA_PD7 was 0.963 ($p < 0.001$); that of PB_CD3 was 0.924 ($p <$

0.001); and that of PB_PD7= 0.943 (p < 0.001). The interrater reliability indicated a strong agreement (Cohen's κ > 0.7) between raters.

Rasch analysis was carried out to investigate the psychometric properties of the RISHA items. Rasch analysis, which is based on the principle that the probability of answering a question correctly depends on both the difficulty of the items and the ability of the respondents, computes the level of an item within an instrument (Hambleton & Jones, 1993). This method constructs measures in logit scales for both items difficulty and person ability measures, so item difficulty and person ability can be compared with each other on the same scale (Bond, Yan, & Heene, 2020). As some items in RISHA were scored polytomous, this study used partial credit Rasch model (PCM) specifying that each item has its own scoring scale (Masters, 1982). Rasch analysis was conducted by using ConQuest 5.22.4 (Wu, Adams, Wilson, & Haldane, 2007). Further information on how multidimensional Rasch analysis can answer each research questions is presented in the appendix (*Appendix Information S2*).

Findings

Dimensionality of RISHA

Principal components analysis (PCA) of Rasch residuals approach was used to check the possibility of additional dimensions of RISHA items. The eigenvalue of the first contrast is way above the cut-off value of 2.0 (3.15), which indicated additional dimensions present in RISHA items. Hence, the unidimensional Rasch model was unable to explain the variation in students' responses to RISHA. To further investigate this, we carried out PCA of Rasch residuals in the three-dimensional model and six-dimensional models. Given that students' reading comprehension of scientific texts were conceptualised in three dimensions, it is important to identify the most suitable Rasch model for our analysis. It was unknown if students' comprehension of scientific texts vary across the HOS passage and the SSI passage. Six dimensions, with respective content, procedural and epistemic dimensions in each passage, might also be a potential solution for Rasch analysis. By organizing items into content, procedural and epistemic domains, we ran PCA of Rasch residuals in these three unidimensional models. The results of PCA of Rasch residuals for the three unidimensional showed that the eigenvalue of first contrast for epistemic domain was greater than the threshold of 2.0 (e.g., 2.55; ***Table 2***). Hence, six-dimensional model might be a better option. After conducing unidimensional Rasch models with domains separately, the PCA residuals indicated eigenvalues below 2.0 (e.g., content domain of passage A = 1.62; procedural domain of passage A = 1.60; epistemic domain of passage A: 1.56; content domain of passage B: 1.47; procedural domain of passage B: 1.56; epistemic domain of passage B: 1.43; ***Table 2***).

Apart from PCA of Rasch residuals, Chi-squared test of parameter equality and values for the likelihood ratio test can be used to select the best fit model. According to Wu et al. (2007), a better model fit was indicated by a smaller Chi-square values. The values shown in the likelihood ratio test include the final deviance (FD) of each model which refers to the lack of fit between data and model, as well as AIC and BIC (information-based criteria scores) (Sbeglia & Nehm, 2019). Smaller FD and information-based criteria scores provide evidence for a better model fit which takes into account free parameters. As indicated in ***Table 2***, the six-dimensional model [$\chi^2(18, N=161) = 769.74$, $p < 0.01$] had a smaller Chi-square value than the three-dimensional model [$\chi^2(21, N=161) = 789.89$, $p < 0.01$] and unidimensional model [$\chi^2(23, N=161) = 1333.31$, $p < 0.01$], indicating that the six-dimensional model has a better fit over the three-dimensional and unidimensional model. FD and information-based criteria scores of the six-dimensional model (FD = 5774.74, AIC= 5896.74, BIC= 6084.71) were lower than those of unidimensional (FD = 5971.42, AIC= 6053.42, BIC= 6179.76) and three-dimensional models (FD = 5960.43, AIC= 6052.43, BIC= 6194.17). Hence, the six-dimensional model had a significantly better fit over the three-dimensional and unidimensional models.

Item fit and reliability

The EAP/PV reliability indices reflects the reliability of each domain or entire instrument by dividing the variance of individual expected posteriori ability estimates by the software' estimated total variance of latent ability (Wu et al., 2007). When all items were modelled unidimensionally, EAP/PV reliability of the entire RISHA assessment was found to be 0.809. Person separation reliability of the entire RISHA assessment was found to be 0.796. As an item separation reliability statistic, EAP/PV reliability statistic is calculated by dividing the individual variance expected a posterior ability estimate by the total estimated latent ability

Model	Model categorisation	PCA of Rasch residuals	Chi-square test for equality and likelihood ration test				
			χ^2	*df*	FD	AIC	BC
Model with one dimension	All items (24)	3.15	1333.31	23	5971.42	6053.42	6179.76
Model with three dimensions	*Domain for reading scientific texts:* (1) Content domain (8 items): PA_CD1 to PA_CD4, PB_CD1 to PB_CD4 (2) Procedural domain (8 items): PA_PD5 to PA_PD8, PB_PD5 to PB_PD8 (3) Epistemic domain (8 items): PA_ED9 to PA_ED12, PB_ED9 to PB_ED12	(1): 1.63 (2): 1.70 (3): 2.55	789.89	21	5960.43	6052.43	6194.17
Model with six dimensions	*Domain for reading scientific texts and types of scientific text* (1) Content domain for HOS (4 items): PA_CD1 to PA_CD4 (2) Procedural domain for HOS (4 items): PA_PD5 to PA_PD8 (3) Epistemic domain for HOS (4 items): PA_ED9 to PA_ED12 (4) Content domain for SSI (4 items): PB_CD1 to PB_CD4 (5) Procedural domain for SSI (4 items): PB_PD5 to PB_PD8 (6) Epistemic domain for SSI (4 items): PB_ED9 to PB_ED12	(1): 1.62 (2): 1.6 (3): 1.56 (4): 1.47 (5): 1.56 (6): 1.43	769.74	18	5774.74	5896.74	6084.71

