Preprint: “Developing a Construct Map for Teacher Attentiveness” to appear in *Quantitative Measures of Mathematical Knowledge: Researching Instruments and Perspective.*

Cite as:

Carney, M. B., Totorica, T., Cavey, L. O., & Lowenthal, P. R. (2019). Developing a construct map for teacher attentiveness. In J. D. Bostic, E. E. Krupa, & J. C. Shih (Eds.), *Quantitative measures of mathematical knowledge: Researching instruments and perspectives* (pp. 152-178). New York, NY: Routledge.

DEVELOPING A CONSTRUCT MAP FOR TEACHER ATTENTIVENESS

Michele Carney
Boise State University

Tatia Totorica
Boise State University

Laurie Cavey
Boise State University

Patrick R. Lowenthal
Boise State University

Abstract

Construct maps are important tools in educational assessment and can serve multiple purposes related to development and validation, as well as score interpretation and use. This chapter outlines a process for developing a construct map from the qualitative ordering of teachers’ responses to open-ended assessment items. The construct of interest pertains to a teacher’s ability to attend to what students say and do, which is a key component of many recommendations for instructional practice within mathematics education. The instrument we are developing is designed to measure teachers’ attentiveness to student thinking in quantitative reasoning problem situations. A key aspect of our instrument development process is the development of a construct map that hierarchically orders qualitatively different levels of teacher attentiveness. In this chapter we describe our process for developing the construct map with the intent of providing an example to others who may be interested in engaging in the development of construct maps.

Author Note: This research was supported in part by grant #*1726543 Preparing Secondary Mathematics Teachers with Video Cases of Students’ Functional Reasoning* from the National Science Foundation.

Introduction

Construct maps, as described by Wilson (2005) in *Constructing Measures*, are used to describe an idea or concept that can be thought of as consisting of hierarchically ordered and qualitatively distinguishable categories along a continuum (e.g., low to high levels of knowledge for a particular topic). Construct maps are important tools in educational assessment and can serve multiple purposes related to development and validation, as well as score interpretation and use. For example, construct maps are particularly useful for demonstrating how a theoretical construct has been operationalized into an assessment. Operationalizing a construct into an assessment necessarily bounds the construct into a smaller subset of ideas than is typically envisioned at the theoretical level. The construct map helps test developers and end-users better understand what is and is not included within the assessment of a construct. Relatedly, Shepard (2017) has stated, "To support student learning, quantitative continua must also be represented substantively, describing in words and with examples what it looks like to improve in an area of learning (p. 1)." Construct maps can be used to assist in the development of these substantive qualitative descriptions of test performance, which, in turn, provide meaningful score interpretation and use. Construct maps can also be helpful in the iterative improvement of an assessment, which might include informing continued item development efforts as testing needs change. These are just a few of the potential applications of construct maps for the improvement of assessments and their applications in education.

While Mark Wilson's work at the Bear Evaluation and Assessment Research center is extremely well-regarded, there are few examples within the measurement literature at-large specifying how to go about creating construct maps, particularly for open-ended assessment items. The purpose of this chapter is to provide an example of developing a construct map for teacher attentiveness to student thinking, referred to as *attentiveness*. Attentiveness is a particularly complex construct to assess as it is related to the pedagogical content knowledge a teacher uses to respond to a student in ways that both take the student's ideas seriously and enables the student to build on their own ideas. Providing an example of this type of work for others demonstrates how complex constructs such as attentiveness can be mapped along a continuum via the use of responses to open-ended assessment items.

Literature Review

This section provides a review of the literature that informs the attentiveness construct, describes why it is important to assess attentiveness, and highlights key considerations in the assessment of attentiveness. The iterative relationship between construct map development, instrument development, and analysis and validation of an assessment tool for attentiveness is outlined.

