Paper 3
High School Students' Sense Making with Contour Maps of Hydrologic Systems

Beth A. Covitt
University of Montana

Bess Caplan
Cary Institute of Ecosystem Studies

Agatha Podrasky
University of Montana

Yidan Zhang
University of California, Berkeley

Amanda Morrison
Colorado State University

Alan Berkowitz
Cary Institute of Ecosystem Studies

Garrett Love
North Carolina School of Science and Mathematics

Kristin L. Gunckel
University of Arizona

Paper prepared for the national meeting of
NARST: A Worldwide Organization for Improving Science Teaching and Learning Through Research

Atlanta, Georgia
March 11, 2018

Introduction

One challenge for education aimed at developing scientific literacy is that literacy required for informed participation in debates about socio-scientific issues requires individuals to coordinate multiple facets of knowledge and practice from different disciplines. This complex nature of required knowledge and practice is reflected in the Next Generation Science Standards (NGSS Lead States, 2013), which, within its three strands, integrates aspects of related disciplines including computation, quantitative analysis, engineering, and communication. In this paper, we explore one area of interdisciplinary performance from among the many possible combinations; namely high school students' capacity to make sense of and answer questions with contour maps depicting hydrologic systems. We'll use the acronym HMSM to represent "H-ydrologic contour M-ap S-ense M-aking" throughout the paper.

We are interested in this area for several reasons. First, HMSM represents an aspect of scientific literacy that will be valuable for individuals beyond their time as K-12 students. Contour maps of hydrologic systems (including topographic maps that show surface water, maps of water table elevations, and maps of contamination plumes in groundwater) are common representations used to communicate with the public about socio-environmental problems such as flooding or groundwater contamination and depletion. Similarly, hydrologic contour maps are typical outputs from computational models that are used to investigate issues in these systems. Thus, HMSM is important for understanding and responding scientific information about socio-environmental issues that individuals are likely to encounter in their future lives. Finally, we are interested in HMSM (and the related focus of computational modeling of hydrologic systems) because we believe these may be under-represented areas of study in the K-12 science curriculum. Today, the work of hydrologic systems scientists requires computational modeling. However, K-12 students' school learning experiences rarely involve contour map sense making and computational modeling of hydrologic systems. The rarity of computational modeling in K-12 hydrology education is not surprising given the rarity of computational thinking in K-12 science education in general (Sengupta, Kinnebrew, Basu, Biswas, and Clark, 2013).

In a current NSF-funded research project called *Comp Hydro*, we are exploring high school students' knowledge and practice related to the nexus among hydrology, data, and computation. We are finding, for example, that when they first encounter computational modeling of hydrologic systems, high school students tend to focus on the visual representations produced by models (sometimes even making the assumption that the visual representations ARE the computer models themselves). While this may reflect over-simplification as a sense-making strategy students adopt in a new learning domain, this novice perspective also points to the idea that students intuitively know they need to understand the structure and function of a physical system being modeled in order to make sense of computational models of that system. Although a visualization (e.g., map) produced by a computational model is not the model itself, students' focus on visual representations may provide a leverage point for helping them develop knowledge of system structure and function needed to take on more complex sense making, including computational modeling to explain and predict events in hydrologic systems.

Our exploration of students' HMSM is situated in the context of the larger *Comp Hydro* project. The project's core research goal is to develop inter-connected discourse-based learning progressions articulating ways students engage in performances at the nexus among hydrology, data sense making, and computational thinking and modeling. This study of HMSM is contextualized in efforts working with teachers in four[1] states to develop and tests instructional units that engage students in learning within this interdisciplinary nexus.

This paper examining HMSM represents an initial step toward exploring the relationship between HMSM and computational thinking. In particular, we are curious about whether HMSM includes aspects of both scientific and computational thinking that may be necessary prerequisites for being a literate consumer of hydrologic systems computational model outputs. Makers of hydrologic systems contour maps rely on the assumption that the map users have understanding of the represented system's structure

[1] Our project works with teachers in Arizona, Colorado, Maryland, and Montana, but we do not yet have a data set from Colorado that pertains to this paper.

and function. In turn, system structure and function are key elements of computational thinking and modeling embodied in concepts such as abstraction, problem representation, model domains, and model calibration and evaluation (Grover & Pea, 2017; Wing, 2006).

Thus, we are working to position our project to explore how developing hydrologic systems science expertise can help students build computational thinking competence, and vice versa. For example, students may develop knowledge and practice concerning systems abstraction through learning to make inferences about real world systems from two-dimensional system representations encoded in contour maps, and vice versa. Abstraction is a core element of computational thinking. Similar synergies between science and computation have been noted by scholars including Sengupta, et al., (2013) who comment on the centrality of abstraction in both computational thinking and scientific expertise.

This paper reports on one strand in the *Comp Hydro* project's efforts to develop interconnected learning progressions for hydrology, data sense making, and computational thinking. We do so by examining responses to four items asking students to demonstrate HMSM. First, we break down and describe elements of HMSM required to coordinate knowledge and practice in scientifically literate ways. Then, through examining less successful approaches students adopt, we identify and organize reasoning challenges common in this performance domain. Identifying and organizing patterns of increasingly more sophisticated knowledge and practice forms the basis of a discourse-based learning progression, which, once refined and validated, may be used to develop instruction that is both responsive to students' common informal approaches to HMSM performances and that outlines goal knowledge and practice that instruction can be designed to help students reach (National Research Council, 2012).

Research Questions

We address the following research questions:

1. What patterns (indicators) are evident in students' HMSM performances? What is the relative frequency of different sense making patterns?
2. How do patterns reflect a range in sophistication in students' HMSM?
3. What do initial IRT analyses suggest concerning the relative difficulties of the HMSM item steps?

