



Concept Mapping as a Mechanism for Assessing Science Teachers' Cross-Disciplinary Field-Based Learning

Joanna K. Garner, Avi Kaplan, Stephanie Hathcock & Bradley Bergey

To cite this article: Joanna K. Garner, Avi Kaplan, Stephanie Hathcock & Bradley Bergey (2019): Concept Mapping as a Mechanism for Assessing Science Teachers' Cross-Disciplinary Field-Based Learning, Journal of Science Teacher Education

To link to this article: <https://doi.org/10.1080/1046560X.2019.1625573>



Published online: 11 Jul 2019.



Submit your article to this journal [↗](#)



View Crossmark data [↗](#)



Concept Mapping as a Mechanism for Assessing Science Teachers' Cross-Disciplinary Field-Based Learning

Joanna K. Garner ^a, Avi Kaplan ^b, Stephanie Hathcock ^c, and Bradley Bergey ^d

^aThe Center for Educational Partnerships, Old Dominion University, Norfolk, Virginia, USA; ^bDepartment of Psychological Studies in Education, Temple University, Philadelphia, Pennsylvania, USA; ^cSchool of Teaching, Learning and Educational Sciences, Oklahoma State University, Stillwater, Oklahoma, USA; ^dDepartment of Secondary Education and Youth Services, The City University of New York, Queens, New York, USA

ABSTRACT

Two common goals of science teacher professional development (PD) are increased content knowledge (CK) and improved readiness to teach through inquiry. However, PD assessment challenges arise when the context is structured around inquiry-based, participant-driven learning, and when the content crosses scientific disciplines. This study extended the use of concept mapping as an assessment tool for examining changes in the content knowledge of 21 high school science teachers who participated in a field-based environmental science summer institute. The scoring rubric focused on documenting concepts, links, and map organization and scope in an attempt to capture development of cross-disciplinary knowledge in ways that correspond with theories of expertise development. The analysis revealed significant gains from pre-PD to post PD maps in the sophistication of links between concepts and in the number of additional, participant-generated scientifically valid concepts. Relative to the initial maps, post PD maps also manifested more complete clustering of concepts. Findings are discussed in reference to previous studies on teachers' learning and implications for future research using concept mapping as a means of assessing teacher PD.

KEYWORDS

Concept mapping; science teacher professional development; field studies; rubrics

Contemporary trends in science curriculum standards and pedagogy emphasize cross-disciplinary concepts and themes as well as the use of inquiry-based instructional methods (Achieve, 2013; Bodzin, Klein, & Weaver, 2010). Authentic, field-based inquiry offers students a mechanism for learning science content and developing an appreciation for the interconnected nature of science and its application to real world settings, and can impact student interest through increases in perceived relevance of the material (Chesapeake Bay Foundation, 2012; Czerniak & Johnson, 2014; North American Association for Environmental Education, 2012). However, leveraging the pedagogical opportunities of inquiry, particularly in a field setting, can be challenging for teachers due to the high demands on their content knowledge. In order to support students' inquiry-based investigations and their developing appreciation for the cross-disciplinary nature of applied science, teachers must be able to make connections and apply scientific inquiry strategies to the study of complex, real-world settings (Crawford, 2000, 2007; Fishman, Marx, Best, & Tal, 2003; Nagle, 2013). That is, teachers must possess

CONTACT Joanna K. Garner  jkgarner@odu.edu  The Center for Educational Partnerships, Old Dominion University, Norfolk, VA 23508, USA

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uste.

© 2019 Association for Science Teacher Education

knowledge that is also cross-disciplinary and interconnected beyond isolated concepts or principles.

Teachers' content knowledge has been identified as an important factor in improving student learning outcomes (Abell, 2007; Gess-Newsome et al., 2017; Van Driel, Berry, & Meirink, 2014). As a result, science teacher professional development (PD) initiatives often provide teachers with opportunities to learn science content as well as pedagogy. Whereas a large body of literature has documented the impact of PD activities on teachers' attitudes and instructional intentions (Borko, 2004; Desimone, 2009; Enderle et al., 2014; Moyer-Packenham, Bolyard, Oh, & Cerar, 2011; Russell & Hancock, 2007), less is known about how PD impacts change in teachers' content knowledge (Luft & Hewson, 2014). Furthermore, evaluations of PD interventions often employ self-report measures (Abell, 2007), content-based pre- and post-tests with items drawn from standardized science tests (Van Driel et al., 2014), or measures of knowledge and attitudes pertaining to particular topics (e.g., climate change, Hayhoe, Bullock, & Hayhoe, 2011). Although useful, such measures may not be well suited for evaluating the open-ended learning that takes place in cross-disciplinary, inquiry-focused PD programs where teachers' prior knowledge varies substantially and the to-be-learned content is not entirely pre-determined at the outset.

In this article, we describe the use of concept maps for evaluating changes in the scope and interconnectivity in teachers' content knowledge while maintaining the ability to identify consistent features across individuals and concepts (Brown & MacIntyre, 1993). We provide an approach to working with subject matter experts that results in an example of how concept mapping can serve as an assessment tool that is systematic yet sympathetic to the wide variation in teachers' prior and emerging knowledge. The result is a process that is congruent with contemporary constructivist perspectives that view teachers' learning during PD as knowledge (re)construction (Daley, Shaw, Balistrieri, Glasenapp, & Piacentine, 1999; Luft & Hewson, 2014).

Concept maps, knowledge, and learning

A concept map is a two-dimensional, hierarchically arranged, graphical representation of relations between concepts (Novak, 1990, 2005; Novak & Gowin, 1984). Concepts connote ideas and are typically presented within a circle. Labeled lines drawn between circles represent part-whole, sequential order, or cause and effect relations between these terms (Herl, Baker, & Niemi, 1996). Connected concepts and links create propositional phrases and sentences (e.g. "x causes y") and allow the map to be "read" from concept to concept (Novak & Cañas, 2004).

Concept maps are thought to parallel the structure of a learner's knowledge network. According to cognitive and constructivist theories, knowledge acquisition includes the construction of hierarchical arrangements among central and peripheral concepts with proposition-like linkages between them (Anderson, 2000; Herl et al., 1996; Jacobs-Lawson & Hershey, 2002). Concepts – language labels or symbols that capture regularities in the world – and meaningful links between them, form the building blocks of knowledge networks (Novak & Cañas, 2004; Rumelhart, Lindsay, & Norman, 1972).

The process of learning involves relating new concepts to existing ones, with meaningful learning occurring when ideas and the relations among them are reconstructed (Ausubel, 1968). Concepts and their relations become reorganized in hierarchical ways that facilitate the assimilation and interpretation of new information (Bransford, Brown, & Cocking, 2000). As

knowledge is gained, the individual's network grows in its capacity to specify the type of relationship between ideas (e.g. part-whole, cause and effect, similarity, etc.) and support transfer from one context to another (Chi, Glaser, & Farr, 1988; Dansereau & Holley, 1982; Yarden, Marbach-A., & Gershoni, 2004). As a result of learning, iterations of an individual's concept map may come to include new hierarchical map configurations of concepts and links, while unnecessary concepts may be removed (Alexander, 2003; McClure, Sonak, & Suen, 1999; Novak & Cañas, 2004; Novak & Gowin, 1984; Ruiz-Primo & Shavelson, 1996; Tsai, Lin, & Yuan, 2001).

Teachers' learning of subject-matter content knowledge may produce changes in the interconnections and organization of their networks of scientific knowledge (Alonzo, Kobarg, & Seidel, 2012). It may manifest in different ways for individuals in the same learning environment, which in turn could manifest in unique positions and interrelations among concepts in the representation map. However, consistencies across groups of individuals who share experiences in a learning environment may also appear, including changes to the structural features of knowledge networks such as organization, hierarchical structure, or links between ideas. These patterns of uniqueness and similarity are likely outcomes when learners have the opportunity to direct their own learning yet experience a shared activity, such as when a heterogeneous group of science teachers works collaboratively on tasks that require scientific inquiry.

Science teachers' subject matter knowledge

There are many studies describing how pre-service and in-service teachers develop their pedagogical content knowledge – their knowledge and capacity to teach particular content with specific methods – but less is known about how teachers' *subject matter knowledge* develops and changes over time. Yet, content or subject matter knowledge of, for example, biology, chemistry, and mathematics concepts and principles, is foundational for science teachers (Rollnick, Bennett, Rhemtula, & Dharsey, 2008). Teachers' subject matter knowledge contributes to pedagogical knowledge and influences instructional decision making (Kirschner, Borowski, & Fischer, 2011; Shulman, 1986). Strong content knowledge is associated with less reliance on teacher-centered strategies such as lecture and rote memorization (Abell, 2007) and use of high quality teaching behaviors such as the provision of clear explanations (Van Driel et al., 2014). Teachers' subject matter knowledge has also been associated with higher student achievement (Supovitz & Turner, 2000).