The Attentiveness Construct

Attentiveness builds upon ideas from professional noticing (Jacobs et al, 2010), formative assessment (Black & Wiliam, 2009), mathematical knowledge for teaching (Ball, Thames, & Phelps, 2008; Shulman, 1987), and progressive formalization (Freudenthal, 1973; Gravemeijer &

van Galen, 2003; Treffers, 1987) and is defined as the *ability to analyze and respond to a particular student's mathematical ideas from a progressive formalization perspec*tive. It most closely parallels ideas from professional noticing--a set of interrelated skills for teaching that involve attending, interpreting, and responding to student ideas--but differs in that the focus with attentiveness is on *individual* student's thinking and ways in which student ideas can be built upon to become progressively formal. Attentiveness can be seen as a component of high-quality professional noticing but does not include many of the classroom-level components often attributed to professional noticing. Attentiveness can also be viewed as a significant contributor to formative assessment, as it helps bridge the processes of "engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding" and "providing feedback that moves learners forward" within the seminal formative assessment framework developed by Black and Wiliam (Black & Wiliam, 2009). Bounding attentiveness to focus on an individual student's thinking allows the construct to be operationalized at a grain size appropriate for traditional forms of assessment (e.g., tests); whereas professional noticing and formative assessment, due to their classroom context, are harder to assess using traditional forms of assessment.

Attentiveness can be thought of as making use of components within the construct of mathematical knowledge for teaching (MKT), an oft-cited construct in the teacher education literature that refers to content and pedagogical knowledge specific to the work of teaching mathematics (Ball, Thames, & Phelps, 2008; Shulman, 1987). MKT encompasses both traditional mathematical knowledge related to instructional content, as well as knowledge that is specifically related to designing and managing students' classroom experiences. This includes knowledge of how students' mathematical ideas develop, how to promote the development of students' ideas, how to recognize common conceptualizations (both informal and formal), as well as how to identify and create mathematical tasks which elicit important mathematical ideas (Gravemeijer, Cobb, Bowers, & Whitenack, 2000; National Research Council, 2001; Stein, Engle, Smith, & Hughes, 2008; Stein, Smith, Henningsen, & Silver, 2000).

Attentiveness makes use of components within MKT because attending to students' thinking requires a teacher to 1) examine student work in relation to the mathematical intent of the task, 2) situate the student's work within a larger progression of student understanding for the topic, and 3) press students to generalize, formalize, or revise their ideas. In this way, attentiveness is bounded by employing a particular pedagogical lens (progressive formalization) which emphasizes the importance of building upon students' ideas (Freudenthal, 1973, 1991; Treffers, 1987). In progressive formalization, students initially apply their informal mathematical knowledge to a mathematically demanding task (Freudenthal, 1991). Students gradually develop more sophisticated ways of reasoning, eventually connecting their conceptual models to established conventions through the support of the teacher as they work through a sequence of mathematical tasks. From a progressive formalization perspective, a teacher's ability to analyze student work and respond in ways likely to support students in developing more sophisticated mathematical understanding is paramount to that teacher's practice.

Implementing progressive formalization requires the teacher to recognize the valid mathematical ideas within students' informal or incorrect answers, and to make connections from students' informal ideas to more formal mathematics. Likewise, the teacher must be able to interpret incorrect answers that reveal a disconnect within the student's understanding and build on what the student *does* understand to help bridge to the targeted mathematical idea. In this way, attentiveness can be viewed in relation to high-leverage or 'core' practices -- regular routines that teachers use to engage students in productive work (Grossman, Hammerness, & McDonald, 2009). The *Learning to Teach in, from and through Practice Project* (Lampert et al., 2013) has identified eliciting and responding to students' contributions as a significant practice that can be developed in teacher candidates. Attentiveness requires careful attention to students' ideas while keeping the important mathematical goals of the task in mind, similar to the type of ambitious teaching described by Lampert and her colleagues.