Existing Research Concerning Understanding of Contour Maps (and General Maps as Relevant)

Investigators have examined map perceptions and uses in efforts that began some decades ago (e.g., Boardman, 1989; Filippakopoulou, Michaelidou, and Nakos, 1998; Liben and Downs, 1993; Matthews, 1984) and that have continued in more recent years (e.g., Atit, 2016; Clark, et al., 2008; Ishikawa and Kastens, 2005; Lee, 2017; Newcombe, et al., 2015; Rapp, Culpepper, Kirkby, and Morin, 2007; Ooms, et al., 2016). Through these efforts, researchers have identified common challenges people encounter when they seek to make sense of and use maps. Below, we describe some challenges that have been identified, focusing in particular on those that are relevant to our interest domain of HMSM. Researchers have found that people encounter challenges with:

- Visualizing three-dimensional landscapes from two-dimensional representations (Boardman, 1989; Clark, et al., 2008; Rapp, et al., 2007)
- Interpreting maps that show information about continuous dimensions (e.g., elevation) using categorical representation (e.g., contour lines). Students may read shifts in elevation associated with contour lines as step-like rather than continuous (Atit, 2016; Boardman, 1989).
- Grasping meaning of elevation numbers on topographic maps that do not show the ocean (Atit, 2016)
- Decoding correspondence between patterns of contour lines and the shape of surfaces on Earth. This is challenging because there are no visible boundaries in Earth structures that align with contour lines on maps (Atit, 2016).
- Making inferences about unlabeled, implicit data on maps (e.g., contour lines that are not marked with a specific elevation number) (Boardman, 1989)

- (Similarly) extrapolating beyond labeled data (e.g., challenges associated with appreciating that land within adjacent and further contours in a certain direction continue to gain or lose elevation compared with labeled area) (Boardman, 1989)
- Decoding the relationship between closeness of lines and steepness of slope, sometimes confusing steepness of slope and highness of elevation (Boardman, 1989; Rapp, 2007)

Further, investigators have also identified a variety of promising approaches for addressing these challenges including:

- Using gestures (Atit, et al., 2016)
- Engaging students in orienteering exercises (Boardman, 1989)
- Engaging students in learning with multiple map formats (Taylor, Renshaw, and Choi, 2004)
- Engaging students in constructing self-explanations for their map sense-making (Kastens and Liben, 2007)
- Adjusting map features with approaches such as groupings, shadings, and stereo visualizations (Filippakopoulou et. al, 1998; Ishikawa and Kastens, 2005; McGuigan, 1957; Rapp, et al., 2007)
- Engaging students in spatial tasks that require visualization, manipulation, construction, and/or sketching (Lanca and Kirby, 1995; Liben, Kastens, and Christensen, 2011; Lowrie and Logan, 2006; Wallgrün, 2010)

Reasoning for Making Sense and Answering Hydrologic Systems Questions with Contour Maps

Our task in this paper is related, but different from the map understanding and use research efforts described above. In particular, to begin developing a discourse-based learning progression for HMSM, we are (1) focusing on a very particular slice (i.e., HMSM) of the larger domain of applied map sense making, and (2) seeking to articulate and organize increasingly more sophisticated approaches in a progression characterizing different ways students approach these map tasks.

Discourse-based learning progressions seek to use grounded evidence of how students make sense of a domain to articulate a span of increasingly more sophisticated knowledge and practice (National Research Council, 2007). A discourse is a way of talking, thinking, and acting that identifies a socially meaningful group (Gee, 1991). Students bring with them to school a primary (home) discourse, which is a way of talking, thinking and acting that is a natural consequence of being a member of their home and community. When confronted with new tasks they must perform at school, students often initially apply the ways of thinking and solving problems from their primary discourse. One type of primary discourse people commonly apply to science performances when they do not have access to scientific discourse is called force-dynamic reasoning. Force-dynamic reasoning is a type of discourse that is rooted in grammar and language and that has been described by linguists Talmy (1988) and Pinker (2007). Force-dynamic discourse involves viewing the world as a place where actors with purposes confront antagonists; events and phenomena are viewed and explained as results of these confrontations.

Often, the ways of thinking and talking associated with primary discourse are not well suited to science performances in school. Thus, a core goal of science education is to help students learn a new discourse (similar to learning a new language) that they can apply as appropriate to performances that call for scientific ways of talking, thinking and acting. Students do not unlearn or give up their primary discourse when they learn scientific discourse (much like they do not forget a native language when they learn a second language). Ideally, though, they learn when scientific discourse is called for and develop the capacity to apply it as needed. Briefly, scientific discourse may be characterized as a way of thinking and acting that involves reliance on scientific models and principles and that follows accepted norms for engaging in science practices like those enumerated in the NGSS (e.g., Covitt, Dauer, and Anderson, 2017).

We begin by describing an initial take on an "upper anchor" for a hypothesized HMSM learning progression. An upper anchor defines the knowledge and practice associated with a goal scientific performance in a domain. We developed this initial upper anchor based on previous research in map sense making (described above), our own past work on a hydrologic systems learning progression and sense making with hydrologic systems representations (Covitt, Gunckel, and Anderson, 2009; Gunckel, Covitt, Salinas, and Anderson, 2011), and initial analyses from our current research project. In the results section, we will provide evidence from high school students' actual performances, noting the extent to which students demonstrate upper anchor types of knowledge and practice as well as documenting the extent to which and describing how students provide performances representative of other types of discourse including primary types of discourse and intermediate types that could represent pathways toward upper anchor knowledge and practice.

In order to make sense of and consistently and successfully answer questions with contour maps of hydrologic systems, individuals need to be able to coordinate multiple facets of knowledge and practice. The following list articulates our initial set of facets of upper anchor reasoning for HMSM. Bulleted items represent examples of each facet and are not complete lists.[2]

1. Map language (signs, symbols, words, and numbers) (Weeden, 1997)
   - *Map key symbology* (e.g., compass rose, scale)
   - *Map feature symbology* (e.g., contour lines on groundwater map represent water table not land surface)
   - Insets
2. Spatial reasoning, perspective, orientation, scale and dimensions
   - *Categorical encoding* (e.g., on contour maps, continuous information - elevation, concentration, etc. - is categorically encoded, Atit, 2016)
   - *2D/3D Inferences* (Interpreting topographic maps requires inferring three-dimensional space from two-dimensional representations. The reverse is true for making topographic maps.)
   - *Orientation* (The map compass can be used to orient cardinal directionality on a map)
   - *Perspective and Dimensionality* (Map view perspective is two-dimensional, but can be used with symbology to represent multiple dimensions in space or along other variables - e.g., concentration, rainfall, pressure, temperature)
3. Quantitative approaches
   - *Estimation* (Estimate distances and use dimensions and units to estimate a unit value for a point on map - e.g., elevation of point A or concentration of sample drawn from well at point B)
   - *Gradient* (Estimate slope from contour lines - this requires reasoning along an added dimension to coordinate a change in elevation represented by a contour line along a horizontal distance on a map)
4. Hydrologic systems principles
   - *Water flow* (Water flows from high to low potential energy for groundwater or elevation for surface water)
   - *Groundwater structure* (Groundwater exists in pore spaces in sediments and materials, which needs to be understood to make inferences about how water is moving below ground through a system, for example, not in an underground river, and consequently to do things like make inferences about relative velocity of groundwater versus surface water)
   - *Groundwater-surface water interface* (Surface water and groundwater are connected and surface water is the manifestation of the water table where the water table elevation is higher than the land surface elevation)