Abbreviations: PCA- principle component analysis; χ^2- chi-square values; *df*- degree of freedom; FD-Final deviance; AIC-Akaike information criterion; BIC-Bayesian information criterion

Table 2

Comparison of the structure and outcomes of different Rasch models before and after rescaling of epistemic domain

variance (Rauch & Hartig, 2010). Similar to classical test theory, EAP/PV reliability statistic can be interpreted in a similar way to Cronbach's alpha (Wu et al., 2007), with acceptable values above 0.70 (Grigg & Manderson, 2016). As the value of EAP/PV for the entire instrument was well above the cut-off value of 0.7, representing a good item separation reliability. EAP/PV reliability statistics of multiple dimensions in RISHA were also examined in ConQuest software. EAP/PV reliabilities of all dimensions (***Table 3***) were above 0.7, which justified the reliabilities of the domains of measurement.

Table 3

Item separation reliability and means for domains of reading scientific texts

Dimension	***EAP/PV reliability***
Content domain (HOS)	0.797
Procedural domain (HOS)	0.796
Epistemic domain (HOS)	0.726
Content domain (SSI)	0.758
Procedural domain (SSI)	0.830
Epistemic domain (SSI)	0.779

Abbreviations: HOS: history of science; SSI: socio-scientific issues

The item fit measurement for the six-dimensional model was further examined. Three parameters computed by Conquest (Wu et al., 2007) indicated item-fit measurement: item-total correlation (point biserial coefficient) and weighted mean squares item fit statistics (WMNSQ) (***Table 4***). The classical item-total correlation has a range between 0 and 1, with a higher value indicating a stronger discrimination of that specific item. All items were falling into the range between 0.212 and 0.605, indicating that the items were of acceptable discrimination. The values of WMNSQ and MNSQ were used as measures of Rasch fit. We chose to use a more stringent threshold for WMNSQ, with values ranging from 0.8 to 1.2 indicating acceptable range of fit (Boone, Staver, & Yale, 2013). A WMNSQ of 1.0 justifies expected fit for responses to the Rasch model. Values greater than 1.0 signal the presence of more variation than expected in Rasch model, while values lower than 1.0 show data overfit. WMNSQ values of all items were within the range of between 0.87 and 1.12.

Table 4

Items	**Item point-biserial correlation**	**MNSQ**	**Weight MNSQ**	**Location (SE) (logit)**	**Threshold 1 (logit)**	**Threshold 2 (logit)**
Content domain in reading HOS passage						
PA_CD1	0.416	0.94	0.96	0.060 (0.123)		
PA_CD2	0.519	1.02	1.00	-0.496 (0.093)	-1.063	0.071
PA_CD3	0.418	1.08	1.09	-0.052 (0.093)	-0.583	0.479
PA_CD4	0.312	1.00	1.01	0.487 (0.180)		
Procedural domain in reading HOS passage						
PA_PD5	0.605	0.83	0.91	-1.138 (0.106)		
PA_PD6	0.434	0.96	1.01	-0.379 (0.107)		
PA_PD7	0.402	0.91	1.12	1.760 (0.110)	0.890	2.630
PA_PD8	0.465	1.01	1.08	-0.243 (0.186)	-1.146	0.660
Epistemic domain in reading HOS passage						
PA_ED9	0.212	0.98	1.04	-0.083 (0.114)	-1.333	1.166
PA_ED10	0.319	1.03	1.06	0.425 (0.119)	-1.604	2.455
PA_ED11	0.285	0.96	1.01	0.058 (0.116)	-1.433	1.548
PA_ED12	0.369	0.84	0.91	-0.399 (0.202)	-2.160	1.361
Content domain in reading SSI passage						
PB_CD1	0.380	1.00	1.02	-0.702 (0.128)		
PB_CD2	0.440	1.09	1.10	0.306 (0.109)	-0.811	1.424
PB_CD3	0.555	0.86	0.87	1.455 (0.121)	-0.328	3.239
PB_CD4	0.443	0.93	0.94	-1.059 (0.207)		
Procedural domain in reading SSI passage						
PB_PD5	0.441	1.01	1.01	-0.707 (0.101)		
PB_PD6	0.378	1.01	1.01	0.172 (0.102)		
PB_PD7	0.566	0.88	0.96	0.610 (0.090)	0.059	1.160
PB_PD8	0.483	1.11	1.14	-0.074 (0.169)	-0.826	0.678
Epistemic domain in reading SSI passage						
PB_ED9	0.451	1.02	0.97	0.136 (0.103)	-0.712	0.984
PB_ED10	0.513	0.97	1.01	-0.469 (0.103)	-1.856	0.918
PB_ED11	0.479	0.93	0.96	0.663 (0.109)	-1.038	2.365
PB_ED12	0.475	1.00	1.05	-0.33 (0.182)	-1.693	1.032