Early career teachers, those teaching outside of their area of specialization, and those who must teach with new content or themes within curriculum standards may find their subject-matter knowledge to be “fragmented,” meaning that concepts are not necessarily linked with one another or to important ideas or desired learning contexts within the mandated curriculum (Mäntylä & Nousiainen, 2013). The need for teachers to develop richer and more coherent knowledge has led to continued attention to the inclusion of subject matter learning opportunities in teacher professional development programs (Capps, Crawford, & Constatas, 2012; Luft & Hewson, 2014; Supovitz & Turner, 2000). Specifically, researchers have proposed that content should be organized within and across disciplines (Luft & Hewson, 2014), should promote conceptual knowledge as well as knowledge of scientific practices (Capps et al., 2012), and should be considered within the overall domain of pedagogy (Diamond, Maerten-Rivera, Rohrer, & Lee, 2014).

Cross-disciplinary, field-based inquiry

The implementation of The Next Generation Science Standards (Achieve, 2013), which emphasize inquiry-based activities as well as learning in applied contexts, has been accompanied by the endorsement of tasks requiring cross-disciplinary knowledge and its application as a way to focus students' attention on the types of questions and challenges pursued by practicing scientists. Such attention, in turn, promotes students' appreciation for the need to combine knowledge from multiple domains and perspectives to solve meaningful personal and societal problems (Nagle, 2013). This interest in integrated pedagogical approaches has also generated calls for promoting teachers' integrated cross-disciplinary knowledge so that they can facilitate problem solving and inquiry in authentic learning environments (Achieve, 2013; National Research Council, 1996; Watanabe & Huntley, 1998). Accordingly, some researchers and practitioners have implemented field-based teacher PD aiming to influence teachers' views on environmental literacy and understanding of Nature of Science principles (Van Wylen, Abdella, Dickinson, Engbrecht, & Vandiver, 2013).

As an area of research, the topic of teachers' cross-disciplinary knowledge is still relatively new, and is often limited by a reliance on self-report data as means of providing evidence of effecting change in teachers' knowledge (Rennie, Venville, & Wallace, 2012). For example, Holden, Groulx, Bloom, and Weinburgh (2011) investigated the impact of outdoor field experiences over a two-week institute in which elementary, middle and high school teachers explored topics in biodiversity, and found increases in self-reported knowledge and self-efficacy for field-based science instruction. Similarly, Dresner and Worley (2006) described the impact of teacher-scientist collaborations around topics in ecology following a five week field-based research program, and found that more than three fourths of the teachers self-reported gains in knowledge and skills. One possible reason why self-report data remains so prevalent is that PD initiatives that seek to integrate and apply scientific concepts and authentic, inquiry-based practices across disciplines are confronted with the problem of how to assess changes in teachers' subject matter knowledge, given that the anticipated changes might be unspecified at the outset of the program, and the outcome of teachers' learning may include not only the assimilation of new concepts but a reconfiguration of existing ones and a propensity to apply existing knowledge in new ways. This challenge calls for an assessment method that can detect changes in the quantity and configuration of teachers' cross-disciplinary knowledge, yet can be implemented systematically in a manner that reflects universal principles of knowledge development. For this reason, we turned to concept maps as an assessment tool. This decision necessitated the development of a systematic protocol and scoring rubric.

Assessing teacher learning via concept maps

The use of concept maps as a tool for assessing teachers' learning begins with the premise that a comparative analysis of maps generated by the same individual at different time points allows the researcher to investigate change in knowledge content and structure, i.e., learning (Novak, 2005; Rye, Landenberger, & Warner, 2012). This leads to the question of how to systematically capture change, as methods vary in regard to soliciting and scoring maps. In some but not all cases, respondents have been provided with a list of expert-generated terms. Maps commonly begin with a prompt, but this may involve responding to a topic, such as

“runoff,” or to a prompt question. Each strategy has implications for the interpretation of the resulting data, as the provision of concepts and even examples of link types can promote consistency within and among individuals as they build their maps which facilitates reliable scoring (Jin & Wong, 2014; Yin & Shavelson, 2008). However, it may also constrain or otherwise skew the content and configuration of the resulting maps.

Rubric based approaches to map scoring are common but also vary depending on whether or not the map is compared to an expert-generated reference map, and whether a holistic or analytic, structural-relational approach is taken (Chang, Sung, Chang, & Lin, 2005; Novak & Cañas, 2004; Novak & Gowin, 1984; Ruiz-Primo & Shavelson, 1996). The expert comparison assumes that learning can result in an optimal representation. This approach is appropriate for well-defined topics or learning experiences, but it is less well suited to contexts where learners are engaged in individualized or inquiry based projects. Similarly, although a holistic, qualitative rating provides information about map comprehensiveness and overall accuracy, structural-relational methods that assess the number and quality of concepts and links can generate scores that are sensitive to small scale as well as large scale knowledge change (Besterfield-Sacre, Gerchak, Lyons, Shuman, & Wolfe, 2004). In some cases, researchers have elected to combine these strategies in order to permit “top-down” expert ratings in comparison to a target prompt or question along with “bottom up” scoring for the inclusion of idiosyncratic but relevant ideas (e.g. Greene, Lubin, Slater, & Walden, 2013).

Only a small number of studies have described the use of concept maps as tool for assessing teachers’ mathematics or science content knowledge. Hough, O’Rode, Terman, and Weissglass (2007) gathered pretest and posttest maps from teachers attending an algebra focused summer institute by asking participants to use self-generated terms and a topic rather than a question. Using an analytic rubric, they detected increases in the number of concepts, the hierarchical structure of the maps, and the numbers of cross-linked concepts. Hough and colleagues also analyzed the number of “chunks” or small clusters of concepts that were presented and reported increases in the number of clusters relating to key topics within algebra such as graphing, equations, patterns, and formulas. Their findings are supportive of the idea that PD can influence the way in which teachers’ content knowledge is organized. In a study with pre-service teachers, Austin and Schmidt (2010) used concept mapping to investigate the development of expert-like knowledge features such as noticing and abstracting patters, using an overarching framework to organize information, and using a contextual clue as a means of flexible knowledge retrieval and application. They used concept maps to solicit cross-disciplinary knowledge representations of environmental science, and in doing so concluded that participants were able to create maps that reflected cross-disciplinary connections.

Two prior studies of teachers’ learning of science concepts in an inquiry-based PD setting have successfully utilized concept maps as an assessment tool. Greene et al. (2013) collected concept maps from in-service teachers before and after a two week summer institute in which one of nine different science topics could be investigated. Participants created maps that elaborated on a topic reflecting a central theme in their research activities, such as “fossil remains.” Greene and her colleagues compared holistic, qualitative scoring to quantitative measures of nodes, links, and map complexity. They found pretest to posttest differences according to these measures and concluded that concept mapping could reveal changes in the amount and configuration of teachers’ knowledge, particularly when using a structural-relational approach to scoring.

Finally, in a study by Rye et al. (2012), teachers learned about topics in watershed and geosciences education through a summer institute that included concept mapping as an assessment tool as well as a pedagogical strategy. Although the focus of the study was on teachers' acquisition of knowledge regarding watershed topics and of concept mapping as a tool for assessing problem based learning, the investigators examined changes in teachers' pedagogical and content knowledge by soliciting concept maps at various times during and after the institute. The resulting maps exhibited a blend of science content and pedagogical concepts, and demonstrated the potential for concept maps for externalizing individual and collaborative learning processes that take place during teacher PD.

The present study

In this exploratory study, we used concept maps in the context of a field-based science teacher professional development with the goal of creating a cross-disciplinary mapping protocol and scoring rubric sensitive to the many ways in which biology, physics, chemistry and earth science concepts can be integrated and applied in environmental science-based inquiry projects. The protocol and rubric were aligned with established theories of knowledge construction and expertise development (Bransford et al., 2000; Novak & Cañas, 2004). We took as a starting point that concept maps reflect the structure and relations among concepts within an individual's developing knowledge of a particular domain (Novak & Gowin, 1984), and that there is merit in drawing a parallel between the development of expertise and its associated changes in the breadth, depth and organization of knowledge, and changes in the way that concepts, links, and hierarchical structures may appear within a concept map (Alexander, 2003; Bransford, Brown & Cocking, 2004; Mulder, Lazonder, & De Jong, 2015; Novak, 2005). The assessment development process was designed to allow us to investigate three empirical questions about teachers' content knowledge learning; specifically, we focused on changes in the structural coherence of knowledge, the use of scientifically accurate terminology, and the inclusion of expert-generated and self-generated concepts and terms. Our research questions were as follows:

- (1) Which features of concept map structure changed over time from pre-PD to post PD?
- (2) What changes were evident in the scientific accuracy of participants' maps from pretest to posttest?
- (3) How did participants' use of expert- and self-generated concepts change from pretest to posttest?