Methods

Context

We addressed the research questions within the context of the *Comp Hydro* project, in which high school teachers and students in three states (Arizona, Maryland, and Montana) completed instructional

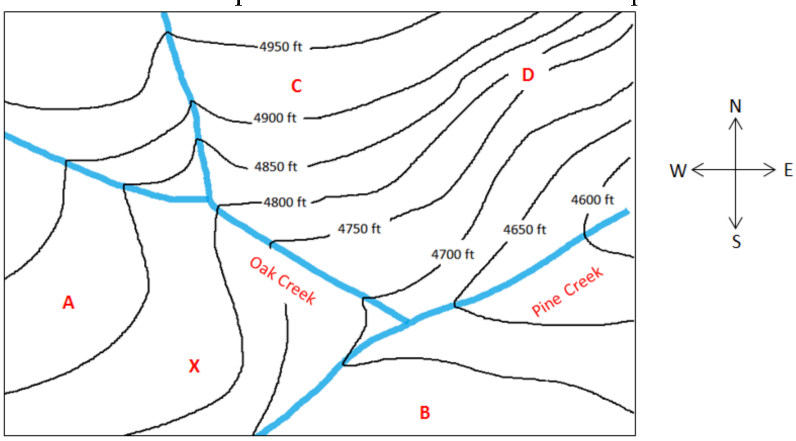
[2] Note too that the four headings are interconnected, and it is not always straightforward to place examples under one particular heading.

units addressing modeling (including computational modeling) of hydrologic systems. A different site-specific unit was implemented in each state. Units in Arizona and Montana addressed problems of groundwater contamination and the unit in Maryland addressed flooding. In all three states, instructional units addressed hydrologic science, data sense making, and computational modeling of systems in an integrated fashion. Unit instruction in all states required about two to three weeks of class time (i.e., the equivalent of between about 10 and 15 one hour lessons).

Instruments and Items

Data were collected through an online system in which participating teachers implemented *Comp Hydro* pre and post-assessments with their students. Responses to the four items shown in Figures 1 through 4 were used as the data sources for this paper. Each item asks students to make sense of a contour map to answer a question about the represented hydrologic system. Each item includes both a forced-choice portion and an open-ended written explanation portion. A student's responses to both the forced-choice and open-ended portions of an item were examined together to yield one response code per item per student.

Use this contour map of a land surface to answer the questions below.



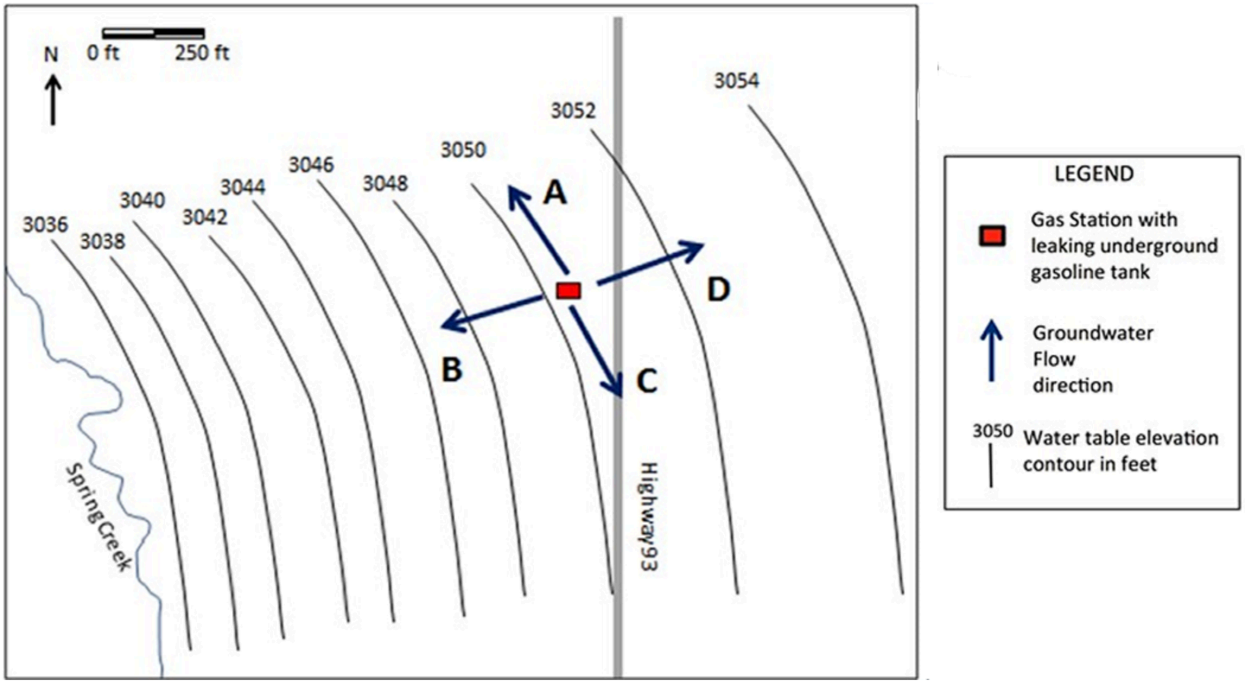
What would be a reasonable estimate for the elevation of land surface at the X on the map?
___4800 ft
___4850 ft
___4815 ft
___4765 ft
Please explain your thinking behind the answer that you chose.

Figure 1. Topographic Interpolation Item

(Using same contour map as in Figure 1)
Where is the slope of the land the steepest?
___A
___B
___C
___D
Please explain why you chose that answer?

Figure 2. Topographic Gradient Item

A gas station has been leaking gasoline from an underground storage tank. A chemical called MTBE that occurs in gasoline has been found to be contaminating the groundwater. The map below shows the site of the gas station and the water table elevation contours



