

Development of a questionnaire on teachers' knowledge of language as an epistemic tool

Gavin W. Fulmer¹  | Jihyun Hwang²  | Chenchen Ding¹ |
Brian Hand¹  | Jee K. Suh³  | William Hansen¹

¹Department of Teaching & Learning,
University of Iowa, Iowa City, Iowa

²Department of Mathematics Education,
Kangwon National University,
Chuncheon, South Korea

³Department of Curriculum and
Instruction, University of Alabama,
Tuscaloosa, Alabama

Correspondence

Gavin W. Fulmer, Department of
Teaching & Learning, University of Iowa,
Iowa City, IA.
Email: gavin-fulmer@uiowa.edu

Funding information

U.S. National Science Foundation, Grant/
Award Number: 1812576

Abstract

We report on the development of a new instrument for measuring teachers' knowledge of language as an epistemic tool in science classes. Language is essential for science learning, as all learning requires the use of language to constitute one's own ideas and to engage with others' ideas. Teachers with knowledge of language as an epistemic tool can recognize the ways that language allows students to generate and validate knowledge for themselves, rather than to replicate canonical knowledge transmitted by other sources. We used a construct-driven development approach with iterations of domain analysis, item revision, teacher feedback, expert review, and item piloting to address the content, substance, and structure aspects of validity. Data from 158 preservice and in-service teachers on 27 preliminary items were collected. Findings from Rasch measurement modeling indicate a single dimension fits the items well and can distinguish teachers of higher and lower knowledge. We revised and selected 15 items for an updated instrument. This contributes to ongoing measurement projects and provides a potential instrument for future, broader use by the field to gauge teachers' knowledge of language as an epistemic tool.

KEYWORDS

epistemic tools, instrument development, language, Rasch models, teacher knowledge

1 | INTRODUCTION

A key theme emerging from the Next Generation Science Standards (NGSS; NGSS Lead States, 2013) and from subsequent work on their interpretation and enactment is the importance of classroom environments where students do not merely receive scientific knowledge but actively generate and validate scientific knowledge by engaging with phenomena and using the practices of science to explore disciplinary ideas (Duschl & Bybee, 2014; Krajcik et al., 1998; Pruitt, 2014). This has underscored the trend in science education to support learning environments focused on knowledge generation rather than knowledge replication (Settlage & Southerland, 2019). Classrooms in which students are engaged in generating and validating knowledge can be characterized as epistemically rich. Teachers and students in these classrooms use a variety of epistemic tools to help them to generate ideas as knowledge statements, explore the meanings of knowledge statements made by others or drawn from source documents, and to test the validity of their own and others' knowledge statements.

(Boon, 2012; Boon & Van Baalen, 2019), whether as a physical object, a symbolic representation, or use of spoken or written language (Kelly & Cunningham, 2019). The “generation” of this knowledge includes the private knowledge of each student as well as public knowledge shared among the class, and “validation” extends beyond replicating to critiquing and judging any knowledge statement to explore its validity.

Language is essential for science learning on multiple fronts. As the NRC's (2012) *Framework* document stated, “every science or engineering lesson is in part a language lesson... [because] students should be able to interpret meaning from text, to produce text in which written language and diagrams are used to express scientific ideas, and to engage in extended discussion about those ideas” (p. 76). In this passage, it is clear how language has its important social role (Vygotsky, 1978) as it serves students for communicating their own and receiving others' ideas (Duschl, 2005). Language is also needed to access scientific concepts, because experiencing and thinking about scientific phenomena requires engaging with the language of science (NRC, 2000, p. 122; Yore & Treagust, 2006; Yore & Hand, 2010)—to develop students' scientific literacy in both a fundamental sense of using language to read and write about science, and in a derived sense of being knowledgeable and educated about science itself (Norris & Phillips, 2003).

Thus, learners need to have language to allow them to connect new experiences with natural phenomena to their existing knowledge (Wang, Wang, Tai, & Chen, 2010). This occurs whenever a person engages with language in any of its forms: spoken, heard, written, read, or drawn (Chen, Park, & Hand, 2016; McDermott & Hand, 2016; Prain & Hand, 2016a). Thus, language is of paramount importance for students in learning about science concepts and the scientific enterprise: scientific language requires multiple, inter-related modes and purposes (NRC, 2012, p. 75); students have to develop and use language to express their ideas about concepts as they work across everyday and scientific terminologies

(NRC, 2012, pp. 96–100). Taken together, language thus serves multiple and inextricable roles in science learning.

For students to experience classrooms that are epistemically rich and engage with language in generative ways would require their teachers to know about, and to encourage the use of, language to support this (Bae, Cikmaz, & Hand, 2019; Yore & Treagust, 2006).

Despite the importance of language as an epistemic tool and the need for teachers to know about and support students' use of language to generate and validate science knowledge, we are still severely limited in our capacity to gauge teachers' knowledge of language as an epistemic tool. We posit that teachers who understand language as an epistemic tool will be more likely to provide students with opportunities to use language in generative ways as they engage in scientific practices and learn scientific concepts. But to test such an assertion requires an appropriate measure of teachers' knowledge of language as an epistemic tool—no such measure exists currently. Lacking a validated questionnaire on teachers' knowledge of language as an epistemic tool prevents the field from studying how teachers may differ or can grow in understanding this essential feature of science learning. Creating and validating a measure of teachers' knowledge of language as an epistemic tool contributes to the growing body of research on how teachers are preparing for epistemically rich classroom environments that match the vision of the NGSS (NGSS Lead States, 2013) and to the body of research on the fundamental sense of literacy put forward by Norris and Phillips (2003).

The aim of this study is to develop an instrument to measure science teachers' understanding of language as an epistemic tool.

In the following sections we introduce the theoretical basis for our design process—with iterations of literature review, item writing, expert review, data collection, analysis, and item revision—and then expand on the results of that process.

2 | RESEARCH CONTEXT FOR THIS STUDY

The research presented in this paper is an initial step of a larger project (*Moving beyond pedagogy: Developing elementary teachers' adaptive expertise in using the epistemic complexity of science*, funded by National Science Foundation grant number DRL-1812576) to examine potential growth in teachers' knowledge of epistemic tools as they take part in an extended professional development program on the Science Writing Heuristic (SWH; Hand & Keys, 1999) approach. The larger project builds on Desimone's (2009) conceptual framework for studying effects of professional development—on teachers' knowledge, then to classroom practices, and then to student outcomes (Desimone, 2009, p. 185). The SWH approach is an immersive argument-based inquiry approach (Cavagnetto, 2010) that emphasizes language as the core of learning science (Bae et al., 2019). It involves students in posing questions, gathering data, making claims

with evidence, and debating their assertions (Chen et al., 2016). Students have opportunities to negotiate their understanding of big ideas using oral and written language, which promote engaging in epistemic practices for both science and language (Chen, Hand, & McDowell, 2013). The SWH approach has been demonstrated to yield substantial improvements in students' understanding of science through laboratory work (Hand, Wallace, & Yang, 2004), growth in students' general critical thinking (Hand, Shelley, Laugerman, Fostvedt, & Therrien, 2018), and in preservice teachers' use of language and argument in their science writing (Yaman, 2017, 2018).

The instrument development process reported here was conducted prior to the first wave of teacher professional development. The immediate goal was to generate a questionnaire suitable for use with other questionnaires, interviews, and classroom observations to study the teachers' growth in knowledge of language and their implementation of immersive argument-based inquiry approaches in their classrooms. Using the instrument outlined in this paper, the larger project aims to report in future on how teachers grow in their understanding of language as an epistemic tool over time.

3 |

They did not look at the number of categories in the Likert scale, etc.....

Our study is based on unified validity theory, in which the validity of a proposed instrument is evaluated based on empirical and theoretical support for interpretations based on that instrument. This theory posits that many educational and psychological measures of import are

Unified validity theory has been the basis of the *Joint Standards for Educational and Psychological Testing* since 1985 (APA/AERA/NCME, 2014; Plake & Wise, 2014). Under this theory, decisions about validity are based on an extended analysis of the intended measures and a validity argument based on the evidence available for their development, application, and interpretation (Kane, 2001). No single "validity coefficient" could provide suitable evidence for the multiple aspects of validity (Messick, 1989a), so

As the instrument is in the early stages of development, we did not investigate aspects that are relatively *external* to the instrument (i.e., external criterion validity, consequential validity, or generalizability; Lissitz & Samuels, 2007), but focus on the relatively more *internal* aspects of validity to establish our validity argument: Content, Substance, and Structure.

Content refers to the contents of the items or tasks themselves. It is imperative to determine that the content of the task matches the proposed domain appropriately. This requires the instrument to adequately represent the construct, be clear on the boundaries for the construct, demonstrate that the tasks are representative and functionally important to the domain. We address the content validity by performing a domain analysis (Mislevy & Haertel, 2006), a process to gather substantive information about the construct or activity, reinforced by review from knowledgeable experts (Tran, Griffin, & Nguyen, 2010).

Substance refers to the rationale for the cognitive or affective processes of the construct to be measured. For our purposes, this requires considering how respondents would think about the items when they read the item and need to select a response. We address the substance

validity with a combination of expert review and feedback from experienced teachers, which allow us to support how teachers would respond to the statements posed in the questionnaire.

Structure refers to the appropriateness of the scoring process and the interpretation of response patterns with respect to the domain (McNamara, 2001). For our purpose, this means examining whether the responses demonstrate the items are functioning well across the range of the construct, and that the pattern of responses is consistent. Items would be considered to function well if they elicit responses that add relevant information to the measure by being consistent but not redundant with the rest of the instrument (Boone, 2016). A consistent pattern of responses is observable when individuals' responses are relatively predictable across items, such as when an individual's estimated level of the measured trait is quite high compared to an item that is easy to endorse versus low compared to an item that is difficult to endorse (Liu, 2010). For example, Sondergeld and Johnson (2014) created a questionnaire on teachers' and community members' awareness and community support for STEM, and calibrated it using a Rasch measurement model. A respondent with a moderate level of STEM awareness would be much more likely to endorse the statement, "Students with postsecondary education are more likely to secure a career in a STEM field," which has a low estimated item difficulty; but the same respondent is much less likely to endorse the statement, "The state standardized tests used in this region's K-12 schools adequately assess STEM knowledge and skills," which had a high estimated item difficulty. So, it is necessary to examine the pattern of responses and judge if it is sensible based on what is known about the underlying construct. We address the structure validity by conducting a Rasch measurement modeling analysis to test whether the responses are consistent with this model of measurement.

4 | INSTRUMENT DEVELOPMENT METHODS

We followed a construct-driven design process informed by our theoretical framework on validity theory. Figure 1 provides a flowchart of the iterative steps. Below we summarize the methodology for each of these steps.

4.1 | Domain analysis and review of construct literature

Domain Analysis "concerns gathering substantive information about the domain to be assessed" (Mislevy & Haertel, 2006, p. 7), which can include gathering information on the appropriate concepts, terminology, and forms of representation that are appropriate for the content—and how relatively high or low levels of knowledge may be distinguished from each other. Domain analysis typically involves reviewing standards documents (e.g., in disciplinary or educational content reviews), reviewing relevant research literature, and soliciting input from domain experts.