Methods

Design

The design of the study was a single cohort, pretest posttest design. Since all participants completed the institute, the nature of the PD context, format, and research purpose meant that a comparison group was not assigned. The mapping prompt was designed to allow participants to complete the activity regardless of the specific content of the scientific investigation project. This was important because participants were

permitted to work in interdisciplinary teams to design their own field-based investigations once the institute was underway. Similarly, the scoring rubric was designed to allow the scientist facilitators and researchers to obtain information about the overall impact of institute as well as the particular concepts and questions on which participants were focused.

Participants

Participants were 21 high school science teachers ($n = 15$ female, $n = 6$ male) from a large, suburban school district in the mid-atlantic region of the United States. Participants were enrolled in a two week summer institute that was part of an ongoing collaborative project between two institutions of higher education and the school district. A voluntary demographic survey was made available to participants. The survey data indicated that participants ranged in age from 27 to 56 years of age and that 90% were Caucasian. They reported from 1 to 28 years of high school science teaching experience ($M = 13$ years). Teachers reported that they had most recently taught one of four disciplines: Earth Science ($n = 8$), Biology ($n = 4$), Chemistry ($n = 7$), or Physics ($n = 2$). Approximately 50% of respondents indicated that they taught at least one Advanced Placement course. Participants were compensated for their attendance at the institute but were not provided with any additional compensation for generating concept maps.

Context

The summer institute included eight days of activities spread over two weeks. Teachers explored new science content and discussed strategies for implementing guided inquiry and field-based science lessons. Within the umbrella topic of earth and environmental science, they worked in self-selected small groups to design, execute, analyze and present a field-based investigation that could either examine and illustrate the processes and components of a tidal saltmarsh or examine and illustrate interconnections among science principles and concepts as they related to an urban greenspace. Investigation design began on day one, and continued with data collection on day two followed by data analysis and interpretation that continued through the end of the first week. Ahead of the concept mapping activity, approximately half of the teachers ($n = 11$) self-selected to work on an investigation that could be conducted in a wooded area, and the other half ($n = 10$) self-selected to work on an investigation that could be conducted in a nearby wetland area. Participants designed their investigations based on consensus agreements within their group, which required the groups to balance collective expertise in particular subject areas and data collection methods with interests that arose from considering the opportunities to study specific sites and topics.

Five faculty scientists with expertise in biology, chemistry, physics, oceanography, and geology provided content and process support for the teachers. At the beginning of the institute, they provided satellite and topographical map-based overviews of the field sites followed by a visit to each location. Scientists assisted participants as they generated data collection strategies for their investigation. Support continued during data analysis and interpretation, through individual, small group, and large group instruction. To facilitate the design of their data collection protocols, participants were shown an array of tools and scientific equipment choices available to them, including surveying equipment, soil and

water quality testing kits, hand-held probeware with realtime graphing technology, microscopes, filtering equipment, and a kiln for drying specimens. These were available prior to teachers' commitments to their research questions, which focused on various descriptive analyses of water and soil quality based on proximity to a water source, relative biodiversity in areas surrounding a particular location such as a tree or water source in a wooded area, and the presence of particular pollutants in standing water versus flowing water within the urban green space setting.

During the first week of the institute, teachers spent time planning their investigations, familiarizing themselves with the equipment, and collecting and analyzing data. After participants had collected and analyzed their data, additional half-day sessions prompted teachers to consider the links between the science concepts encountered in the field based inquiry projects and the state curriculum standards, and the relevance of Nature of Science principles including science as a human endeavor, the distinctions between observation and inference, the epistemological stance that scientific knowledge is subject to change, and science involving creativity and critical thinking (Lederman & Lederman, 2004). Participants' final products were presented in a conference-style poster session on the last day of the institute. The session took place after the posttest concept maps had been completed.

Concept map task design

The team of five faculty scientists served as the instructors and facilitators at the institute and also served as subject matter experts for the purpose of developing the concept map task. Prior to the institute, this team met with two educational researchers (and primary authors) and the lead science administrator for the school district to discuss the map prompts and create a list of reference concepts. During the first step of the concept map task design process, prompt questions were created to match the types of environments and scientific subject domains in which participants were anticipated to engage. This step resulted in a wetlands investigation prompt: "How do scientific concepts and processes interact within a wetlands system?" and an urban green space prompt: "How are Earth Science, Chemistry, Biology and Physics concepts inter-connected within an urban green space such as a campus setting?" The task was designed to allow participants to freely arrange concepts of their choosing on a large piece of paper, and draw lines between concepts to indicate links among them. The resulting prompt and instructions offered the opportunity to gauge changes in participants' knowledge relating to the instructional objective of an intervention (Novak & Cañas, 2004).

Each science faculty member generated a list of 10–15 concepts that represented scientific processes or phenomena pertinent to the prompt questions. The aggregate list was distributed and consensus was reached on the inclusion of 42 concepts in the list. The list was included in the instructions for map construction that were given to the participants (see Appendix A). The provision of a list of concepts with the stipulation that participants could add their own, provided a solution to the problem of artificially reducing inter- and intra-individual differences when the concepts that can be included in the map are constrained a-priori.

Procedure

Following informed consent procedures, concept maps were created on the first (pre-PD timepoint) and on the final (post PD timepoint)_days of the institute. Although participants

completed their investigations in small groups, maps were completed individually and were treated as individual work products. At each time point, participants were given a small packet of instructions along with large pieces of paper, colored markers, pencil, and a stack of multi-colored sticky notes.

Instructions included a one-page overview of concept mapping, two example concept maps on unrelated topics that exhibited alternative map structures of network and hierarchical tree and labeled links, the focus question prompt for the map, and a list of 42 concepts (see Appendix A). Participants were told that they could use some, all, or none of these concepts, and that they could add their own if they wished. Participants worked independently from one another and created their map on a large piece of paper. The same list of concepts was provided for each prompt. One hour was allotted for the mapping task. The map scoring rubric was not provided to participants.

Rubric development

The researchers and science faculty members engaged in a structured discussion to develop a rubric scoring method for the concept maps with dimensions that included evaluations of four units-of-analysis: concepts, links, clusters, and overall map. These dimensions were selected in order to capture the different facets of the underlying cognitive knowledge structure: Concepts, including concept number, centrality, and accuracy, pertaining to the overall degree to which the knowledge structure is complete and appropriate; Links, representing the degree to which knowledge is interconnected or fragmented; Clusters, representing the degree of organization of various portions of the knowledge structure; and Overall map quality, representing a total score for the holistic knowledge structure.

Concept centrality ratings

Since generating a single “expert” map against which to compare participants maps was deemed inappropriate for the current inquiry-based multi-disciplinary learning, we had the expert scientists independently rate their perceived centrality of each concept on the expert list to their particular domain of study, as it related to the prompts. The three point centrality scale ranged from “not central” to “highly central.” Ratings were aggregated and averaged across the experts, with a weighted rating for each concept on the list that included double weight for the earth/environmental science faculty members’ ratings, as this reflected the earth and environmental science emphasis in the overall PD. For example, whereas concepts such as water and oxygen/dissolved oxygen were rated as being central and were awarded centrality ratings of 3.00 and 2.06 respectively, concepts such as Newton’s laws and velocity were not weighted as being central and were awarded ratings of 0.93 and 0.48 respectively.¹ The concept list and centrality ratings, and their percent change in usage from pre-test to posttest, are provided in Appendix A.

Inter-rater reliability for concept centrality ratings

Inter-rater reliability was calculated using Fleiss Kappa (Fleiss, 1971) to provide the ratio of the proportion of agreement attainable above chance compared to the proportion of agreement

¹The ratings reflect the relative and particular importance of the concepts for the overall topic of cross-disciplinary study of a field setting. They do not necessarily reflect absolute or “correct” values.

actually achieved beyond chance. Kappa values range from ≤ 0.00 , indicating no agreement beyond what would be expected by chance, to 1.00, indicating perfect agreement. In accordance with Fleiss (1971), descriptors were assigned at intervals of 0.20 between 0.00 and 1.00. Fair agreement was obtained for 45% of the concepts, indicating between 21–40% agreement beyond chance. Moderate agreement was obtained for 38% of the concepts, indicating between 41–60% agreement beyond chance. Substantial agreement was obtained for 14% of the concepts, indicating between 61–80% beyond chance, and near perfect agreement was achieved for 6% of the concepts, indicating between 81–100% beyond chance.

Map scoring

The science experts received training on the rubric, and established inter-rater agreement through blind, independent co-scoring of 30% of the maps, and resolving discrepancies through a researcher-facilitated discussion. Then, two scientists blind scored the remaining 70% of the maps using the rubric, which combined structural, relational, and holistic methods and that assessed the concepts, links, and overall organization of each map. Elements within the rubric and a rationale for each element are presented in Table 1.