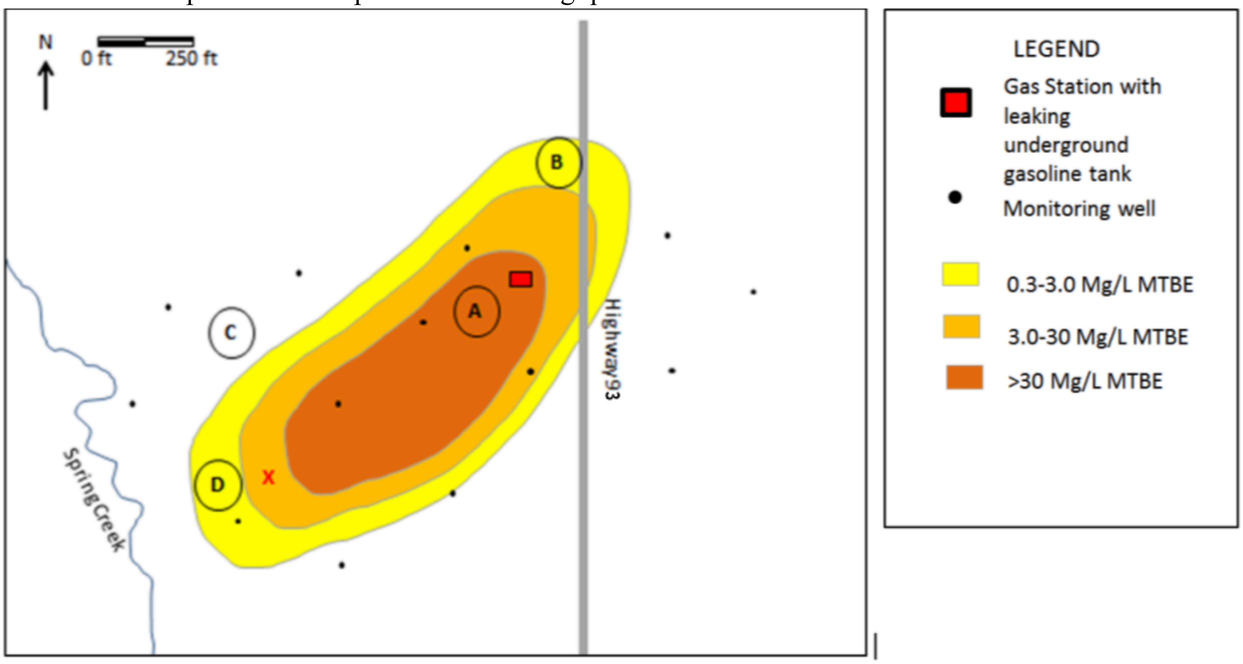
Which of the four arrows on the map best shows the direction contamination will flow from the leaking storage tank?

___Arrow A
___Arrow B
___Arrow C
___Arrow D

Please explain why contamination will flow in that direction?

Figure 3. Ronan Water Flow Arrows Item

The map below shows a computer-generated picture of the contamination (pollution) plume from the leaking storage tank. The plume map was created using MTBE concentration data from the monitoring wells on the map. Use this map for the following questions.



What would be a reasonable estimate of the concentration of MTBE from a groundwater sample taken from a well at the X?

___0.2 mg/L
___3.0 mg/L
___14 mg/L
___ 31 mg/L

Please explain why your choice is the best estimate of concentration.

Figure 4. Ronan Interpolation Item

Participants and Sample

All students were high school students. Data are drawn from students of 15 teachers in 3 states (6 teachers from Arizona, 5 from Maryland, and 4 from Montana). The instructional units were taught as part of the curriculum within a range of courses including biology, computer science, Earth science, engineering design, environmental science, health, human ecology, integrated science, and natural resources management.

The data sample for this paper came from post assessment responses. Post-assessment (as opposed to both pre and post-assessment) data were used in order to examine a data set representing a diverse set of responses including some more and some less sophisticated responses, using the assumption that at least some but not all students would develop more sophisticated ways to respond to these questions as a result of *Comp Hydro* instruction. Assessments were different but had overlapping items across the three states. Assessments were designed to address hydrologic science, data sense making, and computational thinking relevant to the unit created for each state. Item overlap was utilized to facilitate use of IRT analyses. Following these assessment design characteristics, Arizona students completed all 4 items examined in this paper, Maryland students completed 2 of the items, and Montana students completed 2 of the items (see Table 1).

Table 1
Distribution of Assessment Items Completed by State

Assessment Items	AZ	MD	MT
Topographic Interpolation	X	X	
Topographic Gradient	X	X	
Ronan Water Flow Arrows	X		X
Ronan Interpolation	X		X

The following process was followed to obtain a sample of students for analysis from the three states. For each state, we downloaded a file with all post-assessments taken by students (see Table 2). With data files from each state, we began by selecting the subset of post-assessments for which students provided complete responses to all of the questions examined in this paper (4 items for Arizona, 2 items each for Maryland and Montana). An individual item response was considered complete if it contained both a forced-choice response and a written explanation. From the number of complete post-assessments from each state we used a random number generator to select a subset of 75 post-assessments from each state. Thus, the full sample included responses from 225 students. There were 150 responses for each item: 75 Arizona students' responses plus 75 responses from either Maryland or Montana students, depending on the item.

Table 2
Assessment Sample by State

State	# of post-assessments	# of post-assessments w/completion of items	Random sample
Arizona	625	327	75
Maryland	118	78	75
Montana	99	85	75

Data Analysis

Indicator and Level Development and Inter-rater Reliability

The coding scheme was developed through an iterative process involving six project team members and follows previously used learning progression coding methods (Gunckel, Covitt, Salinas, and Anderson, 2012). Members met regularly to discuss student responses once post-assessments became available. Initially, group members would individually make notes about ways of reasoning (proto-indicators) for subsets of responses and then came together to discuss. After a series of discussions and indicator list development at every other week meetings, one researcher created an initial complete indicator exemplar for each of the items. (See Tables 4 through 7 for examples of refined indicator exemplars.) The indicators were organized into levels in the exemplars based on the group's iterative discussions concerning what makes a more or less sophisticated response.

The sample of responses for coding was then created and divided into three inter-rater reliability coding development sets, and posted on a shared spreadsheet where coders could complete codes individually while hiding other coders' columns. The first two inter-rater reliability sets had 30 responses per item in each set and the final set had 90 responses per item. The research group then used the three coding sets to conduct three sets of inter-rater reliability scoring. After each coding set was scored individually by the researchers, the group met and discussed items that were coded differently, refined the indicator exemplar as needed, and came to consensus on indicator codes for responses not initially agreed upon. Inter-rater agreement was examined through coding responses in three successive rounds for each item (Table 3). Two coders completed the final inter-rater reliability coding set.