Abbreviations: HOS: history of science; SSI: socio-scientific issues; PA: passage A; PB: passage B; CD: content domain; PD: procedural domain; ED: epistemic domain

Item fit and item characteristics

Difference in Students' Performance in Domains

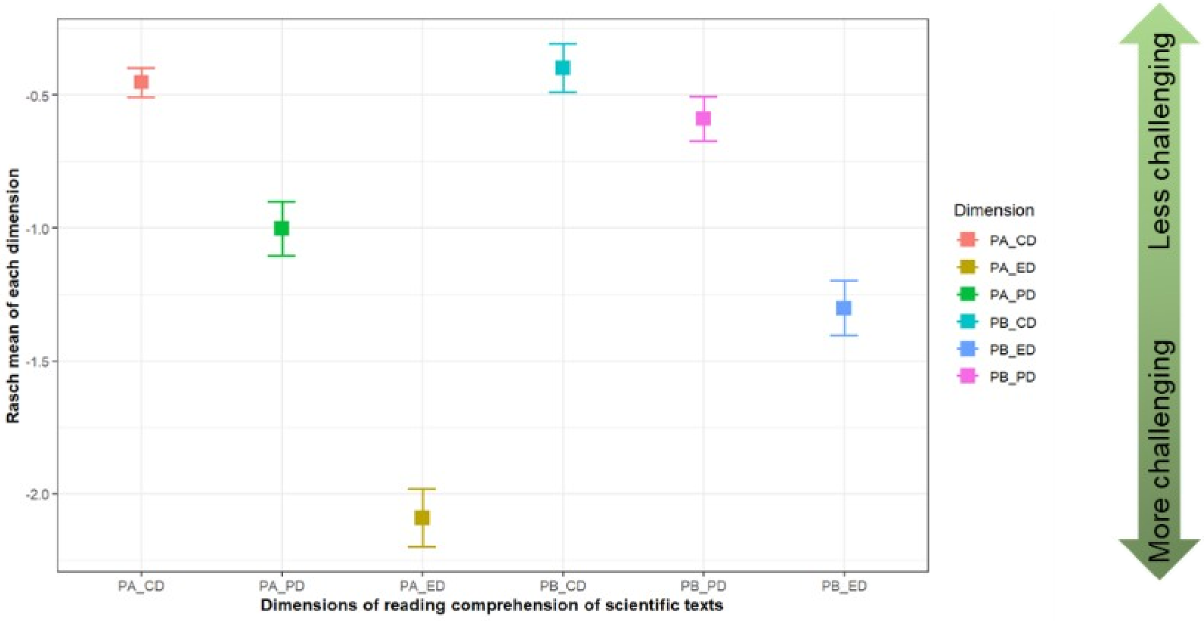
The Wright map (***Appendix Figure S1***) presented the average difficulty of each item with reference to students' performance in six domains of reading scientific texts. It shows the order of item difficulty, the vertical scale, ranging from -7 to +4 logits, show estimates of students' ability and values of task difficulty. The left six columns, with average every 1.9 students are represented by an "X", indicating student ability in reading scientific texts for each domain. RISHA items were evenly distributed around the mean which is default to be 0.00 logits. These items targeted persons relatively well. The means of each dimension were lower than 0.00 logits, indicating that the items were slightly more difficult for students in the population to answer. The most difficult item was PA_PD7, whereas the easiest item was PA_PD5. In the Wright map (***Appendix Figure S1***), students' average ability in *content domains* for both passages (first and forth column) were higher than that in *procedural domains* for both passages (second and fifth column), whereas their ability in *procedural domains* was higher than that in *epistemic domains* (third and sixth column). Moreover, students' average ability in *content domain* in passage A (historical discovery of watertight-bulkhead technology) was lower than that in passage B (socio-scientific text on causes of climate change). Similar observations were found in *procedural* and *epistemic* domains.

To further confirm the differences in students' performances in six domains of reading scientific texts (***Figure 2***), one-way repeated measures ANOVA was conducted. There was a significant difference in students' ability in six domains [$F(5, 276.1) = 28.29$, $p < 0.001$]. Pairwise *t*-test comparisons with Bonferroni correction showed that content domain of passage A (PA_CD) ($M = -0.452$, $SE = 0.055$) had a significantly higher mean then procedural (PA_PD) and epistemic domain (PA_ED) of passage A [$t(161) = 3.65$, $p < 0.001$], whereas PA_PD ($M = -1.002$, $SE = 0.102$) had a significantly higher mean than PA_ED ($M = -2.091$, $SE = 0.109$) [$t(161) = 11.4$, $p < 0.001$]. On the other hand, content domain of passage

B (PB_CD) had a significantly higher mean (*M* = -0.398, *SE* = 0.092) than epistemic domain of passage B (PB_ED) [*t*(161) = 6.63, *p* <0.001]. However, content domain of passage B (PB_CD) did not have a significantly higher mean than its procedural domain (PB_PD) [*t*(161) = 2.12, *p*=0.534]. PB_PD (*M* = -0.588, *SE* = 0.083) had a significantly higher mean than PB_ED (*M* = -1.302, *SE* = 0.103) [*t*(161) = 6.93, *p* <0.001]. The most challenging domain of reading scientific texts for students was epistemic domain for both passages. There was not any difference in students' performance in content and procedural domain across two passages. However, students' performance in epistemic domain in passage A (PA_ED) was significantly lower than that in passage B (PB_ED) [*t*(161) = -3.99, *p* <0.001]. This shows that students' epistemic understanding of scientific text in debates on climate change was better than that of the text on history of watertight-bulkhead technology. Epistemic understanding of scientific texts can depend on specific contexts in the passage.