There have been multiple calls for mathematics pedagogy focused on these concepts. For example, *Principles to Actions* (National Council of Teachers of Mathematics, 2014) calls for teachers to *elicit and use evidence of student thinking* as one of eight mathematics teaching practices; similarly one of the high leverage practices from TeachingWorks focuses on *working with individual students to elicit, probe, and develop their thinking about content* (TeachingWorks, 2018). Underlying these various calls is the need for teachers to be able to meaningfully interpret student work and build upon student thinking in ways that leads to their understanding of important mathematical ideas. While the end goal is for teachers to interpret and respond to student thinking within the context of a classroom setting (with multiple competing priorities), attentiveness, with its narrowed focus on interpreting and responding to individual student work samples, is an appropriate starting place for teacher educators seeking to build teacher candidates' ability to engage in the wide range of important pedagogical practices focused on the progressive formalization of student thinking.

Assessing Attentiveness

Our instrument development work is, in part, a response to the need to hold teacher preparation and professional development programs accountable for the work they do in preparing and supporting teachers (e.g., Council for the Accreditation of Educator Preparation, 2013; Grossman, Hammerness, & McDonald, 2009; Lampert, 2009). High quality instruments, designed for research and program evaluation, are essential to this endeavor. While commercial assessments like the PRAXIS have long been used for individual evaluation of content knowledge, there has been a growing interest in developing instruments of constructs central to teaching mathematics (e.g., Learning Mathematics For Teaching, 2005). Such instruments could enable educators to examine changes over time for a single program, or to make comparisons across multiple programs. For teacher educators whose work focuses on developing mathematics teachers' ability to analyze and respond to student thinking, tools are needed to help identify whether teachers' engagement in course and program activities influences their attentiveness.

An important consideration when constructing instruments to measure a complex construct such as attentiveness is the meaningful operationalization of the construct. Performance assessments enacted in practice provide an authentic means of assessing constructs such as attentiveness; however, they are difficult to implement at scale and costly in terms of time and money to score accurately. More traditional forms of assessment provide a more efficient means of assessing attentiveness, but can be compromised in terms of authenticity and connections to practice. Given both a program-level focus and an end goal of teachers effectively analyzing and responding to student thinking in practice, a scalable means of assessing attentiveness is needed. Therefore, we opted to develop a more traditional form of assessment, with a focus on authenticity in the operationalization of attentiveness.

Our operationalization of attentiveness in the Attentiveness in Quantitative Reasoning Inventory (QRI) involves identification of key disciplinary ideas for the mathematics topic, in this case quantitative reasoning, in conjunction with common, informal ways students reason about these topics. The key disciplinary ideas and ways of student reasoning are examined through the lens of the following professional noticing categories: analyzing the approach, interpreting understanding, and responding to the student. While we use the professional noticing categories, our operationalization differs slightly from professional noticing in that our focus is on teacher candidates' ability to be attentive to an individual student's thinking, which we see as a precursor to the broader perspective of professional noticing in a classroom context. In addition, our analysis of the assessment item responses includes a focus on analyzing the approach, interpreting understanding, and responding to the student in ways that builds upon their thinking and allows for a progressive formalization of their understanding.

The current version of the Attentiveness in QRI makes use of a constructed-response item format. This version of the instrument is useful for program level assessment but can be time consuming to analyze and score if administered on a large scale. This limitation is further addressed in the discussion section under further assessment development.

Construct Map Development

In Wilson's (2005) book, *Constructing Measures,* creating a construct map is the recommended first step in assessment development. An important decision involved in this first step is determining whether to first articulate the qualitative ordering of the responses or that of the respondents. Wilson states,

> In creating a construct map, the measurer must be clear about whether the construct is defined in terms of who is to be measured - the respondents - or what responses they might give – the item responses. Eventually both will be needed, but often it makes sense in a specific context to start with one rather than the other. (p. 38)

In the development of the instrument described in this chapter, we chose to focus on the qualitative ordering of item responses. While we do not describe the response space for respondents in detail here, it is an important aspect of supporting score interpretation and use which is addressed at the end of the chapter.

In the next section, we address our methods and results associated with articulating the item response portion of the construct map for teacher attentiveness, with a focus on providing detailed examples of how to develop the qualitative categories and ordering of the responses.

The first step was a review of the literature to identify potential shifts in responses as attentiveness increases. The next steps were to code and categorize the responses received from teachers and teacher candidates to further delineate and hierarchically organize the categories of attentiveness. Finally, this information was organized into a construct map diagram for attentiveness.