Table 3
Inter-rater Percentage Agreement

Item	IRR Set (# Responses)	# Coders	% Indicator Agreement	% Level Agreement
Topo Interpolation	1(30)	6	53	70
Topo Interpolation	2(30)	4 to 5	87	97
Topo Interpolation	3(90)	2	96	100
Topo Gradient	1(30)	4	57	90
Topo Gradient	2(30)	3 to 4	83	97
Topo Gradient	3(90)	2	99	100
Ronan Arrows	1(30)	3 to 4	50	57
Ronan Arrows	2(30)	5	87	90
Ronan Arrows	3(90)	2	78	84
Ronan Interpolation	1(30)	4	37	73
Ronan Interpolation	2(30)	5	73	90
Ronan Interpolation	3(90)	2	84	94

Item Response Theory Analysis

Item Response Theory (IRT) is a model-based test theory in psychological measurement. The IRT approach of measurement is sample-independent of the estimation of item and person parameters (Embreston & Reise, 2000). In other words, the item parameters and the person parameters can be compared directly because they are placed on the same measurement scale. Several IRT models including the Rasch model (Rasch, 1960), the rating scale model (Andrich, 1978) and the partial credit model (Wright & Masters, 1982) have been proposed to analyze different types of data. Partial credit model (Wright & Masters, 1982) with weighted mean likelihood estimate method (Warm, 1989) was applied to this study. With the partial credit model application, item parameters and person ability estimates were calibrated to be on the same logit metric. The infit mean squares of all four items and their item steps are in between the standard range of .75 to 1.33, suggesting that the instrument has a good item fit. The person ability estimation of the sample has a range of -4.70 to 3.41 logits, with the standard error of measurement for each person ranging from .84 to 1.96.

Results

***Research Question 1:* What patterns (indicators) are evident in students' HMSM performances? What is the relative frequency of different sense making patterns?**

Patterns in students' HMSM for each of the four questions are shown in Tables 4 through 7. Patterns are represented by indicators showing different ways of thinking that students used in their item performances. Relative frequencies for indicators are shown in the Ind. % column of each table.

Table 4
Topographic Interpolation Item Indicators and Response Examples

	Reasoning Indicators	**Indicator Response Examples**	**Ind. %**
4.1	(4815) Above or slightly above reasoning or higher (than 4800)	4815 ft. It is right about 4800	15
4.2	(4815) Described reason as in-between, estimated, interpolated.	4815 ft. based on its location in between the elevations of 4800 to 4850	29
3.1	(Not 4815) Above or slightly above or higher than reasoning	4850 ft. I think that it has to be more than 4800	2
3.2	(Not 4815) In-between, estimated, interpolated reasoning	4765 ft. Because it seems to be in between 4800 and 4750.	5
3.6	(4815) Proximity (close to, near to) to 4800 reasoning	4815 ft. The answer is the closest to the point of x	4
2.1	(Any elevation) Number reasoning	4800 ft. the numbers are acending as they go up	3
2.2	(Any elevation) Water flow direction	4800 ft. The X on the map flows to 4800ft	5
2.3	(Any elevation) Line reasoning	4800 ft. The line is 4800 so I just followed the line	5
2.4	(Any elevation) Creek reasoning	4815 ft. I think that X is not that far out of oak creek its only a couple of feet out.	3
2.5	(Any elevation) Other reasoning	4800 ft. because it seems large space.	6
2.6	(Not 4815) Proximity (near or close to something other than creek)	4800 ft. Because X is close to 4800 ft	11
2.7	(Any elevation) Low area of map reasoning	4765 ft. The x is like in a low sea level.	2
2.8	(Not 4815) Elevation (no explanation) reasoning	4800 ft. I believe that if the land surface is 4800 ft than so is the elevation	1
1.1	Unclear reasoning	4815 ft. it goes to the north	2
1.2	IDK / No reason / Guess	4800 ft. i don't know about this question	7

*Missing indicator numbers were deleted or combined with other indicators during code development process.
**Percentages may not add to 100% due to rounding.

Table 5
Topographic Gradient Item Indicators and Response Examples

Reasoning Indicators		Indicator Response Examples	Ind. %
4.1	(D) Change of elevation is fast or over a small distance reasoning (use this if ALSO 4.2 or 4.3)	D. It decreases 50 ft in a smaller distance than any other place.	5
4.2	(D) Closer/more lines = steeper reasoning	D. Closer the lines are the steepest it will be.	4
4.3	(D) Closer/more lines reasoning	D. the lines are closest together at point d	6
3.1	(A, B, or C) Change of elevation is fast or over a small distance reasoning	A. The elevation levels in that area go down quickly in a short amount of space.	1
3.2	(A, B, or C) Closer/more lines = steeper reasoning	C. I chose C because the values around C were close together, indicating it would be steep.	1
3.3	(A, B, or C) Closer/more lines reasoning	C. because the lines are really close to each other.	1
3.4	(D) Steep line or steepest by look reasoning	D. its D because it looks very steep.	3
3.6	(D) High reasoning	D. d because its the highest point	1
2.1	(A, B, or C) Steep line reasoning or steepest by look or steepest w/no reason	A. The contour lines are very steep and curved to support my answer	16
2.2	(A, B, or C) Lines far apart reasoning	B. It has bigger gaps between contour lines.	4
2.3	(A, B, or C) High or big number reasoning	C. This is the highest elevation in the map.	17
2.5	(Any letter) Undifferentiated or other line reasoning	C. That is wear the line drops the most.	3
2.6	(Any letter) Low elevation or number reasoning	B. The elevation in that area is lower.	4
2.7	(Any letter) Proximity reasoning	A. it is the closest to 4900 ft	13
2.8	(Any letter) Water flow reasoning	B. because all are flowing in that direction	3
2.9	(Any letter) Other reasoning	B. there is a hill on B witch puts B as the steepest	6
1.1	Unclear reasoning	C. is falling straight	4
1.2	IDK / No reason / Guess	C. I don't know	9

*Missing indicator numbers were deleted or combined with other indicators during code development process.
**Percentages may not add to 100% due to rounding.

Table 6
Ronan Water Flow Item Indicators

	Reasoning Indicators	Indicator Response Examples	Ind. %
4.1	(B) High to low potential energy (may also mention additional explanations)	B. Because it goes from high to low potential energy.	5
4.2	(B) High to low elevation, to decreasing elevation (may also mention e.g., perpendicular to or through contours, gravity)	B. Because the elevation is lowering going to the direction of arrow B	25
4.3	(B) Gravity or gravity plus down or downhill or slope	B. It will flow that way because gravity will pull it down hill to the lowest point until it stops.	1
3.2	(A, C, or D) High to low elevation, to decreasing elevation (may also mention additional explanations - e.g., perpendicular to or through contours)	A. It is the direction in which the elevation is decreasing.	5
3.3	(B) Down, down slope, downhill, downwards, down to river, lower (without mentioning elevation), slope (w/out gravity) (may also mention e.g., perpendicular to or through contours)	B. it flows downhill better.	25
3.4	(B) Elevation w/out any "decreasing" explanation	B. elevation	4
3.5	(A, C, or D) Gravity or gravity plus down or downhill or slope	C. bc of gravity and slope	6
3.6	(B) Perpendicular to or through contours only	B. I picked B because it would flow through all of the contours, and it would be easier to track and see it's water flow.	1
3.7	(B) To water, creek, or river	B. It will head towards the Creek	3
2.1	(A, C, or D) Down, down slope, downhill, downwards, down to river, slope (w/out gravity)	A. Water always flows to a lower level,	4
2.2	(A, C, or D) Elevation w/out other explanation	A. because of elevation	1
2.3	(A, C, or D) Gravity or gravity plus down or downhill or elevation or slope	A. it travels down because of gravity.	1
2.4	To different or higher or steeper elevation	C. that is the direction of the different elevations	3
2.5	By or to highway, or location of gas leak, or proximity	D. It is by the highway	4
2.6	Because map shows, or mentions lines, or parallel to line	A. it says its heading arrow a direction	5
2.7	Other	A. It would be flowing North East.	4
2.8	(A, C, or D) To water, creek, or river	C because it spread to the water	1
1.1	Unclear reasoning	D. D would be the most different and most useable.	2
1.2	IDK / No reason / Guess	C. idk	1