For this stage, we began with a literature review to parse out how language serves as an epistemic tool in science classrooms, and to develop a definition of teachers' knowledge of language as an epistemic tool in the broad sense. Our search and review focused on the combination of "teachers' knowledge" and "classroom use of language," and many variants of these terms. To ensure the review was thorough, we searched in major research databases (e.g., EBSCO Education Source), skimmed titles of the major journals in our field (e.g., *Journal of Research in Science Teaching*, *Science Education*), and invited recommendations from other

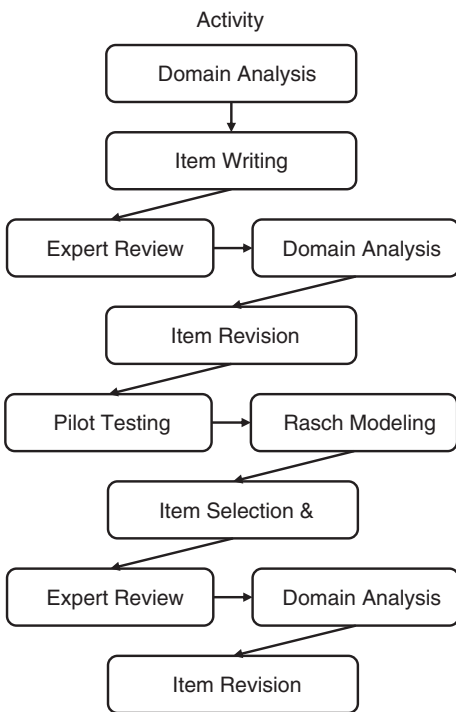


FIGURE 1 Flowchart of our construct-driven design process

scholars. We furthermore incorporated a targeted search of literature on the roles of language in student learning, so that we could draw on these ideas when modeling the range of teachers' knowledge of language. Using this corpus of literature, we focused in depth on the descriptions of what teachers know, or should know, about language use in the classroom to support learning (i.e., to serve as an epistemic tool), and from this to describe what teachers with good understanding about language as an epistemic tool would be able to do with this knowledge. That included identifying potential sub-domains from the literature. In our summary, below, we present the domain definition and sub-domains thematically.

In addition to evidence from literature review, we interviewed four domain experts affiliated with the project. The goal was to elicit and detail their experiences of supporting teachers and observing epistemically rich classrooms—where teachers support students in generating ideas and testing the validity of their own and others' ideas. The purpose of these interviews was to flesh out the construct definition developed from the literature and to help us characterize the ways that differences in teachers' knowledge of language as an epistemic tool would be observable. Our interviews focused on the salient features of epistemically rich classrooms and what positions or views they see as differentiating teachers with relatively higher or lower knowledge of language as an epistemic tool.

In the following sections, we present the construct definition on teachers' knowledge of language as an epistemic tool, review the four sub-domains identified from our review, and then characterize how teachers with relatively low versus high knowledge may differ based on the review and interviews with our experts. In the following sections for each sub-domain, we first provide a review of the sub-domain with respect to its role in learning, and then present how teachers' knowledge of the sub-domain can be operationalized as observable indicators—which then informs the characteristic statements that were the basis for item creation.

Construct definition and sub-domains. Our literature review culminated in writing an overarching definition of teachers' knowledge of language as an epistemic tool and in specifying four sub-domains. Our constructed definition of teachers' knowledge of language as an epistemic tool is:

Based on our review, we assert that a teacher with knowledge about language as an epistemic tool will understand that language is not only a tool for communication in its fundamental sense (Norris & Phillips, 2003) to share knowledge, publicly, but also is needed for knowledge generation in one's own mind, privately (Pinker, 2010). Knowledgeable teachers will also recognize that language is more than just traditional textbooks and use of scientific terminology, as it involves everyday (home) language and multiple modes and forms of representation as students develop their understanding of concepts and draw on multiple language forms and modes to express it (Lemke, 1990; Postman & Weingartner, 1969; Warren, Ballenger, Ogonowski, Rosebery, & Hudicourt-Barnes, 2001).

We unpack this construct definition into four sub-domains: (a) language is essential; (b) language is constitutive; (c) language involves process and product; and (d) language includes multimodal representations. These sub-domains are interrelated but distinguishable, so our review seeks to define each separately while acknowledging relevant overlaps among the sub-domains. Figure 2 presents how we conceptualized these four sub-domains as constituents of the construct of teachers' knowledge of language as an epistemic tool. For the present paper we focus on the knowledge of language and position this work to inform subsequent research of its connection to other knowledge areas and to classroom practices (Desimone, 2009) within the context of the larger study.

Language is essential. Our review demonstrates that language is essential for science as a shared body of knowledge, and we begin with this sub-domain because it is widely discussed in the field: "Every science lesson is a language lesson" (Wellington & Osborne, 2001, p. 2). That refers in part to the special terminology and forms of science language (e.g., Halliday & Martin, 2003), but can be taken further to note how language is needed (a) to communicate

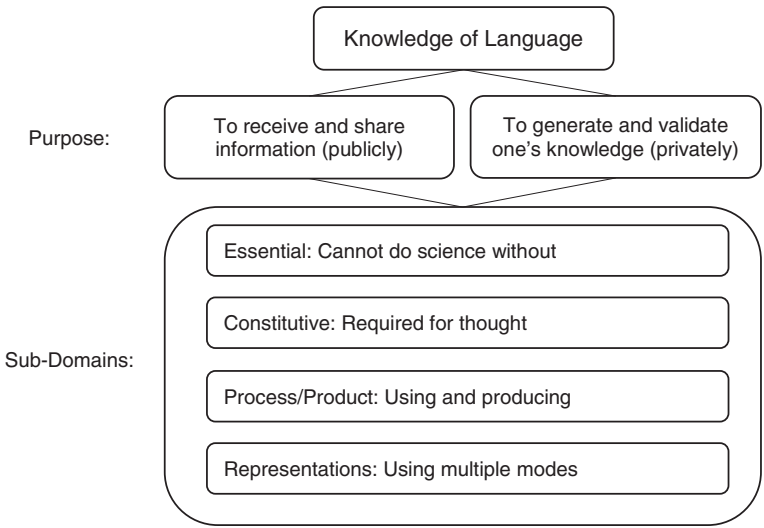


FIGURE 2 Summary of language construct sub-domains

scientific knowledge, (b) to generate new scientific knowledge, and (c) for scientific knowledge to exist at all.

Language's essential role for communicating scientific knowledge has been regularly acknowledged (Avraamidou & Osborne, 2009; Duschl, Schweingruber, & Shouse, 2007). Language is needed to state and communicate any specific scientific knowledge using specialized terminology or modes (e.g., text, graphs, and formulas), particularly when used in the special ways that scientists use it with respect to terminology and grammar (Halliday & Martin, 2003, p. 3). Language is also required to generate new scientific knowledge, and scientists use language not only to describe nature but also to generate concepts, theories, and relationships about nature (Halliday & Martin, 2003). Scientists as well as students need to use language procedurally to record data and report findings, but also conceptually to negotiate meaning, to build arguments, to imagine and compare analogies, or to hypothesize and challenge claims (Carlsen, 2013).

These views are sometimes summarized by the notion that knowledge is language (Carlsen, 2013; Postman & Weingartner, 1969), because language is always required to create, record, or transmit knowledge (NRC, 2012; Wellington & Osborne, 2001).

Drawing on the ideas of language in communicating, creating, and being knowledge, we shift to teachers' knowledge of language's essential role in classrooms. Language is employed as an epistemic tool when teachers and students are using any combination of everyday and scientific language as they investigate, negotiate, and theorize about science. Here we emphasize two aspects of this: the different roles of language for communicating and creating knowledge, and the importance of using everyday language as well as scientific terminology.

On the first aspect, no educator would agree to teach science without any language—spoken word, written text, symbols, graphs, pictures, etc. Yet teachers may unintentionally think of language as being associated with particular subject areas, such as English Language Arts or Foreign Languages, and miss opportunities to incorporate explicit language instruction in their science classes (Graham, Capizzi, Harris, Hebert, & Morphy, 2014; Kiuahara, Graham, & Hawken, 2009). In this sense, language should not be viewed as a standalone topic for study but as integrated with science. Therefore, it is imperative for teachers to understand that students cannot learn anything without language to communicate and express their ideas (Duschl et al., 2007; Duschl & Osborne, 2002) and to generate shared knowledge (Prain & Hand, 2016a). Knowledgeable teachers would recognize that learning science requires using language, such as explicitly incorporating writing and speaking as well as reading or listening. Writing in science effectively supports students to learn science meaningfully (Glynn & Muth, 1994; Hanrahan, 1999); it provides a means to communicate what students have learned to peers or parents (Keys, Hand, Prain, & Collins, 1999), and as a means to clarify and deepen existing knowledge (Hand et al., 2004). Speaking enables students to be explicit about their ideas during discussion with peers, which combined with writing can support students' understanding of science concepts and to generate their explanations (Rivard & Straw, 2000; Syh-Jong, 2007).

On the second aspect, teachers would know that "scientific language" refers not only to specific terminology or usage that is special to science (Halliday & Martin, 2003; Wellington & Osborne, 2001) but also includes incorporating language forms and usage from students' everyday and familiar language into classroom environments (Lynch, 2001; Warren et al., 2001). Using familiar language associated with everyday experience can be easier for students to access (Hand et al., 2003; Wallace, 2004), and provides easier connections between the discussion of

science phenomena with how they already communicate their ideas about nature (Arons, 1983; Lemmi et al., 2019; Wiser & Amin, 2001). In so doing may reduce the sense of barrier in using scientific language (e.g., Brown, 2006) when they can talk through their ideas about natural phenomena in everyday language with peers (Warren et al., 2001). Furthermore, findings show that focusing on conceptual discussions using everyday language before focusing on specific scientific terminology provides opportunities for students to improve their understanding of science (Brown & Ryoo, 2008; Brown & Spang, 2008; Schoerning et al., 2015).

Language is constitutive. Our review demonstrates how language is necessary for science as the basis for individuals' cognition. The notion that language is constitutive addresses how the learner uses language as the "stuff of thought" (Pinker, 2010) when they construct their own understandings.

This is not to claim that other, non-linguistic animals are incapable of thought (cf. Corballis, 2017), but that as children gain language they begin to use it to support their interpretations and inferences about the natural world (Gelman & Coley, 1990), and that using language provides symbolic and representational capabilities that both support human cognition (Carruthers, 2002) and influence cognitive preferences and patterns (Haun, Rapold, Janzen, & Levinson, 2011). Language thus serves as what students are using "in their own heads" to generate such knowledge; knowledge does not exist in the students' heads without any language. This view is mirrored in works from the social constructivist perspective that argue conceptual development cannot be separated from the development of language (Vygotsky, 1962; Wellington & Osborne, 2001). Vygotsky (1962) describes this knowledge acquisition in two ways, the first being the spontaneous knowledge one constructs from their environment and the second being formal knowledge constructed from language originating from a teacher or other authority figure. In the classroom, this regularly involves a negotiation among students, the teacher, and the knowledge resources available from prior experience or new investigations (Chen, Benus, & Hernandez, 2019).

because it highlights how language is also needed when the student constructs scientific ideas and engages in reasoning (Hand & Prain, 2006; Martin & Hand, 2009; Yore & Treagust, 2006). For example, writing is a knowledge constitutive process because the students continually review and recreate their understandings as they write (Prain & Hand, 2016a). To create or revise a written document requires students not only to retrieve and represent what they know, but also to synthesize their knowledge "as if for the first time" (Galbraith, 2009, p. 61) to each new set of writing goals or purposes. This is how science writing emphasizes students' generation of their own ideas of science concepts rather than replication of others' scientific understanding (Hand, Prain, & Yore, 2001; Pelger & Nilsson, 2016; Yore, Bisanz, & Hand, 2003). Similarly, discussion in the science classroom involves sharing different interpretations, persuasion or argumentation, evaluation, and negotiation with peers in the moment, which requires the extemporaneous revision and synthesis of knowledge (Rivard, 2004; Yore et al., 2003).