Methods and Results

Context

The National Science Foundation funded the Video Case Analysis of Student Thinking (VCAST) project. The purpose of the VCAST project is to develop instructional materials with the potential to increase secondary mathematics teacher candidates' ability to analyze and respond to student thinking in quantitative reasoning contexts. In an effort to evaluate the influence of the VCAST intervention on teacher candidates' attentiveness, the Attentiveness in QRI was administered in the fall of 2017 to candidates enrolled in an upper division mathematics course addressing functions and modeling at the secondary level. The Attentiveness in QRI was also administered in the summer of 2018 to assist in the evaluation of a professional development institute which utilized the VCAST materials and focused on functions and modeling for secondary teachers.

Participants

The Attentiveness in QRI was administered to a total of 42 respondents: 17 secondary mathematics teacher candidates and 25 secondary mathematics teachers. The teacher candidates were enrolled in an upper division functions and modeling mathematics course designed for future secondary mathematics teachers in the United States. The mathematics teachers teach grade levels ranging from middle school to high school in the United States and were enrolled in a state-funded professional development course offered over the summer. The Attentiveness in QRI responses were collected prior to and following an intervention designed to improve teacher candidates' and teachers' attentiveness. The timing of the instrument administration--prior to and following intervention--was intended to elicit a broad range of responses along the attentiveness continuum.

Instrument

The Attentiveness in QRI is being developed using the multi-phase process and general item framework described by Carney, Cavey, and Hughes (2017). See the Appendix for a description of the development process and purpose statement for the instrument. The instrument used for data collection for this study involved collecting constructed-responses (typically short paragraph) to item prompts. The instrument includes three quantitative reasoning tasks with two student work samples per task along with prompts related to (a) the mathematical intent of the task, (b) describing the student approach, (c) describing the student understanding, and (d) describing how the test taker would respond to the student based on the student work sample. The prompts vary slightly across items, depending on the task and student work sample.

More specifically, in terms of the instrument, clarification of the differences between the assessment item, the task, the student work sample, and prompts is provided. The term *assessment item* refers to a student task, its related student work samples, and item prompts. There are three total assessment items in the Phase 3 Attentiveness in QRI. *Task* refers to the quantitative reasoning problem presented to secondary (grades 6-12) mathematics students. There are three tasks (one for each assessment item) embedded in the Attentiveness in QRI, each with a different mathematical focus: (a) estimating a line of best fit, (b) proportional reasoning, and (c) estimating total distance traveled from a time and speed graph. The tasks were designed to be accessible across ability levels and to elicit a broad range of student approaches. In other words, they were designed so those without formal mathematical knowledge could still work productively toward correct solutions. *Student work sample* refers to the solution process (written and/or verbal) generated by the secondary mathematics student. Two authentic student work samples are presented with their accompanying prompts. The student work samples feature elements that can be considered correct, elements that can be considered incorrect, and/or relatively informal approaches to solving the task. This intentional selection of these types of student work samples, and the tasks which elicit them, is informed by a pedagogical lens of progressive formalization and the need for teachers to be able to understand and build upon students' informal understandings. *Prompts* refers to the wording used within the instrument to generate a response from the test taker. The prompts in the Attentiveness in QRI vary across assessment items. For instance, the prompts associated with estimating a line of best fit and estimating total distance traveled tasks primarily focus on the following four areas:

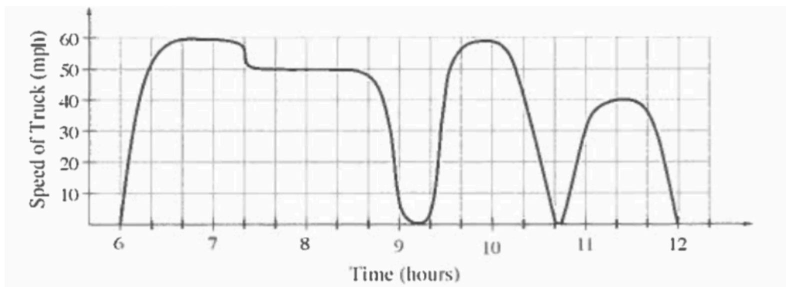
- The important mathematical idea(s) the task is targeting. (describing intent)
- Description of the approach used by the student. (describing student approach)
- Description of what the student response reveals about their understanding of the important mathematical idea(s) in this task. (describing student understanding)
- Description of how the test taker would respond to the student. (teacher response)

The proportional reasoning task, on the other hand, only has two prompts focused on explaining the similarities and differences in understanding between the two student work samples.