Figure 2

Students' Rasch logits mean in different domains of RISHA



Differential Item Functioning

Differential item functioning (DIF) was examined for any differences in performance in items between subgroups. According to Zwick, Thayer, and Lewis (1999), negligible DIF has a value below 0.43 logits; slight to moderate DIF has a value between 0.43 and 0.64 logits; moderate to large DIF has a value larger than 0.64 logits. Analysis of DIF indicated that DIF contrasts for gender ranged from 0.002 to 0.366 logits, so all items exhibited negligible DIF. Moreover, DIF contrasts for language status ranged from 0.025 to 0.420 logits, so all items exhibited negligible DIF for students of different language status. The items did not exhibit substantial subgroup differences in terms of students' gender and language status.

Discussion and Conclusion

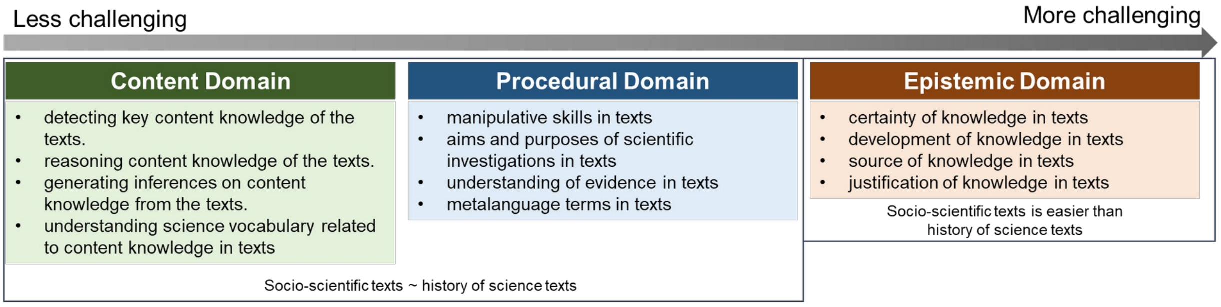
Developing an instrument that measures students' disciplinary and epistemic reading of scientific texts is crucial. In particular, science educators have recently called from moving on from simple view of reading to a more holistic view of reading (K. S. Tang et al., 2022; Yore & Tang, 2022). This study developed a Reading in Science Holistic Assessment (RISHA), which can be viewed as comprising content domain, procedural domain, and epistemic domain of reading. Although numerous research studies investigated students' reading of scientific texts (Chen et al., 2022; Fazio et al., 2022), they did not use a specific instrument to measure students' disciplinary and epistemic reading of scientific texts.

Reading scientific texts in different domains in various contexts

In our initially proposed framework, content, procedural and epistemic domains of reading scientific texts centres around holistic reading of scientific texts. This adds on to previous research studies that just consider content domain (e.g., Oliveras, Márquez, and Sanmartí, 2013) or epistemic domain alone (e.g., Chen et al., 2022). The findings also contribute to science education literature that these domains exhibit a hierarchical level. The content domain appears to be the least challenging while the epistemic domain appears to be the most challenging (**Figure 3**). It could be attributed to the reason that science/literacy lessons just focused on students' identification or reasoning of texts, instead of discussing certainty, development, source, and justification of scientific information in the texts. Another potential reason was that students played a passive role in receiving information from scientific texts, instead of critically evaluating where knowledge comes from and paying attention to the scientific investigation that generates the knowledge. We herein propose that science teacher could start teaching how to read scientific texts from the easier domain. Likewise, science teachers can initially teach students how to detect, reason, infer content knowledge and scientific vocabularies. Afterwards, teachers can work with students to pay attention to what the method and purpose of a scientific investigation are described in the texts, and how these scientific investigations generate evidence to support a claim and the metalanguage associated with the investigation. Eventually, the teacher could ask students to reflect on the certainty, development, source, and justification of knowledge in scientific texts.

Figure 3

Hierarchy of different domains of reading scientific texts



More importantly, students' performance in epistemic domain is less ideal in a HOS text compared to SSI context. The current finding further supported the study by Khishfe (2022) that students generally acquired a better epistemic understanding in SSI contexts compared to HOS context. An alternative account for this finding was that the SSI context provided in RISHA, global warming, was more familiar to students compared to the history of watertight-bulkhead technology. The concept of 'adherence" of epistemic understanding (Leung, 2020) might explain this. As students felt that HOS texts were more distant to them compared to SSI texts, students were less likely to activate their epistemic resources as these resources were context-specific (Strømsø et al., 2008). For example, compared to the context of global warming, students might think that the discovery of watertight-bulkhead technology happened in the 13th century and cannot appreciate how knowledge was formed, developed and justified at ancient times. Therefore, this provided further counterevidence of using a generic questionnaire to measure students' epistemic belief about scientific texts (Yang et al., 2016; Chen et al., 2022). Most importantly, we propose that for science teachers who would like to develop students' holistic reading of scientific texts, they could start from SSI texts rather than HOS texts as students might develop a stronger epistemic adhere to the texts.