Drawing on the ideas of language serving as the "stuff of thought" for students, we shift to teachers' knowledge of language's constitutive role in classrooms. There is no one right way for learners to construct their understanding of science ideas—everyone has different experiences,

privately held prior knowledge, and idiosyncratic use of language (e.g., Anderson, 1992). Teachers, therefore, need to understand how using language provides opportunities for each student to express their previous knowledge and current ideas, and to use language to combine different representations to effectively and accurately convey ideas as they build new knowledge (Yore, Chinn, & Hand, 2008; Yore & Treagust, 2006). Teachers can provide openings for students to express their own ideas, in writing or speaking, and thus to restate and revisit their knowledge regularly (Cavagnetto, Hand, & Premo, 2020; Cervetti, Barber, Dorph, Pearson, & Goldschmidt, 2012).

Language involves process and product. Our review highlights how language use in classrooms engaging it both as a process and as product of knowledge generation. This underscores the use of language to create as well as represent. Language is not only a final product to use at the conclusion of a lesson or experiment to share one's ideas in a written text or a spoken statement; it also works as a process where actively speaking with others or writing out one's ideas allows those ideas to be refined and clarified (Norris & Phillips, 2003; Yore et al., 2003). Writing and revising helps students construct their understanding of science concepts rather than just to memorize scientific knowledge. For example, Wang (2020) found that students using argumentative structures helped them to express their ideas to peers to clarify their claims and related evidence, and to improve the connections between claims and evidence. Teachers can guide this process to strengthen the students' understanding of the science content and its underlying epistemic processes (Benedict-Chambers, Kademian, Davis, & Palincsar, 2017).

There are many classroom examples of using language as both process and product. At the beginning of a unit, a teacher may invite students to share their prior understanding and to pose questions about a phenomenon (Choi, Hand, & Norton-Meier, 2014; Hand, Cavagnetto, et al., 2016; Hand, Norton-Meier, Gunel, & Akkus, 2016; Villanueva & Hand, 2011). This communicates what they already know and helps to clarify its relevance to the current phenomenon, and to create a set of shared questions to guide an investigation. The questions that emerge seem like a product, but the discussion uses language as a process for students and the teacher to elicit, activate, and organize knowledge. When designing an investigation, students in groups will talk through their ideas, draw diagrams, and write up plans to help them determine appropriate procedures. These can be reviewed and revised after input, giving students further chance to reflect on their reasoning. Later, when students are summarizing their investigations and claims, the process of writing an argumentative summary or talking through their reasoning helps the student to review and reorganize what they experienced and to see connections between their investigation, their evidence, and the claims they are making. And, in situations where students' prior understanding may conflict with science ideas (Huang, 2006; Sinatra & Broughton, 2011), students have to use language to negotiate ideas among prior knowledge and new experiences until they gain a clear understanding through discussion or argument (Cavagnetto & Hand, 2012; Chin, Yang, & Tuan, 2016; Demirbag & Gunel, 2014). Of course, these language processes often result in language products—where students communicate their science ideas to others through different representations, such as graphs, diagrams, tables, and spoken discourse (Chen et al., 2013; Hand & Prain, 2006; Yore et al., 2008)—but the process itself is important.

Many of the above examples—opening questions and discussions from the teacher, draft investigation plans—are widely recognized aspects of effective formative assessment strategy (Black, 2017; Black & William, 2009). This perspective is consistent with the notion of language as process. Using such strategies can help teachers create classroom environments that welcome students' ideas, including competing ideas or potential “wrong answers” (Black, 2017, p. 296).

This sets the stage for students to serve as resources for each other and gives them the position to take ownership of their own learning (Black & Wiliam, 2009; Ko & Krist, 2019). Therefore, when teachers emphasize the processes of language, and guide students to use writing and talking to revisit their ideas, students will express their ideas more freely regardless of perceiving them as right or wrong; this may result in student taking initiative to talk and explain their understanding, which makes students more engaged in learning (Cavagnetto et al., 2020). Meanwhile, during interaction with classmates and teachers in presenting their ideas, they may find out flaws in their presentations and then have a chance to revise them.

Drawing on these ideas, we shift to teachers' knowledge of language when used as process and product. Teachers should know that language can consist of familiar language products—such as lab reports, oral presentations, and so on—through which students communicate their knowledge and reasoning. These can allow teachers to make informed decisions about individual students or the class as a whole. Teachers should also know that using language provides a process for students to refine their ideas and to construct individual understanding (Yore et al., 2004; Yore & Hand, 2010). Teachers who over-emphasize language as a product are likely to focus on replication of knowledge. Examples of replicative uses of language as a product include filling in worksheets or similar templates, and completing pre-prepared data tables or figures (Chin et al., 2016; Waldrup & Prain, 2006). Examples of generative uses of language include students drafting their own observations, interpretations, and conclusions (Demirbag & Gunel, 2014; Hand et al., 2018); creating their own competing ways of representing data (Prain & Waldrup, 2006; Waldrup, Prain, & Carolan, 2010); journaling reflectively about how their ideas changed from the beginning to the end of an investigation (Balgopal & Montplaisir, 2011); and integrating various modes of representations to communicate their ideas as they are developing them (Tytler, Prain, & Hubber, 2018).

Language includes multiple modes of representation. Our review points to the relevance of incorporating language with multiple modes of representations to address the varied ways that students can represent a science concept, and it builds on ideas raised about language being essential and constitutive for science. Science concepts and evidence come not only as text, but involve varied modes of representations beyond text to include spoken word, pictures, diagrams, graphs, equations, and tables to convey the meanings that students want to share (Lemke, 1990, 2004). In using multimodal representations, students can draw on the combination of modes to express their ideas to others (Yore et al., 2004; Yore & Hand 2010), and in doing so the incorporation of textual, graphical, and mathematical formats (Hand & Choi, 2010; Yore & Hand, 2010) can increase the depth and interconnectedness of students' knowledge (Tang, Delgado, & Moje, 2013; Van den Broek, 2010; Yaman, 2018). This ultimately improves the students' science achievement (Bradbury, 2014; Oyoo, 2012).

For classroom environments that draw on ideas in this sub-domain, teachers should know that using multiple modes of representation helps students to learn more deeply and to express different aspects of their prior knowledge and new experiences (Prain & Hand, 2016b). Students can use multiple modes of representation simultaneously or sequentially as part of sense-making (Olander, Wickman, Tytler, & Ingerman, 2018). As examples, students may use spoken words, text, or even multiple different types of figural representations to help them deepen their ideas with others, such as about thermal expansion (Seah, Clarke, & Hart, 2014), substance exchanges between organs (Olander et al., 2018), or electrical current in circuits (Prain & Waldrup, 2006). Students may have preferences for some representation modes, and the teacher can encourage students to incorporate other modes that could complement it. By using various forms of representation, students can be as creative as they want to share their understanding

with others. Much of the value comes from combining language modes, such as talking, writing, and drawing, to help the students to use various representations helps to make their reasoning clear for themselves as well as others (Tytler, Prain, Aranda, Ferguson, & Gorur, 2020).

With multiple modes of representation in mind, teachers will encourage students to express their ideas through different modes of representations or even to use multiple forms of the same mode. Teachers who are more knowledgeable about the use of multiple modes will also push students to make explicit connections between modes (Cikmaz et al., 2019), such as using text to support figural representations, referring directly to figures in the written text or spoken presentation, and so on. By making these expectations clear, students will gradually learn to explain their science ideas thoroughly using different modes (Tytler et al., 2020). As a result, they will learn science ideas more deeply and will express different aspects of those ideas.

Characteristic statements distinguishing teachers' knowledge levels. To move from the domain analysis to the item writing phase, we need to operationalize the construct of "teachers' knowledge of language as an epistemic tool" from a concept into an observable and measurable response. That required a broad view of the knowledge, characteristics, or potential observations that would demonstrate the intended measure (Mislevy & Haertel, 2006) and how these can be used to distinguish teachers of varying extent of knowledge. We combined the findings from various sources—the construct definition, the literature review on the four sub-domains, and interviews with our domain experts—using two phases described below—that culminated in a set of statements that help to characterize teachers of varying level of knowledge (summarized in Table 1), and which would inform development of self-report items that could distinguish the respondents. We intentionally incorporated some statements that explicitly mention science, such as use of science terminology or an emphasis on scientific accuracy, with others that appear more general, such as using notes to reproduce content or that associate "language"

TABLE 1 Characterizing statements on teachers' knowledge of language as an epistemic tool

Lower level knowledge	Higher level knowledge
Teachers are likely to... • See language as a distinct skill associated with subject matter of English language rather than an important tool in learning science • Think that varied language approaches are not required to learn science • Prefer that students focus their language effort on reproducing notes and material rather than using language to represent ideas in multiple ways that are mutually supporting • Emphasize "accurate" representations of scientific terms and ideas according to the teachers' own thinking (even if the teacher's view may include naïve conceptions)	Teachers are likely to... • Focus on providing opportunities for students to use language to express their own ideas (it allows students to consolidate what they already know and to generate new understanding for themselves) • Allow students to use any combination of everyday and scientific terms, depending on the stage of discussion • Use terminology to support the discussion or debate, but not as the main purpose of the discussion • Emphasize cohesiveness of representations and encourage students to use multiple modes of language (as meaning expressed by non-verbal representations like graphs and charts can never be replaced by words) • Encourage students to write out and plan for themselves, and to revisit and reflect on their previous writings, drawings, and so on

with other subject areas. This fits our operationalization of teachers' knowledge of language as an epistemic tool tying closely to science while recognizing that the extent of a teacher's knowledge is also influenced by an understanding of language more fundamentally (Norris & Phillips, 2003).

First, since the domain analysis established sub-domain definitions, we could draw some characteristic statements from the literature. For example, regarding the sub-domain *language is essential*, Norris and Phillips (2003) wrote that “[a]ny scientific theory requires for its creation and expression the use of text” (p. 231). This statement expresses a view of language being essential for science that a relatively knowledgeable teacher would agree (and could be further adapted into a draft item, LQ12R, in Table 2).

Second, additional statements were generated by members in the research team with extensive classroom observation experience with teachers in SWH projects to help characterize teachers according to relatively high and low extent of implementation. This was based on the research team's prior work (Bae et al., 2019; Cikmaz, Fulmer, Yaman, & Hand, 2019) in supporting teachers' growth in knowledge of the SWH approach (Cikmaz, Hand, Fulmer, & Yaman, 2019). Since teachers with more extensive implementation of the SWH approach are more knowledgeable about language as an epistemic tool, the research team members' experiences in observing and collaborating with teachers as they adopted and implemented SWH was insightful in helping to characterize differences in knowledge.