*Missing indicator numbers were deleted or combined with other indicators during code development process.
**Percentages may not add to 100% due to rounding.

Table 7
Ronan Interpolation Item Indicators

Reasoning Indicators		Response Indicator Examples	Ind. %
4.1	(14mg/L) In between, middle of, or estimated	14 mg/L. Its in between 3.0-30 mg/L	39
3.1	(0.2, 3.0, or 31mg/L) In between, 3 is in 3 to 30, or estimated	3.0 mg/L. The MTBE shows from 3.0-30.	5
3.3	(14mg/L) Used map, key, or color reasoning. Or used map to find area of contamination	14 mg/L. The color on the map is not really to dark or too light rather than its more in between.	3
3.4	(14mg/L) Because of movement/flow	14 mg/L. The water flow is ferther than the spring creek	1
2.1	(0.2, 3.0, or 31mg/L) Used map, key, or color reasoning. Or area of contamination (other than high contamination).	3.0 mg/L. because of the key and color	25
2.2	(0.2, 3.0, or 31mg/L) Because of movement/flow	3.0 mg/L. It seems like its also coming from a different direction	1
2.3	It was the right size or level	14 mg/L. bc that's the right amout of liqwould	3
2.5	Proximity reasoning	3.0 mg/L. because its closer to the x	6
2.6	High, highest, high contamination	31 mg/L. because it's the highest contamination level.	7
2.8	Other	14 mg/L. it doesn't seem the concentrated	3
1.1	Unclear	14 mg/L. it is an accurate	5
1.2	IDK/No reason/Guess	3.0 mg/L. i don't understand	3

*Missing indicator numbers were deleted or combined with other indicators during code development process.
**Percentages may not add to 100% due to rounding.

Research Question 2: How do patterns reflect a range in sophistication in students' HMSM?

Indicators for each item were also divided into four initially proposed levels of increasing reasoning sophistication. IRT analyses, and ideally, learning-focused studies in the future, will be used to refine and validate the arrangement of indicators within ordered levels. Distributions of percentages of responses coded at each level are shown in Table 8.

Table 8
Distribution of Item Responses by Coding Levels

	Topographic Interpolation	Topographic Gradient	Ronan Water Flow Arrows	Ronan Interpolation
Level 4%	44	15	32	39
Level 3%	11	5	43	9
Level 2%	26	67	22	45
Level 1%	9	13	3	8

*Percentages may not add to 100% due to rounding.

Level 1 represents unsophisticated reasoning including "don't know" reasoning, uninterpretable reasoning, or similar. Level 1 responses are generally not related to the facets of HMSM described earlier in this paper.

Level 2 represents responses reflecting informal, unsophisticated, and mostly unsuccessful HMSM. Examples include using proximity (nearness) as a rationale, reasoning about lines and numbers on the map in ways that are inconsistent with what the lines and numbers are intended to represent (e.g. interpreting an elevation contour line as though from cross-section perspective), or inferring that map symbols provide reasons for choices. With respect to facets of reasoning with contour maps, students operating at level 2 are mostly bringing informal rather than scientific principle-oriented approaches to their performances. For example, proximity reasoning, which is evident in many Level 2 responses, is a common force-dynamic reasoning approach we have identified in previous analyses. Reasoning that relies primarily on proximity or nearness often reflects an informal idea that the capacity of one thing to relate to or act on another thing depends on how close together the two things are. While proximity may play a role in cause and effect or other phenomena and events, "nearness" itself is not the causal mechanism or explanation.

Often, students responding at Level 2 are encountering challenges that span multiple facets of map and systems reasoning including making sense of map language (e.g., knowing that a contour line represents water table elevation) and connecting map language to systems structure and principles in ways that demonstrate capacity to visually manipulate spatial dimensions (e.g., interpreting a groundwater contour as a representation of the same elevation of the water table across a span of horizontal space at some depth below the three-dimensional domain shown on the physically two-dimensional map). Map application questions like those used in this paper are difficult for students operating at Level 2 because they require coordination among multiple facets of reasoning that these students are still working on.

Although they do not represent the "upper anchor," Level 3 responses still demonstrate significant knowledge and practice achievements. Students operating at Level 3 have developed understanding of the structure of hydrologic systems and are thinking about and trying to apply scientific principles that govern how hydrologic systems functions in their explanations.

Level 3 responses often fall into several categories. Sometimes, level 3 responses demonstrate problematic, incomplete, or inaccurate attempts to use principled reasoning to make sense of the map. For example, a response might describe the reason for direction flow on a map as high to low elevation, but then choose a direction that actually points toward higher elevation. In other cases, level 3 responses can represent use of rules, sometimes learned in school, that may be right, but that aren't intellectually connected with scientific principles. An example would be using "flow is perpendicular to contour lines" as an explanation for choice of groundwater flow direction. While flow direction is perpendicular to contour lines, this explanation is more like a fact learned in class than like an explanation based on the physics of groundwater flow.

Level 4 responses generally provide responses and explanations that reflect coordination across the facets of HMSM described above (map language, spatial reasoning, scientific principles, etc.) as appropriate. This coordination represents a significant accomplishment of knowledge and practice that can be used in a flexible way to make sense of and respond to questions about hydrologic systems and issues and problems related to hydrologic systems.

Relative frequencies for student responses at each level are shown in Table 8. An interesting pattern emerged showing that for three of the four items, Level 3 responses are less frequent than either Level 2 or Level 4 responses. This pattern is different from patterns we have seen in previous related research, in which we've often found that Level 3 responses on post-instruction assessments are more common than Level 4 responses, suggesting that fewer students attain capacity to create sophisticated scientific model-based responses. There are several possible explanations for this level response frequency pattern. It may be that the arrangement of indicators that we have organized in the levels requires refinement to better delineate distinct approaches to HMSM represented in the different levels.