In addition to the individual rubric component scores, each map received three sub-scores which were based on structural features and one total score. Sub-scores corresponded to the three overall rubric categories of map attributes and correspond to areas of potential knowledge restructuring: Concepts, incorporating the degree to which participants included the provided terms, self-generated terms, the expert-rated centrality of the terms, and the cross-disciplinary scope of the knowledge structures; Links, which took into account both the number and the accuracy of the links; and Overall organization, which took into account the proportion of concepts and links that were contained in coherent clusters as well as the structure (including the hierarchy) of the map. The calculation of scores aimed to reflect dimensions that the research team deemed important to capture knowledge facets central to the particular interdisciplinary domain studied by the teachers. These dimensions are reflected in the formulae for calculation of sub-scores for each of the three knowledge facets as described below.

Structural feature 1: concepts

The concepts sub-score was created using a formula (see scoring column in Table 1). It took into account ratings of the maps' scientific scope from narrow (one discipline) to broad (all four disciplines of Earth/Environmental Science, Biology, Chemistry and Physics), the sum of the centrality scores for all concepts included in the map, the number of concepts drawn from the provided list, and the number of valid concepts added by the participant. These dimensions aimed to reflect the emphasis on the interdisciplinary nature of the knowledge structure – the broader the number and scope of the concepts, the higher the sub-score.

The degree to which the map included highly central terms from the expert-generated list was also calculated in regard to the overall number of concepts chosen by multiplying the number of concepts from the list by a centrality formula. This provided a weighted estimate of the degree to which the participant had selected central concepts from the list without artificially inflating scores due to the selection of many concepts. The concepts sub-score took the resulting product of the number of concepts from the list multiplied by

Table 1. Concept map scoring rubric.

Structural Feature 1: Concepts			
Critical Element	Description	Scoring	Rationale
Scope	The number of scientific disciplines judged to be represented on the map (Biology, Chemistry, Physics, and Earth Science).	1–4	Cross-disciplinary PD content; inclusion of common science principles with credit for individualized prior knowledge and field study content
# of concepts from the list	The number of concepts included in the map that were derived from the expert-generated list.	0–42	
Centrality	The sum of the centrality ratings of each concept used in the map.	0–126	
Concept Centrality Formula	The centrality formula provides an index of the ratio of centrality to selected concepts.	Centrality # concepts from list	
# valid additional concepts	The number of concepts included in the map that were participant-generated.	1 point per scientifically valid concept	
Concepts sub-score	The sub-total score reflecting the concepts portion of the rubric.	# concepts from list + # valid additional concepts Total number of concepts on the list	
Structural Feature 2: Links			
Critical Element	Description	Scoring	Rationale
Type of Link	Links were given incremental points for their position relative to indicating a process or component within a broader system.	Geographical proximity = 1 Equality = 2 Inclusion (part-whole) = 3 Process or system = 4	Scientific accuracy in the construction of relations among concepts; aggregate scoring maintains a penalty for inaccuracies.
Scientific accuracy	The number of accurate and inaccurate links were tallied	# Inaccurate links	
# Labeled	The number of labeled links was tallied	# Accurate links # Labeled links	
# Unlabeled	The number of unlabeled links was tallied	#Unlabeled links	
Links weighted formula	The links weighted formula provides an index that accounts for the relative number of inaccurate links.	#Accurate links – # Inaccurate links	
Links sub-score	The sub-total score reflecting the links portion of the rubric.	(# Labeled links * Type of link) + Links weighted formula	
Structural Feature 3: Organization			
Critical Element	Description	Scoring	Rationale
Degree of Clustering	The proportion of the concepts within the map that are contained within groups of 3 or more concepts that are linked together	None = 0 < 1/3 are clustered = 1 < 2/3 are clustered = 2 All of the concepts are clustered = 3	Clustering reflects propositional and thematic organization of concepts; Scoring incrementally gives credit for sophisticated organization of knowledge structure
Accuracy of clustering	The proportion of the clusters that are rated as scientifically accurate	None = 0 < 1/3 are accurate = 1 < 2/3 are accurate = 2 All of the clusters are accurate = 3	
Organizing Structure	The overall organizational structure of the map was classified dichotomously and exclusively.	Relational list/ hierarchy = 1 Process = 2 System = 3	

(Continued)

Table 1. (Continued).

Organization sub-score	The sub-total score reflecting the proportion of concepts within clusters, the accuracy of the clusters, and the overall structural sophistication of the map.	(Degree of clustering * Accuracy of clustering) + Organizing structure	
Context	The location of the map was noted, if represented.	Wetlands Field/campus Neither	Fidelity check; reflects appropriate interpretation of task

the centrality formula, and added the number of valid additional concepts provided by the participant plus the rating of the cross-disciplinary scope of the map (a scale of 1–4). This resulted with a score that reflected scope and centrality of the concepts in the map. Concept scores ranged from 4.26 to 25.88.

Structural feature 2: links

Links were scored in ways that reflected the propositions that could be constructed from each map portion (Novak & Cañas, 2004). Although links within concept maps are permitted to be labeled or unlabeled (Nesbit & Adesope, 2006), our approach gave credit to participants who were able to articulate particular types of links (Novak & Cañas, 2004) and therefore display an awareness of informational structure and hierarchy (Greene et al., 2013). Link types were determined by the scientists and took into account the various types of links that were plausible given the prompt (i.e. geographical proximity, equality, inclusion or part-whole relationships, and processes or causal attributes). The number of accurate and inaccurate links was tallied separately. The links sub-score included a weighted sub-total for links that allocated more points to more precise links that demonstrated knowledge of an overall ecological system (see Table 1). A one-point penalty loss was also included for each inaccurate link. These decisions aimed to create an ordinal scale that reflected the richness and accuracy of the links in the map. Link scores ranged from 19.00 to 178.50.

Structural feature 3: organization

The organization sub-score was generated by combining scores for the proportion of concepts included within clusters, the accuracy of those clusters, and the overall structure score for the map. A cluster was defined as a group of at least three concepts that were linked together. Scientists determined the accuracy of the clusters and a cluster accuracy formula was calculated by dividing the number of accurate clusters by the total number of clusters. This was then multiplied by the overall proportion of concepts in clusters relative to the proportion of concepts that were not encased within a linked cluster. This ensured that scores were not artificially inflated by a highly clustered but inaccurate map while taking into account the degree to which the overall map was organized using clusters. A holistic structure score was also given such that maps with minimal evidence of a hierarchical structure scored lower (1 point) than maps that depicted processes (2 points). These in turn scored lower than maps that portrayed a system (3 points). This scoring scheme echoes cognitive and constructivist theories of learning from which concept map methodologies have been derived, as it follows the maturation pattern

often observed during the development of content area expertise (Bransford et al., 2000; Chase & Simon, 1973). Organization scores ranged from 3.00 to 9.00

Total score

A total score was calculated via the sum of the concepts, links, and organization sub-scores. This aimed to create an ordinal score that incorporated the three facets of knowledge structure. Total scores ranged from 32.02 to 194.91.

Figure 1 presents an example of two scored maps – one at pretest and one at posttest.

Results

Preliminary analyses revealed no main effect of map prompt. Data therefore reflect analyses of the sample of 21 participants in aggregate. Prior to examining the findings in relation to each research question, we first present a brief overview of descriptive statistics for the individual rubric components. Some of the indices showed marked change, while others remained relatively stable. Statistics are shown in Table 2.

Overall map concept breadth, concept centrality, and concept usage did not change between the two time points of pre-PD and post PD, but participants did include an increased number of accurate links and more higher order (part-whole, process relations) at posttest.

Research question 1. which features of concept map structure changed over time from pretest to posttest?

As a visual examination of the distribution of scores suggested approximation to normality, we conducted variance based comparison tests between the pre- and post-PD distributions. A main effect was found for time point (pre- PD, post PD) on map total score, $F_{(1,19)} = 29.73$, $p < .000$, $\eta_p^2 = 0.61$. Further analyses were then conducted to investigate any changes in the three aggregate sub-scores of concepts, links, and organization. A General Linear Model Analysis of Variance was conducted with time point as a within subjects factor and the aggregate sub-scores for concepts, links and organization as dependent measures. This analyses yielded a significant effect of time point for links only, $F_{(1,19)} = 34.24$, $p < .001$, $\eta_p^2 = 0.64$. Descriptive statistics are shown in Table 3.