During this stage, the team also drew on preliminary interview findings with a new cohort of teachers who had agreed to join the project but not yet begun attending professional development sessions. When asked about the role of language in their classrooms, the new cohort teachers consistently talked about the use of vocabulary and terminology—with some variation in the extent to which they felt it important that students can use their own everyday wording for concepts or whether it is important to emphasize vocabulary. For example, one 4th grade teacher stated that, “I can give them a term and it could be something (inaudible), but if they don't understand it and don't use it in their own words then they don't ever understand it.” This demonstrates a teacher's stance that students' own word choices are important when developing understanding. By contrast, one 5th grade teacher stated, “In science I have never made them memorize it. I use it more as ‘This is what it means, this is how we use it.’ ... [But] a few times I have quizzed them on vocabulary.” This appears to indicate some tension in the desire not to overemphasize vocabulary, such as by enforcing memorization, with a desire for the students to “know” the terminology so they can use it correctly. These preliminary interviews with a small sample of project teachers helped substantiate the characteristic statements presented in Table 1.

4.2 | Item writing and revision

Based on the results of the domain analysis and the characteristic statements described above (and summarized in Table 1), we moved to write potential items that would elicit responses along a spectrum of knowledge (Mislevy & Haertel, 2006). All members in the research team held brainstorming sessions to identify potential contexts or situations and associated statements that would provide evidence that teachers know about language as an epistemic tool. Informed by our construct-driven approach and our prior experience with Rasch measurement

TABLE 2 Original statements for the instrument

ID	Statement
LQ01	Students cannot learn without language
LQ02	Language is important in science class
LQ03	We can't engage in scientific practice in science without language
LQ04	Students write and talk to show they learned science
LQ05	Students are finding out about science by listening, reading, and writing about it
LQ06	Students need to use specific scientific terms accurately
LQ07	Students should be able to recreate their own ideas from what we have discussed in class
LQ08	Knowledge can be represented in different ways
LQ09	Students can use different kinds of representations, for example, drawing and texts, to share their ideas
LQ10 ^a	When the classroom is quiet, I know the students are paying careful attention
LQ11	Students have to talk about and write their ideas to learn science
LQ12	Students are generating new knowledge by talking, writing, drawing, and reading
LQ13	In science class, students can use everyday language and scientific language to explain ideas
LQ14	Students can use everyday language to communicate accurate scientific ideas
LQ15	Students can use everyday language to improve understanding of scientific ideas
LQ16	If students aren't talking, then they may not be learning
LQ17	Students who build connections among their drawings and text are demonstrating scientific ideas well
LQ18	Language is not only used to copy knowledge from the teacher or a textbook, but is also used to generate knowledge
LQ19	When using different representations and text to develop ideas, students need to make sure there is a cohesive link between them.
LQ20 ^a	Language use includes filling in worksheets or similar templates rather than creating their own competing ways of representing data.
LQ21	Without employing language as a learning tool, learning of scientific concepts would lack understanding.
LQ22	As students grow (or students get older), science class focuses more on acquisition of scientific terms rather than expressing scientific terms using easy/everyday language.
LQ23	Generation of scientific knowledge is achieved by employing language as a learning tool.
LQ24	Writing to different audiences is necessary to develop conceptual understanding.
LQ25 ^a	Sharing ideas through writing is more important than talking in science learning.
LQ26 ^a	Reading comprehension is not necessarily related to learning science.
LQ27	Scientific knowledge growth entails language development.

^aReverse items.

modeling, we sought to create items that could distinguish teachers of varying knowledge. To do so, we attempted to write some items that could be endorsed by teachers with both relatively high and relatively low knowledge of language as an epistemic tool, as well as other items that would only be endorsed by teachers with relatively higher knowledge of language as an

epistemic tool. After creating preliminary statements, expert teachers who had previously participated in our SWH teacher workshops reviewed and advised about all the draft statements. This iteration of expert review focused on removing any ambiguity or confusion in the wording of the statements. After this initial item writing and feedback, we had 27 statements as seen in Table 2.

4.3 | Expert review

The statements in Table 2 were revised by the authors with input from researchers with expertise in the role of language in science classrooms ($N = 3$), professional developers and teacher educators ($N = 4$), and expert teachers with experience in application of the SWH approaches ($N = 3$). The researchers and teacher educators were provided a summary of the domain analysis findings and the draft questionnaires, and asked to talk through their thinking about how teachers with whom they have worked would interpret each question: is it understandable, does it represent use of language domain concepts appropriately, and so on. For expert teachers, we solicited written feedback on the draft questionnaires and conducted phone interviews with them about their interpretation of the questions.

These steps, together, helped to support the substance validity of this instrument, again drawing on the notions of construct-driven measurement and Rasch measurement modeling. After the conclusion of this process, all items were reviewed, and adjustments made for piloting. The items were all presented with a 5-point Likert-type response scale (1 for “strongly disagree” and 5 for “strongly agree”).

4.4 | Pilot testing & analyses

Data for pilot testing and analyses were collected from 158 teachers from across Iowa, including in-service teachers from multiple districts and pre-service elementary and secondary teachers from the University of Iowa. This selection process allowed for recruiting respondents who had a wide range of prior experience with the SWH approach—from no exposure to moderate exposure.

To that end, the 39 in-service teachers (97% identifying as female; average of 13.5 years of teaching experience) came from districts and schools where there had been partial coverage of a previous SWH project. Therefore, some of these in-service teachers, but not all of them, had prior experience with the SWH approach and its associated pedagogical principles. The 119 pre-service teachers (85% identifying as female) were in various stages of either an elementary teacher education or a secondary science teacher education program, so they also had varied amounts of exposure to SWH and associated principles in their coursework and practicum experiences. For the elementary pre-service program, there is

no explicit attention to SWH in science methods courses, so this is considered little to no prior experience with SWH. For the secondary science pre-service program, SWH is widely discussed but is not a required instructional approach, so this could be considered moderate prior experience with SWH.

Our analyses of the pilot data seek to evaluate the structural aspect of validity, by determining whether the 5-point scaling structure works well and whether the pattern of responses to the items matches the intended item difficulty.

Thus, we applied a partial-credit Rasch model to allow each item a unique difficulty scale using the partial-credit model (PCM; Masters, 1982). While lacking sufficient responses in some categories on some items could lead to uncertain item difficulty estimates, our results shown below indicate the PCM is suitable. The probability of a specific respondent, teacher n , selecting the rating level k on item i is denoted as P_{nki} , and is estimated based on the teacher's latent trait, B_n , and the difficulty of that item-step combination, D_{ki} . This is represented in the following equation:

$$\ln\left(\frac{P_{nki}}{1 - P_{nki}}\right) = B_n - D_{ki}$$

This statistical model does not include other parameters for guessing or attenuated item discrimination, which are sometimes applied in other item-response theory applications. As such, the Rasch measurement model upholds a strict interpretation of *good measurement* based on how test-takers respond to items. For our analyses, we estimated a PCM with ConQuest parameterization using the TAM package (Robitzsch et al., 2019) for the R statistical environment (Ihaka & Gentleman, 1996). All items were recoded to accommodate the coding guidelines for the TAM package (retaining 5 points but recoded from 1–5 to 0–4). The latent trait estimates (B_n) of teachers are scaled so that the average estimate of all teachers is zero.

Since 27 items may be too burdensome in practical settings, we sought to identify the most informative items to retain or revise for future use. Item selection under the Rasch measurement approach focuses on items that sample from the domain appropriately, that cover a broad range of item difficulty, and that function well according to the strict definition of “good measurement” consistent with the Rasch measurement model. In it, an item is considered to be informative for measurement when the responses are not redundant with other items, nor too unpredictable to make the estimates unreliable. This decision is based on the mean-square fit statistic for each item, which indicates the extent to which a response on the responses on that item are predictable based on the other items. A mean-square value near 1 is considered “productive for measurement” in that the responses to this item have the expected balance of randomness and predictability; the window for acceptable mean-square fit statistics is set between 0.6 and 1.4 when developing rating scales (Boone, Staver, & Yale, 2014; Wright, 1996). There

are two mean-square fit statistics to review: infit is “inlier-sensitive” or information-weighted, which prioritizes respondents near the center of the distribution; outfit is outlier-sensitive and prioritizes respondents near the upper and lower ends of the distribution (Bond & Fox, 2007; Boone et al., 2014). Items with mean-squares below 0.6 are too predictable and may inflate reliability, whereas items with mean-squares above 1.4 are too unpredictable and could degrade measurement. We could also examine *t*-values for infit and outfit, but these are quite sensitive to sample size so may be misleading with samples of 150 or greater, so only mean-square fit statistics are reported below.

5 | INSTRUMENT DEVELOPMENT RESULTS

Our analyses of the full 27 initial items showed there was fairly good fit for most items to the Rasch measurement model. Table 3 shows the item and item-step fit statistics, and their difficulty estimate. To confirm whether the items are part of a single dimension that we can measure for teachers' knowledge of language as an epistemic tool we performed a principal components analysis on the residuals (i.e., any part of a response that is not estimated by the Rasch model) to test if an additional factor may be present in the data even after modeling the one dimension that we conjectured. Results indicated that Rasch modeled measures explained 54.6% of the variation using a single dimension. After accounting for this dimension, the first additional factor in the residuals had an eigenvalue of 3.6, roughly equivalent to about three questions compared to the original questionnaire of 27 items. This is not considered practically meaningful (Linacre, 2003) and lies within a range that could occur by chance with a test and sample as used for the present study (Chou & Wang, 2010), so we can conclude that the questionnaire was sufficiently unidimensional. Additionally, the full set of 27 initial items showed a separation index of 2.21, with concomitant expected a priori (EAP) reliability coefficient of 0.83, which indicates we are capable of distinguishing teachers of higher versus lower ability, and therefore suitable for a new rating-scale instrument for research purposes.

As seen in Table 3, overall difficulty levels are low with the average of -1.51 and the range from -5.04 to 0.27 . Here, negative values indicate that an item is easy to agree with or endorse. So, respondents in our sample found many of the items relatively easy to endorse. The test information curve (Figure 3) also shows that we have maximum information about respondents whose estimated ability values are between -2 and 0 logits. Since it is necessary for the instrument to cover a wide range of difficulty levels with enough information, we therefore wanted to select items across the whole range of difficulty.

The Wright Map shown in Figure 4 guided our selection of items across this range. It is arranged so that the vertical axis shows the intensity of the teachers' latent trait (on the left side of the figure) alongside the thresholds between response categories for each item (on the right side of the figure). Higher values for respondents meaning a higher level of the trait; higher values for each item indicating it is more “difficult” or harder to endorse. The pattern of the items demonstrates that there are thresholds for each item's steps that progress along the full range of the teacher responses. For each item, its lower thresholds (e.g., Cat1, the step from choosing response option 0 to choosing response option 1) are relatively “easier” to agree with than the higher threshold (e.g., Cat4, the step from choosing response option 3 to choosing response option 4), which is consistent with the expected structure of the trait and with typical Rasch measurement modeling.