Another hypothesis for why we see a relatively higher frequency of Level 4 compared with Level 3 responses in this set of hydrologic systems contour map items is that there are some intellectual synergies among the facets of map reasoning described in this paper. For example, in the project's groundwater instructional units, students engage in integrated experiences involving groundwater elevation contour maps, three-dimensional models of the water table, and experimentation with physical models that demonstrate how groundwater moves from high to low potential energy (which is summarized as elevation in a contour map). It may be that, for some students, once they have opportunities to connect the multiple facets through instructional experience, they are able to answer and explain their answers to questions about the contour maps with some consistency and accuracy. However, based on the percentages of students who remain at Levels 1 and 2 on the items, it is clear that a significant portion of high school students were not successful at coordinating the facets of reasoning with contour maps based on their project learning experience.

Research Question 3: What do initial IRT analyses suggest concerning the relative difficulties of the HMSM item steps?

A Wright map (Figure 5) was produced to represent the relations between the item difficulty and person ability estimates. For each item, the point tt_x is a threshold represents that if the respondent has an estimated ability of the denoted logit level corresponding to the location of the point, then this respondent has a 50% chance of reaching level x or above for this item. For example, for the TG item, respondents with estimated ability level of approximately -2 logits would have a 50 % chance of reaching level 1 or above for this item. The histogram on the left shows the distribution of person ability estimates in the sample, estimated with the weighted mean likelihood estimate method. The Wright map also shows that for each item, the estimated threshold level represents the order of the theoretical difficulty level, suggesting the instrument to have a good internal structure.

The Wright map results also show the pattern of three of the four items having relatively few performances coded at Level 3. For the three items Topographic Interpolation, Topographic Gradient, and Ronan Interpolation, we see that once students reach the threshold of a Level 3 response (step 2) that there is a short distance to their reaching the threshold for a Level 4 response (step 3). This is in contrast to the Ronan Water Flow item where, when students reach the threshold for a Level 3 response (step 2) they still have a considerable distance on the map until they reach the threshold for a Level 4 response (step 3). We plan to continue our discussions and refinements to indicators and assignment of indicators to levels using insights from the IRT analyses.

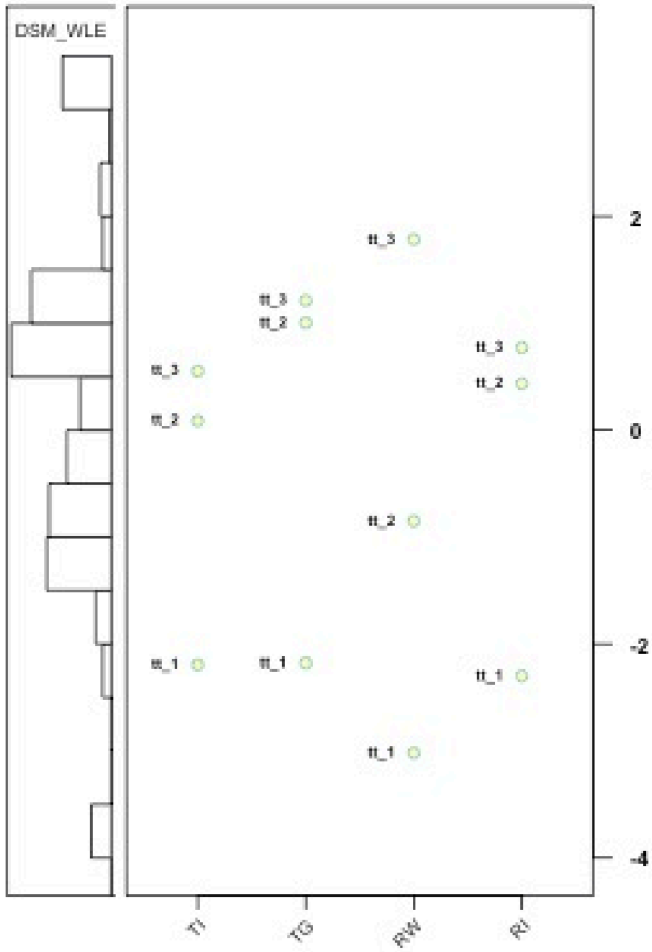


Figure 5. Wright map showing relationship between item difficulty and person ability estimates

Discussion and Conclusion

The analyses presented in this paper represent initial work toward articulating a learning progression for how students make sense with contour maps of hydrologic systems. Analysis of responses demonstrates that after engaging in a relevant high school level learning experience, a significant portion of students retain primary discourse approaches to HMSM. With respect to facets of HMSM reasoning that need to be coordinated to answer questions, students responding at Level 2 are still working on most of the reasoning facets (e.g., using informal principles like proximity rather than scientific principles such as water moving from high to low potential energy or applying informal notions of map symbology such as inferring that contour lines represent a cross section profile rather than equal elevation along a horizontal distance). With the exception of the Ronan Water Flow question, relatively few students provide responses at Level 3 on the post assessments. However, significant portions of students, by the post assessment, develop the capacity to provide Level 4, upper anchor responses that coordinate across the reasoning facets to construct accurate principle-based explanations for the items.

We plan to use the analyses in this paper as a stepping stone to the larger project goal of developing connected and validated learning progressions across the nexus of hydrology, data sense making, and computational thinking and modeling. The work reported here accomplishes several steps in this process, particularly in the area of integrating across data sense encoded in contour maps and hydrologic science. At this nexus, we have identified a set of indicators that can be used to distinguish approaches students use in HMSM. While not focusing on computational thinking specifically, the results presented in this study also have potential connections to computational thinking in areas such as how students make sense of abstraction in systems.

Much work, however, remains to be done. For example, one step we are planning to undertake soon is expanding this analysis to include pre-assessment responses. It was helpful to focus on post-assessment responses first because post responses likely leveraged students' learning experiences in ways that made their responses more comprehensible and easier to classify into indicators. Now that initial indicator development has been completed, looking at pre-assessment responses may provide more insights into patterns of how students approach these questions when they haven't already had relevant learning experiences in school. Given that experiences with contour maps of hydrologic systems in the K-12 domain are relatively rare, we are interested in how students approach these questions in first, untutored encounters. Additional research plans include undertaking work to build and validate learning progression frameworks that link across the full nexus of hydrology, data sense, and computational thinking and modeling.