Link type

Within an administration time point (i.e. at pre-PD, or at post PD), the four types of links sub-scores were not significantly correlated with one another. A repeated measures MANOVA to assess the degree of change between the pre-PD and post PD maps revealed a main effect of time, $F_{(4,17)} = 10.49$, $p < .000$, $\eta_p^2 = 0.71$. Univariate tests revealed that two of the four types of links increased from pre- to post PD. Participants' maps were found to have significantly more links indicating an inclusion or part-whole relationship, $F(1,20) = 6.41$, $p = .02$, $\eta_p^2 = 0.24$, and links indicating a process relationship, $F(1,20) = 5.18$, $p = .03$, $\eta_p^2 = 0.20$. Changes in the number of links representing geographical proximity or equality among concepts were minimal and not statistically significant.

Table 2. Descriptive statistics for formula components.

	Pre-test		Posttest	
	Mean	SD	Mean	SD
Breadth of map scope	3.14	0.96	3.14	0.96
Sum of concept centrality	16.86	6.11	16.72	6.32
Number of provided concepts used	14.43	6.34	13.90	6.31
Number of accurate links	12.24	8.18	20.00	10.34
Number of inaccurate links	1.05	1.35	0.33	1.11
Links: Geographical proximity	0.52	2.18	0.71	1.71
Links: Equality	4.00	5.86	4.19	4.85
Links: Inclusion	3.33	3.02	7.14	8.33
Links: Process	6.10	3.99	9.33	6.40
Proportion of concepts within a cluster	2.71	0.56	3.00	0.00

Table 3. Summary statistics for map features sub-scores and total scores.

	Pre-test		Posttest	
	Mean	SD	Mean	SD
Concepts sub-score	72.75	28.71	106.08	36.63
Links sub-score	56.19	26.16	89.21*	35.98
Organization sub-score	5.87	1.78	6.26	1.81
Total score	72.75	28.71	106.08*	36.33

* $p < 0.000$

Research question 2. what changes were evident in the scientific accuracy of participants’ maps from pretest to posttest?

Rubric components designed to assess the number of scientifically valid added concepts, the number of scientifically accurate links, and the number of scientifically accurate clusters of concepts and links, were included in a GLM Analysis of Variance with time point (pretest, posttest) as a within subjects factor and the three scientific accuracy based sub-scores as measures. All three scores pertaining to scientific accuracy improved significantly from pretest to posttest (see Table 4): for scientifically valid concepts, $F_{(1,19)} = 12.81, p = .002, \eta_p^2 = 0.39$; for scientifically accurate links, $F_{(1,19)} = 15.35, p = .001, \eta_p^2 = 0.43$; for scientifically accurate clusters, $F_{(1,19)} = 42.48, p < .001, \eta_p^2 = 0.68$.

Participants included significantly more self-generated concepts in their posttest maps compared to their pretest maps. The mean number of added accurate concepts doubled from 5.57 to 10.91. Participants also included significantly more scientifically accurate concepts and more scientifically accurate links between the concepts. In the case of accurate links, the mean score increased nearly 70% from 12.24 accurate links to 20.00 accurate links. A nearly four-fold increase was noted in the number of scientifically accurate clusters of concepts from pretest to posttest. The mean at pretest was one accurate cluster but this increased to a mean of

Table 4. Summary statistics for scientific accuracy sub-scores.

	Pre-test		Posttest	
	Mean	SD	Mean	SD
Number of scientifically accurate additional concepts	5.57	5.38	10.91**	6.94
Number of scientifically accurate concepts	20.00	6.37	24.81*	8.31
Number of scientifically accurate links	12.24	8.18	20.00**	10.34
Number of scientifically accurate clusters	0.97	0.10	4.14**	2.28

* $p < 0.05, **p < 0.000$

four accurate clusters at posttest. In sum, the analysis suggested improvements in various facets of the scientific accuracy of participants' knowledge across all three structural areas of the rubric: concepts, links between concepts, and map organization.

Research question 3. how did participants' use of the expert- and self-generated concepts change from pretest to posttest?

We were interested in whether participants used the provided concepts differently at the beginning versus the end of the institute. In order to understand whether teachers used more concepts or more central concepts at the beginning versus the end of the institute, we compared the number of concepts and the centrality scores at pretest and posttest using repeated measures ANOVA.

As a group, participants did not show significant changes in the number of listed concepts that were included in the maps. At pretest, participants selected an average of 14 (33%) of 42 concepts ($M = 14.42$, $SD = 6.33$) with an average of 13 (31%) of the 42 concepts ($M = 13.90$, $SD = 13.90$) selected at posttest. In contrast, participants included almost twice the number of their own (non-listed) concepts to their maps at posttest ($M = 10.90$, $SD = 6.94$) compared to pretest ($M = 5.58$, $SD = 5.38$). The increase in valid additional concepts was found to be statistically significant, $F_{(1,20)} = 12.81$, $p = .02$, $\eta_p^2 = 0.39$.

In order to investigate whether the cohort changed in the likelihood of selecting important disciplinary concepts, we also examined the mean centrality values of the selected concepts over time. We found that participants selected concepts with a mean total centrality rating of 16.86 ($SD = 6.11$) at pretest and 16.72, ($SD = 6.32$) at posttest. This slight change was not statistically significant. On average, maps represented 20% of the total possible centrality at pre- and posttest, meaning that changes in overall total centrality ratings were not evident.

Relatedly, we were interested in the degree to which the whole group of participants was able to select concepts with high centrality ratings. We examined the correlations between the centrality rating of each concept and the percentage of the participants that selected the term. These correlations were positive and linear and stable over time points: at pre-PD, $r_{(42)} = 0.64$, $p < .001$ and at post PD, $r_{(42)} = 0.68$, $p < .001$. Correlations among the 42 listed concepts and the percentages of participants that selected each concept at pre- and posttest revealed that the percentage of participants that selected particular concepts was stable over time between pretest and posttest map creation, $r_{(42)} = 0.88$, $p < .001$. This demonstrates a positive association between concepts' centrality ratings and the relative proportion of the group that selected those concepts. Stability over time points suggests that participants were already familiar with key concepts at the beginning of the institute and were already able to incorporate important concepts into their maps. However, positive changes were found in the overall organization and interrelations of these concepts.

Discussion

This study addressed the question of how to create an authentic assessment of changes in teachers' content knowledge during their pursuit of individualized, open-ended inquiry based professional learning, while remaining sensitive to variations in prior knowledge and respectful of teachers' time in the PD setting. Specifically, we used concept-mapping, and explored the

feasibility of prompting and scoring maps among teachers who possessed varying knowledge at the onset of the PD experience and who participated in independently executed but thematically linked field-based inquiry. We selected concept maps as a method sympathetic to capturing learning in settings in which teachers, as learners, experience reform-oriented science teaching. The study is unusual in its use of direct measures of content knowledge rather than self-report measures, and in its approach of matching the cognitive-constructivist nature of the teachers' field-based scientific inquiry with the method of learner-generated products as assessments of learning. It was sympathetic to the idiosyncratic content of participants' learning and participants' prior knowledge, as was the case in the study of algebra learning by teachers (Hough et al., 2007). In our case, the variation in prior disciplinary expertise among the cohort of teachers and teacher-designed nature of the inquiry also required the design of an assessment protocol that could prompt and evaluate maps containing cross-disciplinary ideas. The study is therefore also atypical in that the scope of the PD and the scope of the assessment used to capture teachers' knowledge were both deliberately cross-disciplinary in nature. It builds on previous studies through the deliberate inclusion of multiple subject matter experts whose conceptual knowledge and emphases were sought throughout the rubric design process. It therefore also represents a brief example of how diversity of perspectives can be leveraged and brought to bear coherently within the context of cross-disciplinary science PD.

While others have found that descriptive, field based studies produced changes in teachers' self-efficacy for teaching environmental literacy content (Trauth-Nare, 2015), our methodology allowed us to detect changes in the structure and organization of teachers' knowledge representations as a result of studying watersheds and urban greenspaces. Teachers increased their capacity to link concepts in meaningful ways as well as the ability to add in scientifically correct concepts, including their own novel but scientifically correct concepts. This suggests that overall, the inquiry-based field studies that were completed during the institute did impact teachers' representations of science concepts and their relations and did not lead to the creation of many new misconceptions or inaccuracies among participants. Those who design, implement and evaluate science teacher professional development programs might find a rubric-driven concept mapping approach to be a useful lens, since it equipped the researchers with multiple ways to consider changes in participants' knowledge.