Performance of students of different language statuses

One important finding of this study was that negligible DIF existed between students whose first language is English and students whose first language is not English. This shows that RISHA can be applied to diverse groups of students with various linguistic backgrounds. As the major aim of RISHA is to measure students' holistic reading of science, this study developed a reading instrument that cater for a diverse population of students. The finding could be attributed to our theoretical framing that holistic reading of scientific texts did not comprise decoding vocabularies and understanding sentence structures which heavily rely on linguistic backgrounds. Another possible reason accounting for the result was that English learners in Hong Kong started to acquire English in pre-school education, hence they acquired disciplinary and epistemic reading in the same fashion as learners with English as their first language.

Limitations and future research directions

Content and language integrated learning has been widely promoted in science classrooms (Pun, Fu and Cheung, 2023). Hence, there is an increasing popularity in the notion of reading-science learning (Holliday, Yore, & Alvermann, 1994). Despite the importance of RISHA in understanding the impacts of reading-science intervention on students' performance in reading scientific texts, this paper did not report on how a teaching intervention based on the framework influences students' responses before and after the intervention. We proposed that RISHA can be applied to measuring students' changes in reading of scientific texts in future research. Another limitation of this study was that RISHA only covered two domains of science, earth science and physics. We acknowledge that reading chemistry text is different from reading a physics text, as chemistry text involves students' translation across sub-microscopic, macroscopic, and symbolic levels while physics text involves students' translation across unseen, seen and symbolic levels. Owing to the limited time for students to complete the assessment in the program, future research can study the differences in students' holistic reading across two domains of scientific texts, particularly multimodal texts (Cheung and Winterbottom, 2021; Cheung and Winterbottom, 2023). Moreover, future research can explore how to modify multiple-choice questions to explore students' thinking processes in reading scientific texts.

Another important limitation is that students' holistic reading of scientific texts depends on the nature or type of history of science and socio-scientific texts. For example, in comparison with non-local issues, familiarity with local issues can facilitate learners' decision-making on local socio-scientific texts (Wiyarsi, Çalik, Priyambodo and Dina, 2023; Tuncay, Yılmaz-Tüzün and Teksoz, 2012). Future research studies can explore students' holistic reading of scientific texts grounded in the same type of contexts, which is either history of science or socio-scientific issues.

Despite these limitations, the current instrument has demonstrated a strong construct validity and reliability. To better understand what factors are linked to students' different domains of reading scientific texts, future research can examine how content knowledge, vocabularies and metacognitive reading strategies predict these domains of reading scientific texts. Previous studies evidenced that content knowledge, vocabularies and metacognitive strategies can predict students' reading scientific texts (e.g., Härtig et al., 2022; Stang Lund et al., 2019). However, the instrument they used did not consider students' disciplinary and holistic reading of scientific texts. By examining the predictors of different domains of reading scientific texts, teacher's intervention can emphasise on specific factors that promote reading-science learning.

References

Abrahams, I., Reiss, M. J., & Sharpe, R. M. (2013). The assessment of practical work in school science. *Studies in Science Education, 49*(2), 209-251.

Ali, S. H., & Ruit, K. G. (2015). The Impact of item flaws, testing at low cognitive level, and low distractor functioning on multiple-choice question quality. *Perspectives on medical education, 4*, 244-251.

Authority, C. D. C. a. H. K. E. a. A. (2017). *Science education: key learning area curriculum guide (primary 1—secondary 6)*. Hong Kong: Curriculum Development Council.

Avsar Erumit, B., & Yuksel, T. (2023). Developing and Using Physical Dynamic Models On Socioscientific Issues to Present Nature Of Science Ideas. *International Journal of Science and Mathematics Education, 21*(4), 1031-1056.

Bernholt, S., Härtig, H., & Retelsdorf, J. (2022). Reproduction Rather than Comprehension? Analysis of Gains in Students' Science Text Comprehension. *Research in Science Education, 53*(3), 493-506. doi:10.1007/s11165-022-10066-6

Boggs, G. L., Wilson, N. S., Ackland, R. T., Danna, S., & Grant, K. B. (2016). Beyond The Lorax: Examining Children's Books on Climate Change. *The Reading Teacher, 69*(6), 665-675.

Bond, T., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences*: Routledge.

Boone, W. J., Staver, J. R., & Yale, M. S. (2013). *Rasch analysis in the human sciences*: Springer.

Chan, H. Y., Cheung, K. K. C., & Erduran, S. (2023). Science communication in the media and human mobility during the COVID-19 pandemic: a time series and content analysis. *Public Health*, *218*, 106-113.