Also planned are learning studies that examine how students' reasoning and discourse changes as a result of participating in instruction that is designed to be both responsive to the ideas and ways of thinking that students bring with them to school, and rigorous in setting goals for principle-based knowledge and practice needed for environmental science literacy. The site-specific *Comp Hydro* units represent our first attempts at attending to responsiveness and rigor based on our previous research and instructional design work. Continuing work will further our understanding of what kinds of interdisciplinary knowledge and practice are required for environmental science literacy, the kinds of knowledge and practice that students bring with them to relevant performances before instruction, and the instructional strategies and designs that can help support students in attaining capacity for the types of performances they will need to address pressing socio-environmental problems in informed and knowledgeable ways.

References

Andrich, D. (1978). A binomial latent trait model for the study of Likert-style attitude questionnaires. *British Journal of Mathematical and Statistical Psychology*, *31*(1), 84-98.

Atit, K., Weisberg, S. M., Newcombe, N. S., & Shipley, T. F. (2016). Learning to interpret topographic maps: Understanding layered spatial information. *Cognitive research: principles and implications*, *1*(1), 2.

Boardman, D. (1989). The development of graphicacy: Children's understanding of maps. *Geography, 74*(4), 321-331.

Clark, D., Reynolds, S., Lemanowski, V., Stiles, T., Yasar, S., Proctor, S., Lewis, E., Stromfors, C., & Corkins, J. (2008). University Students' Conceptualization and Interpretation of Topographic Maps, *International Journal of Science Education, 30*(3), 377-408.

Covitt, B., Dauer, J., & Anderson, C. 2017. The role of practices in scientific literacy. In Schwarz, C., Passmore, C., & Reiser, B. (Eds.), *Supporting Next Generation Scientific and Engineering Practices in K-12 Classrooms*. Arlington, VA: NSTA Press.

Covitt, B. A., Gunckel, K. L., & Anderson, C. W. (2009). Students' developing understanding of water in environmental systems. *The Journal of Environmental Education*, *40*(3), 37-51.

Embretson, S. E., & Steven, P. Reise. 2000. *Item response theory for psychologists*.

Filippakopoulou, V., Michaelidou, E., & Nakos, B. (1998). A study of children's perception of cartographic landform representation. *Proceedings: Maps for Special Users, ICA, Wroclaw, Poland*, 93-104.

Gee, J. (1991). What is literacy? In C. Mitchell & K. Weiler (Eds.), *Rewriting literacy: Culture and the discourse of the other* (pp. 3-11). Westport, CT: Bergin & Gavin.

Grover, S. & Pea, R. (2018). Computational thinking: A competency whose time has come. In S. Sentence, S. Carsten, & E. Barendsen (Eds.). *Computer Science Education: Perspectives on teaching and learning.* Bloomsbury.

Gunckel, K. L., Covitt, B. A., Salinas, I., & Anderson, C. W. (2012). A learning progression for water in socio-ecological systems. *Journal of Research in Science Teaching*, *49*(7), 843-868.

Gunckel, K. L., Mohan, L., Covitt, B. A., & Anderson, C. W. (2012). Addressing challenges in developing learning progressions for environmental science literacy. In A. Alonzo & A. W. Gotwals (Eds.), *Learning Progressions in Science*. Boston, MA: Sense Publishers.

Ishikawa, T., & Kastens, K. A. (2005). Why some students have trouble with maps and other spatial representations. *Journal of Geoscience Education*, *53*(2), 184-197.

Kastens, K. A., & Liben, L. S. (2007). Eliciting self-explanations improves children's performance on a field-based map skills task. *Cognition and Instruction*, *25*(1), 45-74.

Lanca, M., & Kirby, J. R. (1995). The benefits of verbal and spatial tasks in contour map learning. *Cartographic Perspectives*, 21, 3-15.

Lee, J. (2017). *Learning progressions on map understanding: A case study of elementary through college students in Korea* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. Texas State University, San Marcos.

Liben, L. S., & Downs, R. M. (1993). Understanding person-space-map relations: Cartographic and developmental perspectives. *Developmental Psychology*, *29*(4), 739.

Liben, L. S., Kastens, K. A., & Christensen, A. E. (2011). Spatial foundations of science education: The illustrative case of instruction on introductory geological concepts. *Cognition and Instruction*, *29*(1), 45-87.

Lowrie, T., & Logan, T. (2007). Using spatial skills to interpret maps: Problem solving in realistic contexts. *Australian Primary Mathematics Classroom*, *12*(4), 14.

Matthews, M. H. (1984). Environmental cognition of young children: images of journey to school and home area. *Transactions of the Institute of British Geographers, 9*(1), 89-105.

McGuigan, F. J. (1957). An investigation of several methods of teaching contour interpretation. *Journal of Applied Psychology*, *41*(1), 53.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8.* Washington, DC: The National Academies Press.

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

Newcombe, N., Wesiberg, S., Atit, K., Jacovina, M., Ormand, C., & Shipley, T., (2015). The lay of the land: Sensing and representing topography. *The Baltic International Yearbook of Cognition, Logic and Communication, 10*, 1-57.

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

Ooms, K., De Maeyer, P., Dupont, L., Van der Veken, N., Van de Weghe, N., & Verplaetse, S. (2016). Education in cartography: what is the status of young people's map-reading skills? *Cartography and Geographic Information Science*, *43*(2), 134-153.

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature.* New York, NY: Penguin Group.

Rapp, D. N., Culpepper, S. A., Kirkby, K., & Morin, P. (2007). Fostering students' comprehension of topographic maps. *Journal of Geoscience Education*, *55*(1), 5-16.

Rasch, G. (1960). Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.

Roth, W. M., Pozzer-Ardenghi, L., & Han, J. Y. (2005). *Critical graphicacy: Understanding visual representation practices in school science* (Vol. 26). Dordrecht: Springer Science & Business Media.

Sengupta, P., Kinnebrew, J. S., Basu, S., Biswas, G., & Clark, D. (2013). Integrating computational thinking with K-12 science education using agent-based computation: A theoretical framework. *Education and Information Technologies*, *18*(2), 351-380.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12,* 49-100.

Taylor, H. A., Renshaw, C. E., & Choi, E. J. (2004). The effect of multiple formats on understanding complex visual displays. *Journal of Geoscience Education*, *52*(2), 115-121.

Wallgrün, J. O. (2010). Qualitative spatial reasoning for topological map learning. *Spatial Cognition & Computation*, *10*(4), 207-246.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*(3), 427-450.

Weeden, P. 1997. Learning through maps. In D. Tilbury and M. Williams (Eds.), *Teaching and Learning of Geography* (pp. 168-179). London: Routledge.

Wing, J. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33-35.

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