TABLE 3 Item difficulty, thresholds, and item fit

	Threshold location	Threshold location	Threshold location	Threshold location	Item difficulty		
ID	τ_1	τ_2	τ_3	τ_4	δ_i	Outfit	Infit
LQ01	-2.10	-0.69	-0.47	0.28	-0.74	0.93	0.96
LQ02	-2.10	-2.05	-1.55	-0.09	-1.39	0.86	0.84
LQ03	-2.80	-1.08	-0.57	0.64	-0.96	0.80	0.82
LQ04	-3.20	-1.27	-0.86	0.77	-1.14	1.02	0.92
LQ05	-3.60	-0.95	-0.67	0.55	-1.16	1.21	1.04
LQ06	-2.70	-1.24	-0.58	1.48	-0.77	0.99	0.97
LQ07	-4.10	-2.70	-2.12	0.05	-2.21	0.86	0.88
LQ08	-4.30	-4.00	-3.13	-1.97	-3.33	0.69	0.92
LQ09	-6.40	-6.17	-6.07	-1.44	-5.04	0.77	0.90
LQ10	-2.70	-0.93	-0.24	1.22	-0.66	1.51*	1.39*
LQ11	-2.90	-0.68	-0.38	1.10	-0.7	0.88	0.88
LQ12	-4.90	-4.59	-2.28	-0.42	-3.03	0.75*	0.82
LQ13	-4.40	-4.10	-3.13	-0.26	-2.95	0.86	0.86
LQ14	-5.10	-2.31	-1.59	0.41	-2.15	1.00	0.95
LQ15	-4.10	-2.58	-2.29	0.05	-2.23	0.85	0.85
LQ16	-1.40	-0.06	0.52	2.03	0.27	1.27*	1.24*
LQ17	-5.00	-2.16	-1.72	0.09	-2.18	0.84	0.86
LQ18	-2.60	-1.79	-1.15	0.79	-1.17	0.95	0.89
LQ19	-6.10	-1.99	-1.05	0.89	-2.06	0.88	0.83
LQ20	-1.70	-0.60	-0.18	1.27	-0.31	1.65*	1.41*
LQ21	-3.90	-0.87	-0.04	1.48	-0.83	0.83	0.82*
LQ22	-2.20	-1.12	-0.29	1.53	-0.53	1.18	1.11
LQ23	-6.10	-1.97	-1.12	1.14	-2.02	0.77*	0.76*
LQ24	-2.40	-1.52	-0.72	0.80	-0.95	0.94	0.94
LQ25	-1.50	-1.02	-0.20	1.25	-0.35	1.76*	1.43*
LQ26	-1.80	-0.58	-0.14	0.97	-0.38	1.30*	1.06
LQ27	-4.60	-3.40	-1.40	1.24	-2.05	0.81	0.81

* $p < .05$.

5.1 | Item selection and revision

To select items to retain for a shorter instrument, we examined the infit and outfit mean-square fit statistics, as well as the items' difficulty estimates and the Wright map. This allowed us to select items that fit the measurement model well and that ranged across relatively easy through moderate difficulty. Furthermore, because we anticipate using these items with teachers who are taking part in ongoing professional development on the SWH approach, we needed the capacity to distinguish teachers whose knowledge of language as an epistemic tool may

FIGURE 3 Test information curve with the 27 initial statements

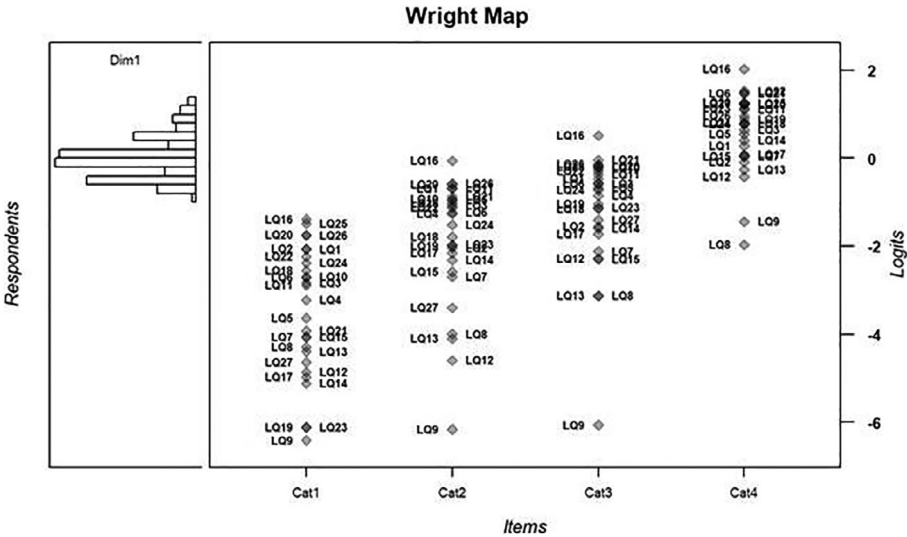
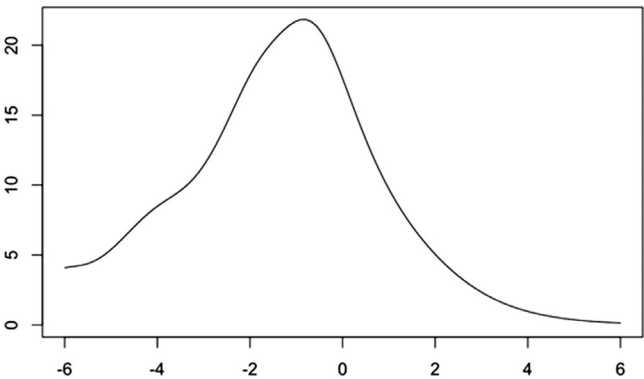


FIGURE 4 Wright map for pilot study data with 27 initial items

subsequently increase over time. For this reason, we also intentionally selected items that were of higher difficulty to include in the retained items for future use. After this step, we also reread all the items for content to ensure that all aspects of the sub-domain are included. Based on these findings, we selected 14 items that covered a range of difficulty and that had excellent fit to the Rasch model. [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] These 15 items are now prepared for wider field testing in our summer workshops and school-year professional development work.

5.2 | Interpretations of instrument response patterns

We examined the pattern of responses on the questionnaire to inform interpretations based on the instrument. We began by creating a Rasch estimated score for language as an epistemic tool

TABLE 4 Selected and revised items for the instrument

ID	Revised statement	Sub-domain
LQ01R	Students cannot think through scientific ideas without language	LE
LQ03R	Students cannot communicate scientific ideas without language	LE
LQ05	Students find out about science by hearing, reading, and writing about it.	LC
LQ06	Students need to use specific scientific terms accurately	PP
LQ07R	Students should be able to communicate their own ideas about what we have discussed in class	LE
LQ11	Students have to talk about and write their ideas to learn science	LC
LQ12R	Producing language—Writing, drawing, talking—Is how students learn scientific knowledge	LC
LQ16R	If students are not talking through ideas, then they may not be learning	LC
LQ17R	Encouraging links across modes of representation (such as written text, drawings, graphs, and classroom discussion) helps students in the process of learning.	MM
LQ18	Language is not only used to copy knowledge from the teacher or a textbook, but is also used to generate knowledge	PP
LQ20R ^a	Filling in worksheets or templates from the curriculum is the most important use of language in science class	LE
LQ22R ^a	As students get older, science class should not encourage students to use everyday language for science concepts	LE
LQ24R	Writing to different audiences helps students to deepen conceptual understanding.	LE
LQ26 ^a	Reading comprehension is not necessarily related to learning science	LE
LQ28 ^a	Using multiple modes of representation would be confusing for students when we are learning science	MM

Abbreviations: LC, language is constitutive; LE, language is essential; MM, language involves multiple modes of representations; PP, language involves process and product. "R" indicates revised items for this version. "a" indicates reverse-coded items.

using the retained 15 items (Table 5). Figure 5 shows a histogram of these scores. Based on visual inspection of the histogram, there appeared “clusters” of respondents with relatively lower, moderate, and higher estimated scores. We created three response clusters—low, medium, and high—by setting two cut-off values, at $-.23$ and $+.23$, which are approximately one-half standard deviation below and above the mid-point of the scale. Figure 6 contains histograms that show the distribution of pre-service and in-service teachers' scores, with the cut-offs as lines forming the three clusters. Though we have a higher respondent pool of pre-service teachers, the distribution is quite similar for in-service teachers and shows that knowledge of language as an epistemic tool varied in both groups.

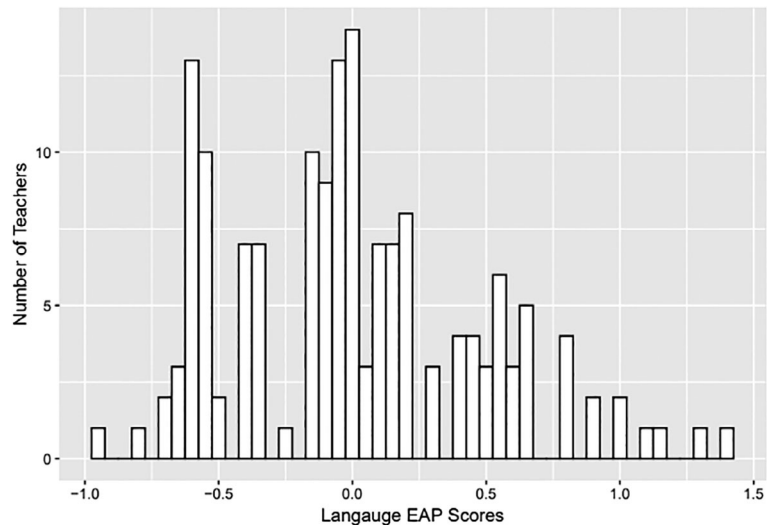
We also examined the raw responses of the respondents, according to whether they were in-service or pre-service teachers, according to the response clusters based on the Rasch estimates (Table 6). Of course, because the Rasch estimated scores are calibrated based on raw response data for 15 selected items, there is an expected relationship between the raw responses and the Rasch estimates. That said, looking to the raw values can provide some insight for interpreting scores on the Rasch estimates. As shown in Table 6, there was a systematic difference in how

TABLE 5 Rasch estimated scores for language as an epistemic tool

Group	Full sample		In-service teachers			Pre-service teachers	
	N	Language score M (SD)	N	Years teaching (M)	Language score M (SD)	N	Rasch estimate M (SD)
Total	158	.009 (.217)	39	13.5	.018 (.229)	119	.006 (.213)
Low	47	−.528 (.202)	11	11.5	−.467 (.228)	36	−.547 (.195)
Medium	71	.001 (.207)	20	12.4	.014 (.217)	51	−.003 (.202)
High	40	.652 (.252)	8	18.9	.696 (.258)	32	.641 (.250)

Note: The pre-service teachers do not have any teaching experience, so their reported years of teaching are zero.