Chen, S.-Y., Chen, C.-H., & Liu, S.-Y. (2022). History of Science Reading Materials as Everyday Homework to Improve Middle School Students' Epistemological Beliefs about Science. *International Journal of Science and Mathematics Education*. doi:10.1007/s10763-022-10285-3

Cheung, C. M. (2017). "Science (S1-3) – Learning and Teaching Resources" (Set C). Retrieved from https://cd1.edb.hkedcity.net/cd/science/multimedia/media1_1_e.htm

Cheung, K. K. C., & Winterbottom, M. (2021). Exploring students' visualisation competence with photomicrographs of villi. *International Journal of Science Education*, *43*(14), 2290-2315.

Cheung, K. K. C., Chan, H. Y., & Erduran, S. (2023). Communicating science in the COVID-19 news in the UK during Omicron waves: exploring representations of nature of science with epistemic network analysis. *Humanities and Social Sciences Communications*, *10*(1), 1-14.

Cheung, K. K. C., & Pun, J. K. (2023). The use of epistemic network analysis in analysing classroom discourse in EMI-science classrooms. In *Qualitative Research Methods in English Medium Instruction for Emerging Researchers* (pp. 33-44). Routledge.

Cheung, K. K. C., & Sonkqayi, G. (2023). Students' science achievement in cognitive domains: effects of practical work and clarity of instruction. *Research in Science & Technological Education*, 1-18.

Cheung, K. K. C., & Winterbottom, M. (2023). Students' integration of textbook representations into their understanding of photomicrographs: epistemic network analysis. *Research in Science & Technological Education*, 41(2), 544-563.

Clough, M. P. (2006). Learners' responses to the demands of conceptual change: Considerations for effective nature of science instruction. *Science & Education, 15*, 463-494.

Conley, A. M., Pintrich, P. R., Vekiri, I., & Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. *Contemporary Educational Psychology, 29*(2), 186-204. doi:10.1016/j.cedpsych.2004.01.004

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

Dagenais, A. (2010). Teaching High School Physics With a Story-Line. *Interchange, 41*(4), 335-345. doi:10.1007/s10780-010-9134-z

Dass, P. M. (2005). Understanding the nature of scientific enterprise (NOSE) through a discourse with its history: The influence of an undergraduate 'history of science'course. *International Journal of Science and Mathematics Education, 3,* 87-115.

Fang, Z. (2006). The Language Demands of Science Reading in Middle School. *International Journal of Science Education, 28*(5), 491-520. doi:10.1080/09500690500339092

Fang, Z. (2008). Going beyond the fab five: Helping students cope with the unique linguistic challenges of expository reading in intermediate grades. *Journal of Adolescent & Adult Literacy, 51*(6), 476-487.

Farber, D. A. (2007). Adapting to climate change: who should pay. *J. Land Use & Envtl. L., 23,* 1.

Fazio, X., & Gallagher, T. L. (2019). Science and language integration in elementary classrooms: Instructional enactments and student learning outcomes. *Research in Science Education, 49*, 959-976.

Fazio, X., Gallagher, T. L., & DeKlerk, C. (2022). Exploring adolescents' critical reading of socioscientific topics using multimodal texts. *International Journal of Science and Mathematics Education*, 1-24.

Flesch, R. (2007). Flesch-Kincaid readability test. *Retrieved October, 26*(3), 2007.

Ford, M. J., & Wargo, B. M. (2012). Dialogic framing of scientific content for conceptual and epistemic understanding. *Science Education, 96*(3), 369-391.

Gardiner, S. M. (2010). Ethics and climate change: an introduction. *Wiley Interdisciplinary Reviews: Climate Change, 1*(1), 54-66.

Grigg, K., & Manderson, L. (2016). The Australian racism, acceptance, and cultural-ethnocentrism scale (RACES): item response theory findings. *International journal for equity in health, 15*, 1-16.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational measurement: issues and practice, 12*(3), 38-47.

Härtig, H., Bernholt, S., Fraser, N., Cromley, J. G., & Retelsdorf, J. (2022). Comparing Reading Comprehension of Narrative and Expository Texts Based on the Direct and Inferential Mediation Model. *International Journal of Science and Mathematics Education, 20*(1), 17-41. doi:10.1007/s10763-022-10302-5

Holliday, W. G., Yore, L. D., & Alvermann, D. E. (1994). The reading–science learning–writing connection: Breakthroughs, barriers, and promises. *Journal of Research in Science Teaching, 31*(9), 877-893.

Jian, Y.-C. (2018). Teaching Fourth-Grade Students of Different Reading Abilities to Read Biological Illustrations and Integrate In-Text Information: an Empirical Experiment. *Research in Science Education, 50*(6), 2269-2282. doi:10.1007/s11165-018-9778-8

Khishfe, R. (2022). Improving Students' Conceptions of Nature of Science: A Review of the Literature. *Science & Education*. doi:10.1007/s11191-022-00390-8

Khishfe, R., Alshaya, F. S., BouJaoude, S., Mansour, N., & Alrudiyan, K. I. (2017). Students' understandings of nature of science and their arguments in the context of four socio-scientific issues. *International Journal of Science Education, 39*(3), 299-334.

Kim, S. Y., & Irving, K. E. (2010). History of science as an instructional context: Student learning in genetics and nature of science. *Science & Education, 19,* 187-215.

Klopfer, L. E., & Aikenhead, G. S. (2022). Humanistic science education: The history of science and other relevant contexts. *Science Education, 106*(3), 490-504.