Specifically, we found a significant increase from pretest to posttest in the ways in which concepts were linked to one another. At posttest, teachers made greater use of links that represented a more sophisticated understanding of the relations among concepts, such as part-whole and process-oriented links. Accordingly, a small increase was noted in the likelihood that concepts would be contained within meaningful groups or clusters and, at posttest, all of the participants included all of their concepts within a meaningful cluster. These changes seem to reflect teachers' improved understanding about the commonalities, differences, and causal relations among scientific concepts, scientific processes, and the scientific disciplines with which they may be most typically associated. Our findings suggest that although their capacity to identify or recognize the meaning of concepts did not shift to a large extent, teachers' ability to create connections between concepts did improve over time points from pretest to posttest. This is in keeping with Greene et al.'s (2013) findings that science teachers' concept maps shifted from including linear chains of concepts to more densely connected concepts after conducting scientific inquiry projects as a form of professional development. This suggests that, in parallel with contemporary notions of meaningful learning for students, interventions to improve science teachers' knowledge might create contexts for participants to

consider the connections between otherwise disparate areas of the curriculum. In this case, concept maps were useful as an assessment tool as they were sensitive to changes in the organization and connectivity of participants' knowledge, and not just the number or diversity of concepts that were correctly identified.

The second and third research questions focused on the degree to which map changes reflected improvements in scientific accuracy and the ways in which participants incorporated the expert-generated lists into their maps. Both questions add to our understanding of the degree to which changes in knowledge reflected a desirable increase in scientific understanding. Whereas teachers in our study included an average of 20 scientifically accurate concepts at pretest and 24 at posttest, the number of scientifically accurate, self-generated concepts increased from approximately 25% self-generated concepts at pre-test to 45% at posttest. Teachers in our study did not necessarily produce more elaborate maps with many more concepts; however, changes were seen in the arrangement and connections among these concepts. This is perhaps reflective of participants' attempts to connect new with existing concepts and provides some evidence of teachers' assimilation of new information with prior knowledge. Future research might explore whether and how knowledge assimilation that takes place in PD setting infuses into instruction upon the teacher's return to the classroom. It may be that under particular circumstances making cross-curricular knowledge connections may be more or less beneficial. How cross-curricular knowledge manifests itself in the classroom may depend on both the teacher's overall level of pedagogical knowledge and beliefs about teaching within a particular subject area, as well as the students' learning needs. Regardless, this study offers a first step towards understanding how concepts can be considered by teachers while they are engaged in cross-curricular, inquiry-based professional development, and finds that even though teachers implemented a variety of inquiry-based science projects, their overall learning of concepts could be successfully captured using group level indices.

In comparing our findings with other studies that have also used concept maps to measure teachers' learning, it is worth considering how varying approaches to concept mapping such as task development, execution, and scoring may influence the outcomes (Nesbit & Adesope, 2006). For example, participants in the study by Greene et al. (2013) demonstrated significant improvements in the number of nodes (concepts) within their maps from pretest to posttest. On average, teachers in that study nearly doubled the number of included concepts from 13 at pre-test to 25 at posttest. However, Greene and colleagues did not provide a reference list of concepts to participants and used an arguably broader, topic prompt instead of a guiding question. In addition, whereas those researchers elected to give a 1–3 rating to reflect the accuracy of propositions on each map, our strategy was to compare concept sub-scores, and thereby take into account the proportion of included concepts designated as scientifically accurate while including a penalty for inaccuracies. In agreement with Hough et al. (2007), who found an increase in clustering after a content-based intervention, our analyses revealed increased scientific accuracy as measured by overall map organization of which clustering was one component. Together, our study, and its predecessors that have examined teacher learning within a PD context in which the content is not predetermined at the outset, show that small scale, independent inquiry can have a beneficial effect on teachers' content knowledge within a relatively short amount of time, and that changes can be revealed in a relatively unobtrusive manner through the implementation of concept mapping.

Limitations

This study supports insights from a small number of previously published studies that have examined the feasibility and effect of using concept mapping as a strategy for assessing changes in teachers' scientific content knowledge. However, it is worth noting that several characteristics of the study do limit its generalizability. Perhaps the most pressing limitation is that the rubric, although carefully developed and based conceptually on existing theory and research, was developed for this particular study and was not previously tested or validated. The subject matter experts did not create an expert map nor did they assess teachers using standardized measures. We cannot be sure, therefore, that the inclusion of a concept within a given map equated to an individual's full comprehension of the concept and its role in answering the overall question. Unfortunately, due to resource limitations, the researchers were unable to investigate this further by, for example, gathering artifacts from the teachers' science projects to triangulate the content of the investigations with the content of their concept maps.

Similarly, although all of the participants in our study were able to produce a map, it is possible that teachers' initial maps may have under-represented their knowledge due to the novelty or perceived difficulty of the mapping task. If the study were to be replicated, prior instruction in concept maps may be necessary to ensure that all participants are equally aware of the task at hand and do not otherwise compromise the depiction of their prior knowledge. A final limitation to note is the variation in inter-rater reliability in centrality ratings for a portion of the concepts. The scientists demonstrated substantial agreement for the rating of centrality of some but not all concepts, perhaps due to the fact that they represented varying disciplines. When designing cross-disciplinary PD work in the future, more extensive discussions ahead of mapping task development might improve scientists' agreement regarding the relative centrality of particular concepts.

Implications

Our findings are in agreement with calls for reconceptualizing theories of teacher learning in order to emphasize the role of prior knowledge and constructivist approaches to knowledge growth, and to generate examples of how such approaches might align the form of the PD – for example, the use of cross-disciplinary field-based inquiry – with its application to the classroom (Luft & Hewson, 2014). Just as studying the natural world has been shown to provide students with an authentic context for science learning in meaningful, relevant, and interdisciplinary ways (Czerniak & Johnson, 2014), teachers' learning may also be deepened and perceived as more meaningful when relationships among ideas are considered and when new knowledge becomes integrated with prior knowledge (Tsai & Huang, 2002). The format of the PD and its assessment through concept maps could support a variety of applications within and across the curriculum, such as the study of biotic and abiotic factors in a local pond, requiring the novel application of existing knowledge of both the physical and the chemical properties of water, a schoolyard investigation of the effects of runoff, requiring the application of knowledge of soil layers and properties, or inquiry into geography and ecology topics such as water resource management. The present study offers insights into ways in which PD might be designed to accommodate diverse teachers' interests and needs without compromising the opportunity to gather common outcomes.

The concept mapping strategy helped reveal the nature of teachers' learning in ways that both support and augment existing literature on teacher learning in professional development settings. For example, our findings suggest that teachers entered the PD context with relevant prior knowledge. Based on the changes that we observed from the first to the second map creation exercise, however, it seems that professional development provided a context for teachers to not only increase the complexity of their knowledge structures as evidenced by the increase in the number of scientifically accurate concepts used at posttest, but also to fine-tune the relations among concepts such that specific link types could be articulated. Future research might explore the impact of combining inquiry-based science teacher professional development with concept mapping assessment methods in order to compare teachers' and students' knowledge change as PD experiences are translated into the classroom. Studying the impact of PD and classroom implementation may open up interesting questions about how students' knowledge representations might be paralleled by those of their teachers, a phenomena that has been previously demonstrated in studies of expertise development (Boshuizen, Bromme, & Gruber, 2004). To our knowledge, cross-disciplinary, environmentally focused research on teacher and student learning has not considered explicitly knowledge reconstruction (Gilman, Hitt, & Gilman, 2015; Trauth-Nare, 2015; Wojdak et al., 2010). With further validation, concept maps might be an appropriate tool for capturing authentic learning that cannot be captured by traditional measures such as self-report or attitude surveys.

We see concept mapping as a useful tool to measure sophisticated and interdisciplinary teacher content knowledge as it can demonstrate the nuanced learning present in individuals as well as provide an overall picture of content gains associated with the PD. Providers looking to demonstrate teacher learning might seek out concept mapping in addition to self-reported or content-based pretests and posttests to provide a more wholistic view of gains associated with the PD. This approach might be particularly suited to cross-disciplinary work that requires input from multiple content area experts in order to account for and leverage variations in teachers' prior knowledge, when teachers can determine the course of their own learning, or when the topic of study varies but the location or setting of the work, and the focus on inquiry, remains consistent (Capps et al., 2012; Wojdak et al., 2010).

By extension, it may be helpful to think of concept mapping in terms of formative assessment. One goal of formative assessment is to consider individual learning, including specific ideas and the progressions made by learners as they go through a unit of study (Bell & Cowie, 2001; Coffey, Hammer, Levin, & Grant, 2011). Although in this study we collected maps at the beginning and end of the PD, the use of concept mapping at various points during an extended period of professional learning and an analysis of the relative changes within each individual's maps, might prove valuable in tracing the emergence of increasingly sophisticated understandings as teachers gradually reorganize their knowledge (Vosniadou & Ioannides, 1998).