FIGURE 5 Histogram of Rasch estimated scores for language as an epistemic tool



the respondents answered the items. [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

6 | DISCUSSION

We sought to develop an instrument about teachers' knowledge of language as an epistemic tool. This fits within the broader push toward more explicit attention to the epistemic basis of

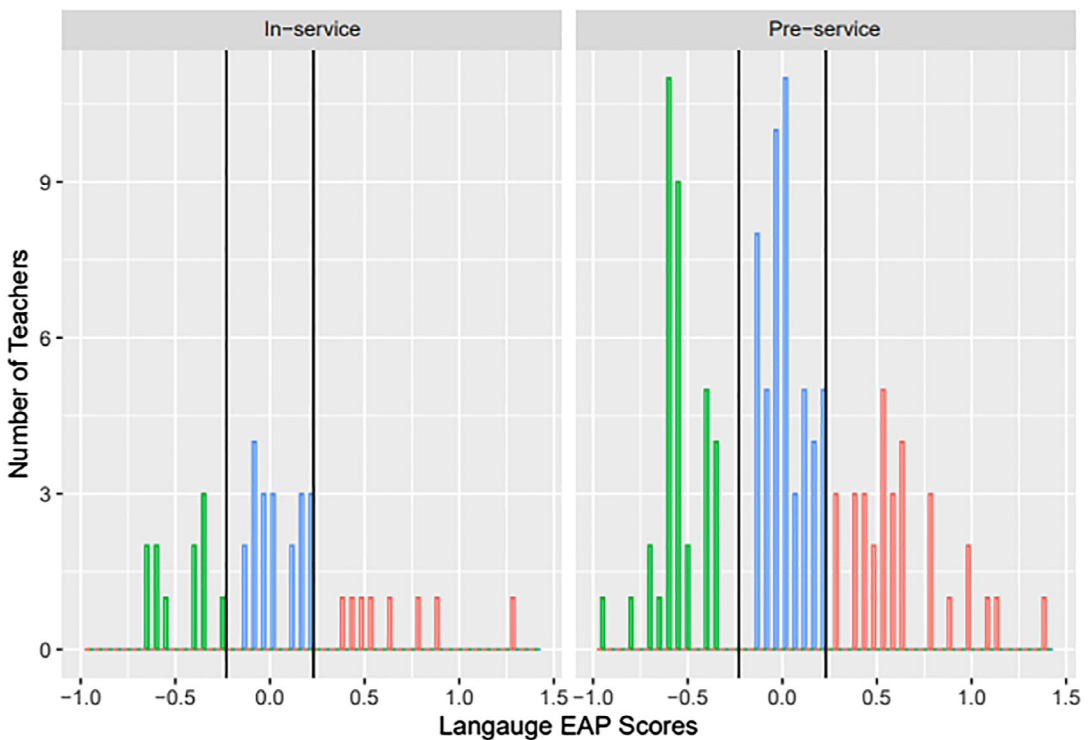


FIGURE 6 Histograms of Rasch estimated scores for language as an epistemic tool showing cut-off points [Color figure can be viewed at wileyonlinelibrary.com]

science knowledge and science learning under the NGSS era and the role of epistemic tools in classroom environments (Settlage & Southerland, 2019). An epistemically rich classroom is one in which students are engaged in generating and validating scientific knowledge, not simply replicating scientific knowledge. Language is the most essential and, perhaps, most impactful epistemic tool available for creating classroom environments focusing on knowledge generation (Chen et al., 2013; Wellington & Osborne, 2001; Yore et al., 2003).

Our domain analysis allowed us to propose a construct definition for teachers' knowledge of language as an epistemic tool: knowledge of the essential nature of language for learning science, including the roles of language for allowing students to create their own knowledge (privately) and to test and validate their own and others' knowledge claims (publicly). We also identified four interrelated sub-domains for language: language is *essential*; language is *constitutive*; language involves both *process and product*; and language include *multimodal representation*. The language sub-domains that we reviewed underscore how language serves to support science learning (Wellington & Osborne, 2001)—not only as being necessary for reading, writing, and communicating science (Duschl, 2005), but also that language is the “stuff of thought” (Pinker, 2010) that students necessarily use whenever they engage in science and engineering practices especially to read, write, draw, speak, or listen about any scientific concept. Furthermore, the process of using language and multiple modes of representation can enhance classroom environments and deepen scientific learning (McDermott & Hand, 2016).

We were able to create an initial 27 items, piloted with 158 teachers in pre-service and in-service settings. Based on Rasch measurement model results, we then selected and revised a

TABLE 6 Mean raw response to initial items for in- and pre-service teachers, by response cluster

	In-service			Pre-service		
	High	Medium	Low	High	Medium	Low
LQ1	4.50	3.55	3.09	4.66	3.98	3.11
LQ2	5.00	4.45	4.09	4.84	4.41	4.03
LQ3	4.50	4.00	3.18	4.66	3.86	3.14
LQ4	5.00	4.25	3.64	4.25	4.02	3.39
LQ5	4.63	4.25	3.64	4.28	3.82	3.47
LQ6	3.88	3.95	3.36	4.16	3.73	3.25
LQ7	4.63	4.45	4.18	4.84	4.45	4.08
LQ8	5.00	5.00	4.55	5.00	4.94	4.72
LQ9	5.00	4.85	4.45	5.00	4.84	4.55
LQ10	4.25	3.75	3.36	3.47	3.55	3.40
LQ11	4.38	3.90	3.27	4.34	3.53	2.87
LQ12	4.88	4.85	4.27	4.97	4.49	4.17
LQ13	4.88	4.55	4.18	4.88	4.59	4.21
LQ14	4.88	4.25	4.09	4.66	4.24	3.98
LQ15	4.88	4.45	4.00	4.78	4.57	4.04
LQ16	3.00	2.70	2.55	3.25	2.73	2.43
LQ17	4.88	4.35	4.09	4.88	4.33	4.06
LQ18	4.38	3.80	3.36	4.63	4.20	3.78
LQ19	4.38	3.80	3.64	4.69	4.14	3.47
LQ20	4.38	4.00	3.18	3.34	3.20	3.25
LQ21	4.00	3.25	3.27	4.41	3.31	2.69
LQ22	3.88	3.20	3.00	3.97	3.71	3.31
LQ23	4.50	4.10	3.73	4.72	3.86	3.47
LQ24	4.63	3.50	3.55	4.41	4.12	3.44
LQ25	3.38	3.60	3.36	3.53	3.63	3.42
LQ26	4.25	3.20	2.64	4.28	3.41	2.86
LQ27	4.38	4.00	3.73	4.56	3.98	3.69

final version of 15 items for use. We drew on the construct-driven instrument development approach to create the items, pilot them, and refine an instrument. This allowed us to document the content validity, substance validity, and structural validity of the items (Messick, 1989b). For content validity, our iterative approach to domain analysis, literature review, and expert feedback allowed us to demonstrate that the instrument is based on a careful definition of the domain and informed by a review of the literature on the domain. We showed that our items' content is sampled from that domain appropriately. For substance validity, we used the extensive input from experienced teachers and professional developers to explore possible interpretations of the item texts. We articulated how teachers of varying knowledge and implementation expertise are likely to respond to the items. For structural validity, we applied

the Rasch measurement model's strict interpretation of good measurement, showing that the item response pattern matches the intended scaling and that the items fit the measurement model. From these, we selected a sample of items that address the range of the latent trait while requiring fewer items, making future uses of the questionnaire less burdensome.

The distribution of responses to the questionnaire also showed that teachers were distributed across the entire scale, among both preservice and in-service teachers, showing that in-service teaching experience was not related to the observed knowledge score for the present sample. This may reveal that knowledge of language as an epistemic tool is not acquired through practice alone but may require explicit attention through teacher education or professional development of some kind (Oliveira, 2010). This is consistent with work that indicates many teachers may think of "language" as it is associated with a particular subject matter like English language, or who may think of using language in science simply to access or recall content, whereas to understand language as an epistemic tool requires recognizing that teachers can understand that language development and science conceptual development are complementary (Seah & Yore, 2017), because students need language to think about science phenomena and to generate ideas for themselves (Galbraith, 2009).

Second, the fact that our Rasch analysis of the questionnaire fit a unidimensional model indicates that there is a single, measurable trait that we can observe in the responses. Although we identified and intentionally incorporated all four sub-domains in understanding language as an epistemic tool—such as language being essential or being both process/product—the teachers' responses on the sub-domains are consistent with a single measurable scale. It is unknown whether this scaling is persistent after teachers undergo professional development that explicitly addresses language as an epistemic tool, but this could be studied over the long term.

Based on these aspects of instrument validity and our interpretations, we assert that the instrument now has capacity for use in future research on teachers' knowledge of language as an epistemic tool, whether at one measurement instance or for examining growth over time. There is much room for further work on this instrument to expand on the current results. In our current work, we are pursuing this by studying how teachers' knowledge of language as an epistemic tool develops over a series of professional development experiences, and the connections to teachers' classroom practices. The other possible uses of such an instrument are wide. For instance, future work could examine how teacher education programs help pre-service teachers develop understanding of language's role in learning across disciplines or as a possible gauge for their expertise in supporting students' use of language in learning.

The instrument could be useful in mixed-methods research as an alternative form of data on teachers' understanding of language. For example, the instrument may be incorporated in parallel with research on teacher noticing (Barnhart & van Es, 2015), which examines how teachers recognize, interpret, and respond to students' thinking. Such work could examine, for instance, whether teachers who are more knowledgeable about language as an epistemic tool are also more apt to notice and interpret students' statements during class. This may fit within a push to identify tasks to help characterize and support development of teacher noticing (Estapa et al., 2018).

6.1 | Limitations and future research

The study is not without limitations. The first set of limitations focuses on interpretations we would wish to make using the instrument. We cannot yet make a complete interpretive argument about how teachers' responses to the present instrument is associated with their classroom actions. This is a focus of the larger research project as described above in the Research Context for this Study section. A set of classroom observations and teacher interviews are under way, in parallel with the ongoing professional development. In so doing, we expect to be able to use the instrument on language as an epistemic tool to help us understand possible differences in teachers' adaptations of the SWH approach in their classrooms.

The second set of limitations relate to the current study's methods. One of these potential limitations is due to sample size. While our sample was reasonably large for the number of items, we applied the Rasch partial-credit model (PCM) because it typically serves very well for uncovering item and scale effects in Likert-type rating scale instruments. However, the Rasch PCM can be sensitive to situations where few or no teachers select one of the item-step combinations (e.g., if no teachers select "Strongly Disagree" on one of the relatively easy items). To counteract this limitation will require future research with much larger data sets and include teachers with quite varied prior experience or knowledge. This will afford the ability to corroborate our instrument findings or to contradict and refine the items further.

Another of these potential limitations is due to respondent selection. We purposefully included teachers with varied levels of previous exposure to the SWH approach, which explicitly addresses the role of language in science learning. That said, our sample of teachers all come from Iowa, so may not be representative of all teachers in the US, let alone other countries.

Hence, the patterns of response that we analyze using our data set cannot be generalized to all teachers unquestioningly. Further research in more diverse settings within the US and in other countries is warranted.

A further opportunity is to compare findings from the revised questionnaire with data on teachers' classroom use of language collected using more in-depth methods qualitative methods. That may involve using approaches such as discourse analysis, classroom observation, or video analyses. These intensive qualitative approaches can provide opportunity for further supporting the interpretations from this questionnaire, or may offer contradictory findings that would impel further revisions or alternative applications of the questionnaire.

Finally, our emphasis for this manuscript was to examine three aspects of validity: Content, Substance, and Structure. Future work could seek to examine the other three aspects outlined by Messick (1989b): external criteria, consequence, and generalizability. External criterion aspects of validity include convergent evidence (testing for variables predicted to correlate with the construct of interest) and discriminant evidence (testing for variables that are expected to be unrelated to the construct of interest). Consequential validity includes evidence of positive consequences of the assessment, negative consequences, or negative impact on individuals or groups, value of the score interpretation, and intended or unintended consequences of the assessment score as important to consider when examining consequential validity. Finally, generalizability looks at the degree to which scores (or constructs) can be generalized or connected across other groups and settings. Generalizability can be a double-edged sword as connecting the interpretation of a particular group of scores to another setting (outside of the study) might diminish the validity of the claims if the connection is speculative or unfounded

(Messick, 1989a). Research to explore these remaining three aspects of validity would, therefore, require a wider data collection with other populations and with multiple covariates, and longer-term data collection to allow for studying potential consequences of an interpretation.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. National Science Foundation under Grant No. 1812576. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors wish to thank the in-service and pre-service teachers for their time and participation in the project; our project site coordinators (Bill Crandall, Lee Freeman, Tracy Jarrett, Kim Wise), experienced teachers and PD leaders (Jason Martin-Hiner, Jill Payne, Josh Steenhoek), and affiliated graduate students (Yejun Bae, Ali Cikmaz) for their input on the instrument; and the JRST editors and reviewers for their feedback on the manuscript.