Kolstø, S. D. (2006). Patterns in students' argumentation confronted with a risk-focused socio-scientific issue. *International Journal of Science Education, 28*(14), 1689-1716.

Kolstø, S. D. (2008). Science education for democratic citizenship through the use of the history of science. *Science & Education, 17,* 977-997.

Lammers, A., Goedhart, M. J., & Avraamidou, L. (2019). Reading and synthesising science texts using a scientific argumentation model by undergraduate biology students. *International Journal of Science Education, 41*(16), 2323-2346.

Lederman, N. G. (2013). Nature of science: Past, present, and future. In *Handbook of research on science education* (pp. 831-879): Routledge.

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching, 39*(6), 497-521.

Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2017). New literacies: A dual-level theory of the changing nature of literacy, instruction, and assessment. *Journal of Education, 197*(2), 1-18.

Leung, J. S. C. (2020). Students' adherences to epistemic understanding in evaluating scientific claims. *Science Education, 104*(2), 164-192.

Li, C., Yu, J., & Li, G. (2023). Development of the Representation of the Nature of Science in Textbooks: the Case of High School Biology Textbooks in Mainland China. *International Journal of Science and Mathematics Education, 21*(6), 1749-1768.

Lin, C. Y., Cheng, J. H., & Chang, W. H. (2010). Making science vivid: Using a historical episodes map. *International Journal of Science Education, 32*(18), 2521-2531.

Lindsey, R., & Dahlman, L. (2023). Climate Change: Global Temperature. Retrieved from https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature#:~:text=Earth's%20temperature%20has%20risen%20by,based%20on%20NOAA's%20temperature%20data.

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174. doi:10.1007/BF02296272

Matkins, J. J., & Bell, R. L. (2007). Awakening the scientist inside: Global climate change and the nature of science in an elementary science methods course. *Journal of Science Teacher Education, 18*(2), 137-163.

Matthews, M. R. (2014). *Science teaching: The contribution of history and philosophy of science*: Routledge.

Mccomas, W. F. (2011). The history of science and the future of science education: a typology of approaches to history of science in science instruction. In *Adapting historical knowledge production to the classroom* (pp. 37-53): Brill.

Millar, R., Lubben, F., Got, R., & Duggan, S. (1994). Investigating in the school science laboratory: conceptual and procedural knowledge and their influence on performance. *Research Papers in Education, 9*(2), 207-248.

Monk, M., & Osborne, J. (1997). Placing the history and philosophy of science on the curriculum: A model for the development of pedagogy. *Science Education, 81*(4), 405-424.

Nigro, R. G., & Trivelato, S. F. (2012). Knowledge, its application, and attitudes associated with the reading of diverse genres of science texts. *International Journal of Science Education, 34*(16), 2529-2564.

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education, 87*(2), 224-240.

Norris, S. P., & Phillips, L. M. (2008). Reading as inquiry. In *Teaching Scientific Inquiry* (pp. 233-262): Brill.

OECD. (2019). *PISA 2018 Assessment and Analytical Framework*.

Oliveras, B., Márquez, C., & Sanmartí, N. (2013). The Use of Newspaper Articles as a Tool to Develop Critical Thinking in Science Classes. *International Journal of Science Education, 35*(6), 885-905. doi:10.1080/09500693.2011.586736

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction, 19*(3), 228-242. doi:10.1016/j.learninstruc.2008.04.003

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill.

Pun, J. K., Fu, X., & Cheung, K. K. C. (2023). Language challenges and coping strategies in English Medium Instruction (EMI) science classrooms: a critical review of literature. *Studies in Science Education*, 1-32.

Quellmalz, E., & Hoskyn, J. (1996). Classroom assessment of reasoning strategies. In *Handbook of classroom assessment* (pp. 103-130): Elsevier.

Rauch, D. P., & Hartig, J. (2010). Multiple-choice versus open-ended response formats of reading test items: A two-dimensional IRT analysis. *Psychological Test and Assessment Modeling, 52*(4), 354.

Roach, L. E., & Wandersee, J. H. (1993). Short Story Science. Using Historical Vignettes as a Teaching Tool. *Science Teacher, 60*(6), 18-21.

Sadler, T. D., Chambers, F. W., & Zeidler, D. L. (2004). Student conceptualizations of the nature of science in response to a socioscientific issue. *International Journal of Science Education, 26*(4), 387-409.

Sadler, T. D., & Dawson, V. (2012). Socio-scientific issues in science education: Contexts for the promotion of key learning outcomes. *Second international handbook of science education*, 799-809.

Sbeglia, G. C., & Nehm, R. H. (2019). Do you see what I-SEA? A Rasch analysis of the psychometric properties of the Inventory of Student Evolution Acceptance. *Science Education, 103*(2), 287-316.

Secko, D. M., Amend, E., & Friday, T. (2013). Four models of science journalism: A synthesis and practical assessment. *Journalism Practice, 7*(1), 62-80.

Senter, R., & Smith, E. A. (1967). *Automated readability index*. Retrieved from

Shepardson, D. P., & Gummer, E. S. (2001). A framework for thinking about and planning classroom assessments in science. *Assessment in science: A guide to professional development and classroom practice*, 83-98.