Finally, we propose that the rubric development process we used is sufficiently flexible to be adapted for use with content from other disciplines. The steps and the resulting rubric provided here might serve as a starting point for other studies in which cross-disciplinary knowledge is focal to the PD aims. This may be of benefit for researchers, facilitators, evaluators, and PD participants, all of whom have a stake in the outcome of teachers' professional learning. Our process leveraged interaction and consensus generation among subject matter expert scientists who worked alongside educational researchers

with expertise in theories of learning and assessment. Key steps included the generation of participant instructions including prompt questions and critical scientific terms, expert knowledge extraction to generate and then implement various rubric components, the establishment of inter-rater reliability among experts for the purposes of map scoring, and analyzing and interpreting the map data. Although effortful, the results of this process included a rigorous and unintrusive method for examining participants' knowledge.

Finally, although anecdotal, it is worth noting that the process was adopted by participating science faculty members and several of the teachers. The selection of concept mapping as an assessment tool seemed to serve as a model strategy for breaking down traditional barriers between project implementation, research on teachers' learning, and the evaluation of a teacher professional development program.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Virginia Department of Education [VA120308].

ORCID

Joanna K. Garner  <http://orcid.org/0000-0001-9105-5194>

Avi Kaplan  <http://orcid.org/0000-0002-2898-0085>

Stephanie Hathcock  <http://orcid.org/0000-0001-9116-4753>

Bradley Bergey  <http://orcid.org/0000-0002-0495-7219>

References

- Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 1105–1149). Mahwah, NJ: Lawrence Earlbaum Associates, Inc.
- Achieve. (2013). *The next generation science standards*. Retrieved from www.nextgenscience.org
- Alexander, P. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, 32(8), 10–14. doi:10.3102/0013189X032008010
- Alonzo, A., Kobarg, M., & Seidel, T. (2012). Pedagogical content knowledge as reflected in teacher-student interactions: Analysis of two video cases. *Journal of Research in Science Teaching*, 49(10), 1211–1239. doi:10.1002/tea.v49.10
- Anderson, J. R. (2000). *Cognitive psychology and its implications* (5th ed.). New York, NY: Worth Publishers.
- Austin, B., & Schmidt, N. (2010). Pedagogy, environmental education, and context: Promoting knowledge through concept mapping. In A. M. Bodzin, B. S. Klein, & S. Weaver (Eds.), *The inclusion of environmental education in science teacher education* (pp. 235–254). New York, NY: Springer Science and Business Media.
- Ausubel, D. P. (1968). *Educational psychology. A cognitive view*. New York: Holt, Rinehart and Winston, Inc.
- Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, 85, 536–553. doi:10.1002/(ISSN)1098-237X

- Besterfield-Sacre, M., Gerchak, K., Lyons, M., Shuman, L. J., & Wolfe, H. (2004). Scoring concept maps: An integrated rubric for assessing engineering education. *Journal of Engineering Education*, 93, 105–115. doi:10.1002/j.2168-9830.2004.tb00795.x
- Bodzin, A. M., Klein, B. S., & Weaver, S. (2010). *The inclusion of environmental education in science teacher education*. New York, NY: Springer.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3–15. doi:10.3102/0013189X033008003
- Boshuizen, H. P. A., Bromme, R., & Gruber, H. (2004). *Professional learning: Gaps and transitions on the way from novice to expert*. New York, NY: Kluwer Academic Publishers.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience and school*. Washington, DC: National Academy Press.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2004). The design of learning environments. In J. D. Bransford, A. L. Brown, & R. R. Cocking (Eds.), *How people learn: Brain, mind, experience, and school* (pp. 131–154). New York, NY: National Academy Press.
- Brown, S., & MacIntyre, D. (1993). *Making sense of teaching*. Buckingham, UK and Philadelphia, PA: Open University Press.
- Capps, D. K., Crawford, B. A., & Conostas, M. A. (2012). A review of empirical literature on inquiry professional development: Alignment with best practices and a critique of the findings. *Journal of Science Teacher Education*, 23, 291–318. doi:10.1007/s10972-012-9275-2
- Chang, K.-E., Sung, Y.-T., Chang, R.-B., & Lin, S.-C. (2005). A new assessment for computer-based concept mapping. *Educational Technology and Society*, 8(3), 138–148.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55–81. doi:10.1016/0010-0285(73)90004-2
- Chesapeake Bay Foundation. (2012). *Environmental literacy plans by state*. Retrieved from <http://bb.cbf.org/page.aspx?pid=924>
- Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Coffey, J. E., Hammer, D., Levin, D. M., & Grant, T. (2011). The missing disciplinary substance of formative assessment. *Journal of Research in Science Teaching*, 48, 1109–1136. doi:10.1002/tea.20440
- Crawford, B. A. (2000). Embracing the essence of inquiry: New roles for science teachers. *Journal of Research in Science Teaching* 37, 916–937. doi:10.1002/1098-2736(200011)37:9<916::AID-TEA4>3.0.CO;2-2
- Crawford, B. A. (2007). Learning to teach science in the rough and tumble of practice. *Journal of Research on Science Teaching*, 44(4), 613–642. doi:10.1002/tea.20157
- Czerniak, C. M., & Johnson, C. C. (2014). Interdisciplinary science teaching. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research on science teaching* (Vol. II, pp. 395–411). New York, NY: Routledge.
- Daley, B. J., Shaw, C. R., Balistrieri, T., Glasenapp, K., & Piacentine, L. (1999). Concept maps: A strategy to teach and evaluate critical thinking. *The Journal of Nursing Education*, 38(1), 42–47.
- Dansereau, D. F., & Holley, C. D. (1982). Development and evaluation of a text mapping strategy. In A. Flammer & W. Kintsch (Eds.), *Discourse processing* (pp. 537–554). Amsterdam, NL: North-Holland.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38, 181–199. doi:10.3102/0013189X08331140
- Diamond, B. S., Maerten-Rivera, J., Rohrer, R. E., & Lee, O. (2014). Effectiveness of a curricular and professional development intervention at improving elementary teachers' science content knowledge and student achievement outcomes: Year 1 Results. *Journal of Research in Science Teaching*, 51(5), 635–658. doi:10.1002/tea.v51.5
- Dresner, M., & Worley, E. (2006). Teacher research experiences, partnerships with students, and teacher networks sustaining factors from professional development. *Journal of Science Teacher Education*, 17, 1–14. doi:10.1007/s10972-005-9000-5

- Enderle, P., Denzau, M., Roseler, K., Southerland, S. A., Granger, E., & Hughes, R. (2014). Examining the influence of RET's on science teachers' beliefs and practice. *Science Education*, 98, 1077–1108. doi:10.1002/sc.21127
- Fishman, B., Marx, R., Best, S., & Tal, R. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education*, 19(6), 643–658. doi:10.1016/S0742-051X(03)00059-3
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382. doi:10.1037/h0031619
- Gess-Newsome, J., Taylor, J. A., Carlson, J., Gardner, A., Wilson, C. D., & Stuhlsatz, M. A. M. (2017). Teacher pedagogical content knowledge, practice, and student achievement. *International Journal of Science Education*, 1–20. doi:10.1080/09500693.2016.1265158
- Gilman, S. L., Hitt, A. M., & Gilman, C. (2015). Training master's-level graduate students to use inquiry instruction to teach middle-level and high-school science concepts. *School Science and Mathematics*, 115(4), 155–167. doi:10.1111/ssm.2015.115.issue-4
- Greene, B. A., Lubin, I. A., Slater, J. L., & Walden, S. E. (2013). Mapping changes in science teachers' content knowledge: Concept maps and authentic professional development. *Journal of Science Education and Technology*, 22, 287–299. doi:10.1007/s10956-012-9393-9
- Hayhoe, D., Bullock, S., & Hayhoe, S. (2011). A kaleidoscope of understanding: Comparing real with random data, using binary choice items, to study preservice elementary teachers' knowledge of climate change. *Weather Climate and Society*, 3(4), 254–260. doi:10.1175/WCAS-D-11-00021.1
- Herl, H. E., Baker, E. L., & Niemi, D. (1996). Construct validation of an approach to modeling cognitive structure of U.S. history knowledge. *The Journal of Educational Research*, 89, 206–218. doi:10.1080/00220671.1996.9941206
- Holden, M., Groulx, J., Bloom, M., & Weinburgh, M. H. (2011). Assessing efficacy through an outdoor professional development experience for in-service science teachers. *Electronic Journal of Science Education*, 12(2), 1–25.
- Hough, S., O'Rode, N., Terman, N., & Weissglass, J. (2007). Using concept maps to assess change in teachers' understandings of algebra: A respectful approach. *Journal of Mathematics Teacher Education*, 10(1), 23–41. doi:10.1007/s10857-007-9025-0
- Jacobs-Lawson, J. M., & Hershey, D. A. (2002). Concept maps as an assessment tool in psychology courses. *Teaching of Psychology*, 29(1), 25–29. doi:10.1207/S15328023TOP2901_06
- Jin, H., & Wong, K. Y. (2014). Training on concept mapping skills in geometry. *Journal of Mathematics Education*, 3(1), 103–118.
- Kirschner, S., Borowski, A., & Fischer, H. E. (2011). *Physics teachers' content knowledge and pedagogical content knowledge: Developing test scales and measuring the relation*. Paper presented at the Annual Meeting of NARST, Orlando, FL.
- Lederman, N. G., & Lederman, J. S. (2004). Revising instruction to teach Nature of Science. *Science Teacher*, 71(9), 36–39.
- Luft, J. A., & Hewson, P. W. (2014). Research on teacher professional development programs in science. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (2nd ed., pp. 889–909). New York, NY: Routledge.
- Mäntylä, T., & Nousiainen, M. (2013). Consolidating pre-service physics teachers' subject matter knowledge using didactical reconstructions. *Science & Education*, 23, 1583–1604. doi:10.1007/s11191-013-9657-7
- McClure, J. R., Sonak, B., & Suen, H. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, 36(4), 475–492. doi:10.1002/(ISSN)1098-2736
- Moyer-Packenham, P. S., Bolyard, J. J., Oh, H., & Cerar, N. I. (2011). Common features of professional development activities for mathematics and science teachers. *Professional Development in Education*, 37(4), 571–589. doi:10.1080/19415257.2010.531597
- Mulder, Y. G., Lazonder, A. W., & De Jong, T. (2015). Key characteristics of successful science learning: The promise of learning by modelling. *Journal of Science Education and Technology*, 24, 168–177. doi:10.1007/s10956-014-9537-1