ORCID

Gavin W. Fulmer  <https://orcid.org/0000-0003-0007-1784>

Jihyun Hwang  <https://orcid.org/0000-0003-3107-7142>

Brian Hand  <https://orcid.org/0000-0002-0574-7491>

Jee K. Suh  <https://orcid.org/0000-0002-7370-3896>

ENDNOTE

¹ These teachers were not recruited to take the original instrument during its piloting phase to avoid confusion when they took the revised instrument at the beginning of the subsequent professional development workshops.

REFERENCES

- American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME), Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Anderson, O. R. (1992). Some interrelationships between constructivist models of learning and current neurobiological theory, with implications for science education. *Journal of Research in Science Teaching*, 29(10), 1037–1058.
- Arons, A. (1983). Achieving wider scientific literacy. *Daedalus*, 112(2), 91–122.
- Avraamidou, L., & Osborne, J. (2009). The role of narrative in communicating science. *International Journal of Science Education*, 31(12), 1683–1707.
- Bae, Y., Cikmaz, A., & Hand, B. (2019). Constructing theoretical foundations of immersive learning environments. Paper presented in the 2019 NARST Annual International Conference. Baltimore, MD.
- Balgopal, M., & Montplaisir, L. (2011). Meaning making: What reflective essays reveal about biology students' conceptions about natural selection. *Instructional Science*, 39(2), 137–169.
- Barnhart, T., & van Es, E. (2015). Studying teacher noticing: Examining the relationship among pre-service science teachers' ability to attend, analyze and respond to student thinking. *Teaching and Teacher Education*, 45, 83–93.
- Benedict-Chambers, A., Kademian, S. M., Davis, E. A., & Palincsar, A. S. (2017). Guiding students towards sensemaking: Teacher questions focused on integrating scientific practices with science content. *International Journal of Science Education*, 39(15), 1977–2001.
- Black, P. (2017). *Assessment in science education*. In *science education* (pp. 295–309). Leiden, Netherlands: Brill Sense.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.

- Bond, T., & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Boon, M. (2012). Scientific concepts in the engineering sciences: Epistemic tools for creating and intervening with phenomena. In U. Feest & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 219–243). Berlin: De Gruyter.
- Boon, M., & Van Baalen, S. (2019). Epistemology for interdisciplinary research: Shifting philosophical paradigms of science. *European Journal for Philosophy of Science*, 9, 16.
- Boone, W., Staver, J., & Yale, M. (2014). *Rasch analysis in the human sciences*. Dordrecht: Springer.
- Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE Life Sciences Education*, 15(4), rm4. <https://doi.org/10.1187/cbe.16-04-0148>
- Bradbury, L. (2014). Linking science and language arts: A review of the literature which compares integrated versus non-integrated approaches. *Journal of Science Teacher Education*, 25(4), 465–488.
- Brown, B., & Ryoo, K. (2008). Teaching science as a language: A “content-first” approach to science teaching. *Journal of Research in Science Teaching*, 45(5), 529–553.
- Brown, B., & Spang, E. (2008). Double talk: Synthesizing everyday and science language in the classroom. *Science Education*, 92(4), 708–732.
- Brown, B. A. (2006). “It isn’t no slang that can be said about this stuff”: Language, identity, and appropriating science discourse. *Journal of Research in Science Teaching*, 43(1), 96–126.
- Carlsen, W. S. (2007). Language and science learning. *Handbook of research on science education*, 57–74.
- Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, 25(6), 657–674.
- Cavagnetto, A. R. (2010). Argument to foster scientific literacy: A review of argument interventions in K–12 science contexts. *Review of Educational Research*, 80(3), 336–371.
- Cavagnetto, A., & Hand, B. (2012). The Importance of Embedding Argument Within Science Classrooms. In: Khine M. (ed.) *Perspectives on Scientific Argumentation*. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-2470-9_3
- Cavagnetto, A. R., Hand, B., & Premo, J. (2020). Supporting student agency in science. *Theory into Practice*, 58(1), 12–24.
- Cervetti, G. N., Barber, J., Dorph, R., Pearson, P. D., & Goldschmidt, P. G. (2012). The impact of an integrated approach to science and literacy in elementary school classrooms. *Journal of Research in Science Teaching*, 49(5), 631–658.
- Chen, Y., Park, S., & Hand, B. (2016). Examining the use of talk and writing for students’ development of scientific conceptual knowledge through constructing and critiquing arguments. *Cognition and Instruction*, 34(2), 100–147.
- Chen, Y.-C., Benus, M. J., & Hernandez, J. (2019). Managing uncertainty in scientific argumentation. *Science Education*, 103(5), 1235–1276.
- Chen, Y. C., Hand, B., & McDowell, L. (2013). The effects of writing-to-learn activities on elementary students’ conceptual understanding: Learning about force and motion through writing to older peers. *Science Education*, 97(5), 745–771.
- Chin, C., Yang, W., & Tuan, H. (2016). Argumentation in a socioscientific context and its influence on fundamental and derived science literacies. *International Journal of Science and Mathematics Education*, 14(4), 603–617.
- Choi, A., Hand, B., & Norton-Meier, L. (2014). Grade 5 students’ online argumentation about their in-class inquiry investigations. *Research in Science Education*, 44(2), 267–287.
- Chou, Y.-T., & Wang, W.-C. (2010). Checking dimensionality in item response models with principal component analysis on standardized residuals. *Educational and Psychological Measurement*, 70(5), 717–731. <https://doi.org/10.1177/0013164410379322>
- Cikmaz, A., Hand, B., Fulmer, G. W., & Yaman, F. (2019). The Impact of Teachers’ Epistemic Orientations on Growth of Representation Competence.
- Cikmaz, A., Fulmer, G. W., Yaman, F., & Hand, B. (2019). Examining growth and interdependence of epistemic tools in different learning environments. Paper presented in the 2019 NARST Annual International Conference. Baltimore, MD.

- Cikmaz, A., Hand, B., Fulmer, G. W., & Yaman, F. (2019). The impact of teachers' epistemic orientations on growth of representation competence. Paper presented in the 2019 NARST Annual International Conference. Baltimore, MD.
- Corballis, M. C. (2017). Language evolution: A changing perspective. *Trends in Cognitive Sciences*, 21(4), 229–236.
- Demirbag, M., & Gunel, M. (2014). Integrating argument-based science inquiry with modal representations: Impact on science achievement, argumentation, and writing skills. *Educational Sciences: Theory and Practice*, 14(1), 386–391.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199.
- Duschl, R., & Bybee, A. (2014). Planning and carrying out investigations: An entry to learning and to teacher professional development around NGSS science and engineering practices. *International Journal of STEM Education*, 1(1), 1–9.
- Duschl, R., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, 38, 39–72.
- Duschl, R., Schweingruber, H., & Shouse, A. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.
- Duschl, R. A. (2005). The high school lab experience: reconsidering the role of evidence, explanation and the language of science. Paper commissioned for the Committee on the Role of the Laboratory in High School Science of the National Research Council. Washington, DC: National Academies. Retrieved from https://sites.nationalacademies.org/cs/groups/dbassessite/documents/webpage/dbasse_073329.pdf.
- Estapa, A. T., Amador, J., Kosko, K. W., Weston, T., de Araujo, Z., & Aming-Attai, R. (2018). Preservice teachers' articulated noticing through pedagogies of practice. *Journal of Mathematics Teacher Education*, 21(4), 387–415.
- Galbraith, D. (2009). Writing about what we know: Generating ideas in writing. In R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.), *The SAGE handbook of writing development* (pp. 48–64). Thousand Oaks, CA: SAGE.
- Gee, J. (2000). Identity as an analytic lens for research in education. *Review of Research in Education*, 25, 99–125.
- Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology*, 26(5), 796–804.
- Glynn, S., & Muth, D. (1994). Reading and writing to learn science: Achieving scientific literacy. *Journal of Research in Science Teaching*, 31(9), 1057–1073.
- Graham, S., Capizzi, A., Harris, K. R., Hebert, M., & Morphy, P. (2014). Teaching writing to middle school students: A national survey. *Reading and Writing*, 27(6), 1015–1042.
- Granger, C. (2008). Rasch analysis is important to understand and use for measurement. *Rasch Measurement Transactions*, 21(3), 1122–1123.
- Halliday, M. A. K., & Martin, J. R. (2003). *Writing science: Literacy and discursive power*. Abingdon, UK: Routledge.
- Hand, B., Alvermann, D., Gee, J., Guzzetti, B., Norris, S., Phillips, L., ... Yore, L. D. (2003). Message from the "Island group": What is literacy in science literacy? *Journal of Research in Science Teaching*, 40(7), 607–615.
- Hand, B., Cavagnetto, A., Chen, Y.-C., & Park, S. (2016). Moving past curricula and strategies: Language and the development of adaptive pedagogy for immersive learning environments. *Research in Science Education*, 46(1), 223–241.
- Hand, B., & Choi, A. (2010). Examining the impact of student use of multiple modal representations in constructing arguments in organic chemistry laboratory classes. *Research in Science Education*, 40(1), 29–44.
- Hand, B., & Keys, C. (1999). Inquiry investigation. *The Science Teacher*, 66(4), 27–29.
- Hand, B., Norton-Meier, L., Gunel, M., & Akkus, R. (2016). Aligning teaching to learning: A 3-year study examining the embedding of language and argumentation into elementary science classrooms. *International Journal of Science and Mathematics Education*, 14(5), 847–863.
- Hand, B., & Prain, V. (2006). Moving from border crossing to convergence of perspectives in language and science literacy research and practice. *International Journal of Science Education*, 28(2–3), 101–107.
- Hand, B., Prain, V., & Yore, L. (2001). Sequential writing tasks' influence on science learning. In P. Tynjala, L. Mason, & K. Lonka (Eds.), *Writing as a learning tool: Integrating theory and practice* (pp. 105–129). Dordrecht, the Netherlands: Springer.