Shymansky, J. A., Yore, L. D., & Good, R. (1991). Elementary school teachers' beliefs about and perceptions of elementary school science, science reading, science textbooks, and supportive instructional factors. *Journal of Research in Science Teaching, 28*(5), 437-454.

Stang Lund, E., Bråten, I., Brandmo, C., Brante, E. W., & Strømsø, H. I. (2019). Direct and indirect effects of textual and individual factors on source-content integration when reading about a socio-scientific issue. *Reading and Writing, 32*, 335-356.

Stang Lund, E., Bråten, I., Brante, E. W., & Strømsø, H. I. (2017). Memory for textual conflicts predicts sourcing when adolescents read multiple expository texts. *Reading Psychology, 38*(4), 417-437.

Strømsø, H. I., & Bråten, I. (2009). Beliefs about knowledge and knowing and multiple-text comprehension among upper secondary students. *Educational Psychology, 29*(4), 425-445.

Strømsø, H. I., Bråten, I., & Samuelstuen, M. S. (2008). Dimensions of topic-specific epistemological beliefs as predictors of multiple text understanding. *Learning and Instruction, 18*(6), 513-527.

Symons, C. (2017). Supporting emergent bilinguals’ argumentation: Evaluating evidence in informational science texts. *Linguistics and Education, 38*, 79-91.

Tang, K.-S. (2021). The interconnections among metadiscourse, metalanguage, and metacognition: Manifestation and application in classroom discourse. *Linguistics and Education, 65*, 100977.

Tang, K.-S., & Rappa, N. A. (2021). The role of metalanguage in an explicit literacy instruction on scientific explanation. *International Journal of Science and Mathematics Education, 19*, 1311-1331.

Tang, K. S., Lin, S.-W., & Kaur, B. (2022). Mapping and Extending the Theoretical Perspectives of Reading in Science and Mathematics Education Research. *International Journal of Science and Mathematics Education*. doi:10.1007/s10763-022-10322-1

Tsai, C.-C., Jessie Ho, H. N., Liang, J.-C., & Lin, H.-M. (2011). Scientific epistemic beliefs, conceptions of learning science and self-efficacy of learning science among high school students. *Learning and Instruction*. doi:10.1016/j.learninstruc.2011.05.002

Tuncay, B., Yılmaz-Tüzün, Ö., & Teksoz, G. T. (2012). Moral reasoning patterns and influential factors in the context of environmental problems. *Environmental Education Research, 18*(4), 485-505.

Wang, J.-R., & Chen, S.-F. (2016). Development and Validation of an Online Dynamic Assessment for Raising Students’ Comprehension of Science Text. *International Journal of Science and Mathematics Education, 14*(3), 373-389. doi:10.1007/s10763-014-9575-4

Wang, J.-R., Chen, S.-F., Fang, I., & Chou, C.-T. (2012). Development of a test assessing elementary students’ comprehension of science texts. *International Journal of Science and Mathematics Education, 10*(4), 955 –973.

Wellington, J., & Osborne, J. (2001). *Language and literacy in science education*: McGraw-Hill Education (UK).

Wilson, M. (2023). *Constructing measures: An item response modeling approach*: Taylor & Francis.

Wiyarsi, A., Çalik, M., Priyambodo, E., & Dina, D. (2023). Indonesian Prospective Teachers’ Scientific Habits of Mind: A Cross-Grade Study in the Context of Local and Global Socio-scientific Issues. *Science & Education*, 1-27.

Wu, M., Adams, R., Wilson, M., & Haldane, S. (2007). ACER ConQuest 2.0: General item response modelling software [computer program manual]. *Camberwell, Australia: Australian Council for Educational Research*.

Yang, F.-Y., Chang, C.-C., Chen, L.-L., & Chen, Y.-C. (2016). Exploring learners’ beliefs about science reading and scientific epistemic beliefs, and their relations with science text understanding. *International Journal of Science Education, 38*(10), 1591-1606.

Yore, L. D., Hand, B., Goldman, S. R., Hildebrand, G. M., Osborne, J. F., Treagust, D. F., & Wallace, C. S. (2004). New directions in language and science education research. *Reading Research Quarterly*, 347-352.

Yore, L. D., Pimm, D., & Tuan, H.-L. (2007). The literacy component of mathematical and scientific literacy. *International Journal of Science and Mathematics Education, 5*, 559-589.

Yore, L. D., & Tang, K.-S. (2022). Foundations, Insights, and Future Considerations of Reading in Science and Mathematics Education. *International Journal of Science and Mathematics Education*. doi:10.1007/s10763-022-10321-2

Zeidler, D. L. (2014). Socioscientific issues as a curriculum emphasis: Theory, research, and practice. In *Handbook of research on science education, volume II* (pp. 711-740): Routledge.

Zeidler, D. L., & Nichols, B. H. (2009). Socioscientific issues: Theory and practice. *Journal of elementary science education, 21*, 49-58.

Zhai, X. (2022). Assessing high-school students' modeling performance on Newtonian mechanics. *Journal of Research in Science Teaching, 59*(8), 1313-1353.

Zhang, J., & Browne, W. J. (2022). Exploring Chinese high school students' performance and perceptions of scientific argumentation by understanding it as a three-component progression of competencies. *Journal of Research in Science Teaching*.

Zwick, R., Thayer, D. T., & Lewis, C. (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement*, *36*(1), 1-28.