- Nagle, B. (2013). Preparing high school students for the interdisciplinary nature of modern biology. *CBE Life Sciences Education*, 12, 144–147. doi:10.1187/cbe.13-03-0047
- National Research Council. (1996). *National science education standards*. Washington, DC: The National Academies Press. doi:10.17226/4962
- Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, 76(3), 413–448. doi:10.3102/00346543076003413
- North American Association for Environmental Education (NAAEE). (2012). *Environmental education materials: Guidelines for excellence*. Washington, DC: Author.
- Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching*, 27, 937–949. doi:10.1002/tea.v27.10
- Novak, J. D. (2005). Results and implications of a 12-year longitudinal study of science concept learning. *Research in Science Education*, 35, 23–40. doi:10.1007/s11165-004-3431-4
- Novak, J. D., & Cañas, A. J. (2004) *The theory underlying concept maps and how to construct and use them*. Technical Report IHMC CmapTools 2006-01 Rev 01-2008. Retrieved from <http://cmap.ihmc.us/docs/theory-of-concept-maps>
- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. New York, NY: Cambridge University Press.
- Rennie, L., Venville, G., & Wallace, J. (2012). *Integrating science, technology, engineering, and mathematics: Issues, reflections, and ways forward*. New York, NY: Routledge.
- Rollnick, M., Bennett, J., Rhemtula, M., & Dharsey, N. (2008). The place of subject matter knowledge in pedagogical content knowledge: A case study of South African teachers teaching the amount of substance and chemical equilibrium. *International Journal of Science Education*, 30(10), 1365–1387. doi:10.1080/09500690802187025
- Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching*, 33, 569–600. doi:10.1002/(ISSN)1098-2736
- Rumelhart, D. E., Lindsay, P. H., & Norman, D. A. (1972). A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 197–246). Oxford, England: Academic Press.
- Russell, S. H., & Hancock, M. P. (2007). *Evaluation of the research experiences for teachers (RET) program: 2001–2006*. Final Report from SRI International. Menlo Park, CA: SRI International.
- Rye, J., Landenberger, R., & Warner, T. A. (2012). Incorporating concept mapping in project-based learning: Lessons from watershed investigations. *Journal of Science Education Technology*, 22, 379–392. doi:10.1007/s10956-012-9400-1
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. doi:10.3102/0013189X015002004
- Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching*, 37(9), 963–980. doi:10.1002/(ISSN)1098-2736
- Trauth-Nare, A. (2015). Influence of an intensive, field-based life science course on preservice teachers' self-efficacy for environmental science teaching. *Journal of Science Teacher Education*, 26, 497–519. doi:10.1007/s10972-015-9434-3
- Tsai, C. C., Lin, S. S. J., & Yuan, S. M. (2001). Students' use of web-based concept map testing and strategies for learning. *Journal of Computer Assisted Learning*, 17(1), 72–84. doi:10.1046/j.1365-2729.2001.00160.x
- Tsai, C.-C., & Huang, C.-M. (2002). Exploring students' cognitive structures in learning science: A review of relevant methods. *Journal of Biological Education*, 36(4), 163–169. doi:10.1080/00219266.2002.9655827
- Van Driel, J. H., Berry, A., & Meirink, J. (2014). Research on science teacher knowledge. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of Research on Science Education* (pp. 848–870). New York, NY: Routledge.
- Van Wylen, D. G., Abdella, B. R., Dickinson, S. D., Engbrecht, J. J., & Vandiver, R. (2013). Interdisciplinarity: The right people, a supportive place, and a program emerges. *CBE Life Science Education*, 12(2), 140–143. doi:10.1187/cbe.13-01-0001

- Vosniadou, S., & Ioannides, C. (1998). From conceptual development to science education: A psychological point of view. *International Journal of Science Education*, 20(10), 1213–1230. doi:[10.1080/0950069980201004](https://doi.org/10.1080/0950069980201004)
- Watanabe, T., & Huntley, M. (1998). Connecting mathematics and science in undergraduate teacher education programs: Faculty voices from the Maryland Collaborative for Teacher Preparation. *School Science and Mathematics*, 98(1), 19–25. doi:[10.1111/ssm.1998.98.issue-1](https://doi.org/10.1111/ssm.1998.98.issue-1)
- Wojdak, J., Guinan, J., Wirgau, J., Kugler, C., Hammond, G., Small, C., ... Baldwin, A. (2010). University facilities as real-world foci of multidisciplinary science learning. *Journal of College Science Teaching*, 39, 8–16.
- Yarden, H., Marbach-A., G., & Gershoni, J. M. (2004). Using the concept map technique in teaching introductory cell biology to college freshmen. *Bioscene*, 30(1), 3–13.
- Yin, Y., & Shavelson, R. J. (2008). Application of generalizability theory to concept map assessment research. *Applied Measurement in Education*, 21(3), 273–291. doi:[10.1080/08957340802161840](https://doi.org/10.1080/08957340802161840)

Appendix A

Table A1. List of Concepts with Centrality Ratings

Concept	Centrality Rating	Percent use at pre-PD	Percent use at post PD	Percent change in usage
Water	3.00	90.5	95.0	+5.0
Soil	2.09	57.0	57.0	0.0
(Dissolved) Oxygen	2.06	67.5	65.5	-2.0
Salinity	2.06	53.5	43.5	-10.0
Deposition	1.43	42.0	38.0	-4.0
Habitat	1.43	47.5	47.5	0.0
Plants/Vegetation	1.43	81.0	81.5	+0.5
Sediment	1.43	43.5	34.0	-9.5
Zones	1.25	29.5	10.0	-19.5
Oxidation/Reduction	1.25	10.0	23.5	+13.5
Permeability	1.25	23.0	28.5	+5.5
Energy	1.12	37.5	37.5	0.0
Nutrients	1.12	52.0	33.5	-18.5
Animals	1.04	57.5	71.5	+14.0
Elevation	1.04	18.0	37.5	+19.5
Gradient	1.04	42.5	47.5	+5.5
Pressure	1.01	14.0	4.5	-9.5
Photosynthesis	0.97	42.5	37.5	-5.0
Specific Heat	0.94	9.5	9.5	0.0
Newton's Laws	0.93	5	13.5	+8.5
pH	0.86	62.5	47.5	-15.0
Infiltration	0.81	5.0	9.5	+4.5
Light	0.77	47.5	57.0	+9.5
Erosion	0.72	42.5	43.0	+0.5
Porosity	0.72	24.0	24.0	0.0
Transportation	0.72	5.0	5.0	0.0
Mudflats	0.71	20	20	0.0
Runoff	0.71	51.5	24.5	-27.0
Calcium	0.69	0.0	4.5	+4.5
Rocks	0.60	32.5	27.5	-5.0
Pollutant	0.60	42.5	32.5	-10.0
Minerals	0.58	42.5	14.5	-28.0
Temperature	0.58	43.5	37.0	-6.5
Density	0.58	15.0	19.0	+4.0
Carbon	0.54	0.0	0.0	0.0
Maps	0.50	5.0	13.5	+8.5
Velocity	0.48	10.0	19.0	+9
Atmosphere	0.46	28.5	27.5	-1.0
Carbon Dioxide	0.46	52.5	27.5	-25.0
Climate	0.46	28.0	18.5	-9.5
Nitrogen	0.46	19.5	14.0	-4.5
Uplands	0.46	5.0	5.0	0.0