- Hand, B., Shelley, M. C., Laugerman, M., Fostvedt, L., & Therrien, W. (2018). Improving critical thinking growth for disadvantaged groups within elementary school science: A randomized controlled trial using the science writing heuristic approach. *Science Education*, 102(4), 693–710.
- Hand, B., Wallace, C., & Yang, E. (2004). Using a science writing heuristic to enhance learning outcomes from laboratory activities in seventh-grade science: Quantitative and qualitative aspects. *International Journal of Science Education*, 26(2), 131–149.
- Hanrahan, M. (1999). Rethinking science literacy: Enhancing communication and participation in school science through affirmational dialogue journal writing. *Journal of Research in Science Teaching*, 36(6), 699–717.
- Haun, D. B., Rapold, C. J., Janzen, G., & Levinson, S. C. (2011). Plasticity of human spatial cognition: Spatial language and cognition covary across cultures. *Cognition*, 119(1), 70–80. <https://doi.org/10.1016/j.cognition.2010.12.009>
- Huang, H. (2006). Listening to the language of constructing science knowledge. *International Journal of Science and Mathematics Education*, 4(3), 391–415.
- Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299–314.
- Kane, M. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319–342.
- Kelly, G. J., & Cunningham, C. M. (2019). Epistemic tools in engineering design for K-12 education. *Science Education*, 103(4), 1080–1111.
- Keys, C., Hand, B., Prain, V., & Collins, S. (1999). Using the science writing heuristic as a tool for learning from laboratory investigations in secondary science. *Journal of Research in Science Teaching*, 36(10), 1065–1084.
- Kiuhara, S. A., Graham, S., & Hawken, L. S. (2009). Teaching writing to high school students: A national survey. *Journal of Educational Psychology*, 101(1), 136–160.
- Ko, M.-L. M., & Krist, C. (2019). Opening up curricula to redistribute epistemic agency: A framework for supporting science teaching. *Science Education*, 103(4), 979–1010.
- Krajcik, J., Blumenfeld, P., Marx, R., Bass, K., Fredricks, J., & Soloway, E. (1998). Inquiry in project-based science classrooms: Initial attempts by middle school students. *Journal of the Learning Sciences*, 7(3–4), 313–350.
- Larsson, P., & Jakobsson, A. (2019). Meaning-making in science from the perspective of students' hybrid language use. *International Journal of Science and Mathematics Education*, 18, 811–820. <https://doi.org/10.1007/s10763-019-09994-z>.
- Leinke, J. (1990). *Talking science: Language, learning, and values*, New York: Ablex Publishing Corporation.
- Leinke, J. (2004). The literacies of science. In E. W. Saul (Ed.), *Crossing borders in literacy and science instruction: Perspectives in theory and practice* (pp. 33–47). Arlington, VA: NSTA Press.
- Lemmi, C., Brown, B. A., Wild, A., Zummo, L., & Sedlacek, Q. (2019). Language ideologies in science education. *Science Education*, 103(4), 854–874.
- Linacre, J. M. (2003). Data variance: Explained, modeled and empirical. *Rasch Measurement Transactions*, 17(3), 942–943.
- Lissitz, R. W., & Green, S. B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology*, 60(1), 10–13.
- Lissitz, R. W., & Samuelson, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher*, 36(8), 437–448.
- Liu, X. (2010). *Using and developing measurement instruments in science education: A Rasch modeling approach*. Charlotte, NC: IAP.
- Lynch, S. (2001). “Science for All” is not equal to “One Size Fits All”: Linguistic and cultural diversity and science education reform. *Journal of Research in Science Teaching*, 38(5), 622–627.
- Martin, A., & Hand, B. (2009). Factors affecting the implementation of argument in the elementary science classroom. A longitudinal case study. *Research in Science Education*, 39(1), 17–38.
- Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174.
- McDermott, M. A., & Hand, B. (2016). Modeling scientific communication with multimodal writing tasks: Impact on students at different grade levels. In B. Hand, M. A. McDermott, & V. Prain (Eds.), *Using multimodal representations to support learning in the science classroom* (pp. 183–212). Dordrecht, the Netherlands: Springer.
- McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing*, 18(4), 333–349.

- Messick, S. (1989a). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5–11.
- Messick, S. (1989b). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). New York, NY: American Council on Education.
- Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, 25(4), 6–20.
- National Research Council (NRC). (2000). *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/9853>
- National Research Council (NRC). (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/13165>
- NGSS Lead States. (2013). *Next generation science standards: For States, by States*. Washington, DC: The National Academies Press.
- Norris, S., & Phillips, L. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, 87(2), 224–240.
- Olander, C., Wickman, P. O., Tytler, R., & Ingeman, Å. (2018). Representations as mediation between purposes as junior secondary science students learn about the human body. *International Journal of Science Education*, 40(2), 204–226.
- Oliveira, A. W. (2010). Developing elementary teachers' understanding of the discourse structure of inquiry-based science classrooms. *International Journal of Science and Mathematics Education*, 8(2), 247–269.
- Oyoo, S. (2012). Language in science classrooms: An analysis of physics teachers' use of and beliefs about language. *Research in Science Education*, 42(5), 849–873.
- Pelger, S., & Nilsson, P. (2016). Popular science writing to support students' learning of science and scientific literacy. *Research in Science Education*, 46(3), 439–456.
- Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences*, 107(Supplement 2), 8993–8999.
- Plake, B., & Wise, L. (2014). What is the role and importance of the revised AERA, APA, NCME standards for educational and psychological testing? *Educational Measurement: Issues and Practice*, 33(4), 4–12.
- Postman, N., & Weingartner, C. (1969). *Teaching as a subversive activity*. New York: Delta.
- Prain, V., & Hand, B. (2016a). Coming to know more through and from writing. *Educational Researcher*, 45(7), 430–434.
- Prain, V., & Hand, B. (2016b). Learning science through learning to use its languages. In B. Hand, M. A. McDermott, & V. Prain (Eds.), *Using multimodal representations to support learning in the science classroom* (pp. 1–10). Dordrecht, the Netherlands: Springer.
- Prain, V., & Waldrup, B. (2006). An exploratory study of teachers' and students' use of multi-modal representations of concepts in primary science. *International Journal of Science Education*, 28(15), 1843–1866.
- Pruitt, S. (2014). The next generation science standards: The features and challenges. *Journal of Science Teacher Education*, 25(2), 145–156.
- Rivard, L. (2004). Are language-based activities in science effective for all students, including low achievers? *Science Education*, 88(3), 420–442.
- Rivard, L., & Straw, S. (2000). The effect of talk and writing on learning science: An exploratory study. *Science Education*, 84(5), 566–593.
- Robitzsch, A., Kiefer, T., Wu, M., Robitzsch, M. A., Adams, W., Rcpp, L., & LSAmITR, R. E. (2019). Package 'TAM'.
- Schoerning, E., Hand, B., Shelley, M., & Therrien, W. (2015). Language, access, and power in the elementary science classroom. *Science Education*, 99(2), 238–259.
- Seah, L. H., Clarke, D. J., & Hart, C. E. (2014). Understanding the language demands on science students from an integrated science and language perspective. *International Journal of Science Education*, 36(6), 952–973.
- Seah, L. H., & Yore, L. D. (2017). The roles of teachers' science talk in revealing language demands within diverse elementary school classrooms: A study of teaching heat and temperature in Singapore. *International Journal of Science Education*, 39(2), 135–157.
- Settlage, J., & Southerland, S. A. (2019). Epistemic tools for science classrooms: The continual need to accommodate and adapt. *Science Education*, 103(4), 1112–1119.
- Sinatra, G., & Broughton, S. (2011). Bridging reading comprehension and conceptual change in science education: The promise of refutation text. *Reading Research Quarterly*, 46(4), 374–393.

- Sondergeld, T. A., & Johnson, C. C. (2014). Using Rasch measurement for the development and use of affective assessments in science education research. *Science Education*, 98(4), 581–613.
- Syh-Jong, J. (2007). A study of students' construction of science knowledge: Talk and writing in a collaborative group. *Educational Research*, 49(1), 65–81.
- Tang, S., Delgado, C., & Moje, E. (2013). An integrative framework for the analysis of multiple and multimodal representations for meaning-making in science education. *Science Education*, 98(2), 305–326.
- Tran, H., Griffin, P., & Nguyen, C. (2010). Validating the university entrance English test to the Vietnam National University: A conceptual framework and methodology. *Procedia Social and Behavioral Sciences*, 2(2), 1295–1304.
- Tytler, R., Prain, V., Aranda, G., Ferguson, J., & Gorur, R. (2020). Drawing to reason and learn in science. *Journal of Research in Science Teaching*, 57(2), 209–231.
- Tytler, R., Prain, V., & Hubber, P. (2018). Representation construction as a core science disciplinary literacy. In *Global developments in literacy research for science education* (pp. 301–317). Dordrecht, the Netherlands: Springer.
- Van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science* (New York, N.Y.), 328(5977), 453–456.
- Villanueva, M., & Hand, B. (2011). Science for all: Engaging students with special needs in and about science. *Learning Disabilities Research & Practice*, 26(4), 233–240.
- Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Waldrip, B., & Prain, V. (2006). Changing representations to learn primary science concepts. *Teaching Science: The Journal of the Australian Science Teachers Association*, 52(4), 17–21.
- Waldrip, B., Prain, V., & Carolan, J. (2010). Using multi-modal representations to improve learning in junior secondary science. *Research in Science Education*, 40(1), 65–80.
- Wallace, C. (2004). Framing new research in science literacy and language use: Authenticity, multiple discourses, and the “Third Space”. *Science Education*, 88(6), 901–914.
- Wang, J. (2020). Scrutinising the positions of students and teacher engaged in argumentation in a high school physics classroom. *International Journal of Science Education*, 42(1), 25–49.
- Wang, J., Wang, Y., Tai, H., & Chen, W. (2010). Investigating the effectiveness of inquiry-based instruction on students with different prior knowledge and reading abilities. *International Journal of Science and Mathematics Education*, 8(5), 801–820.
- Warren, B., Ballenger, C., Ogonowski, M., Rosebery, A., & Hudicourt-Barnes, J. (2001). Rethinking diversity in learning science: The logic of everyday sense-making. *Journal of Research in Science Teaching*, 38(5), 529–552.
- Wellington, J., & Osborne, J. (2001). *Language and literacy in science education*. London: McGraw-Hill Education.
- Weng, L.-J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64(6), 956–972. <https://doi.org/10.1177/0013164404268674>
- Wiser, M., & Amin, T. (2001). “Is heat hot?” Inducing conceptual change by integrating everyday and scientific perspectives on thermal phenomena. *Learning and Instruction*, 11(4–5), 331–355.
- Wright, B. D. (1996). Reliability and separation. *Rasch Measurement Transactions*, 9(4), 472.
- Yaman, F. (2017). Effects of the science writing heuristic approach on the quality of prospective science teachers' argumentative writing and their understanding of scientific argumentation. *International Journal of Science and Mathematics Education*, 16(3), 421–442.
- Yaman, F. (2018). Pre-service science teachers' development and use of multiple levels of representation and written arguments in general chemistry laboratory courses. *Research in Science Education*, 1–32. <https://doi.org/10.1007/S11165-018-9781-0>.
- Yore, L., Bisanz, G., & Hand, B. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education*, 25(6), 689–725.
- Yore, L., Hand, B., Goldman, S., Hildebrand, G., Osborne, J., Treagust, D., & Wallace, C. (2004). New directions in language and science education research. *Reading Research Quarterly*, 39(3), 347–352.

- Yore, L., & Treagust, D. (2006). Current realities and future possibilities: Language and science literacy—Empowering research and informing instruction. *International Journal of Science Education*, 28(2–3), 291–314.
- Yore, L. D., Chinn, P. W. U., & Hand, B. (2008). Science literacy for all students: Language, culture, and knowledge about nature and naturally occurring events. *L1 Educational Studies in Language and Literature*, 8(1), 1–3.
- Yore, L. D., & Hand, B. (2010). Epilogue: Plotting a research agenda for multiple representations, multiple modality, and multimodal representational competency. *Research in Science Education*, 40(1), 93–101.

How to cite this article: Fulmer GW, Hwang J, Ding C, Hand B, Suh JK, Hansen W. Development of a questionnaire on teachers' knowledge of language as an epistemic tool. *J Res Sci Teach*. 2020;1–32. <https://doi.org/10.1002/tea.21666>