Figure 1 presents the item related to estimating total distance traveled from a time and speed graph. This assessment item represents the typical format of our attentiveness items and is very similar in structure to the line of best fit task.

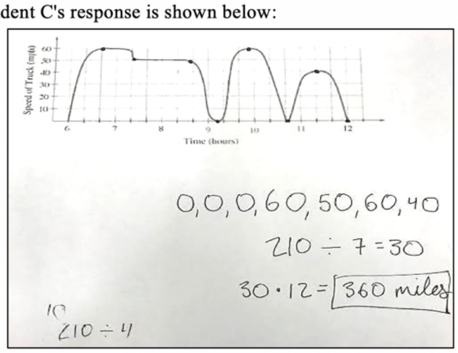
Before responding to the questions below, please work through the following task that was given to a class of high school students.

The Distance Problem.
A graph showing the speed of a truck in miles per hour over a 6-hour period is shown below. Estimate the total distance the truck traveled during this time. Briefly explain how you determined your estimate.



What important mathematical idea(s) is this task targeting?

Student C's response is shown below:

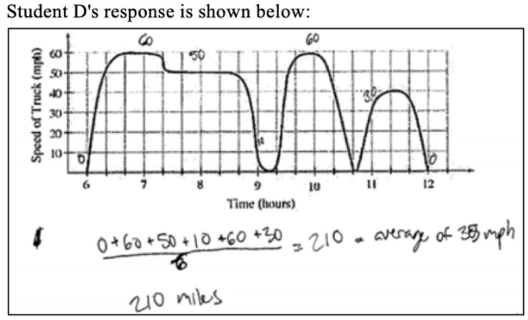


Please describe the approach that Student C used to solve this problem.

Please describe, in detail, what Student C's work indicates to you about their mathematical understanding.

If you were the teacher of Student C, please describe, in detail, how you would respond.

Student D's response is shown below:



Please describe the approach that Student D used to solve this problem.

Please describe, in detail, what Student D's work indicates to you about their mathematical understanding.

If you were the teacher of Student D, please describe, in detail, how you would respond.

Figure 1. Example assessment item consisting of the interpreting a time and speed graph to estimate distance traveled task, student work samples, and related prompts.

Data

In this sample, responses to the assessment prompts were typically one-to-three sentences in length. Certain prompts, such as describing the mathematical intent of the task, lend themselves to shorter phrases, while other prompts, such as describing how the test taker would respond to the student, lend themselves to short paragraph answers.

A total of 39 respondents had two full sets of responses; one set of responses was collected prior to participation in an attentiveness intervention and the other set of responses was collected following such participation. Three additional sets of responses were collected prior to the attentiveness intervention with teachers. This resulted in 81 sets of responses for analysis. Each set consisted of responses to 15 total prompts across the three assessment items, for a total of 1,215 responses to prompts. Our unit of analysis is at the response-to-prompt level.

Prior to analysis, data were de-identified for all respondent information, including respondent name and timing of assessment responses (i.e., pre vs. post attentiveness intervention). A randomly generated ID number was associated with each set of responses so the information could be linked back to respondents following coding.

Construct Map Development Process

The construct development process for teacher attentiveness occurred sequentially in four steps: (1) initial construct map and Attentiveness in QRI development based on the literature review, (2) structural coding of Attentiveness in QRI response data, (3) code mapping to categories, and (4) scoring and construct map diagram. Because completion of each step was a precursor to the next, analysis and results are reported for each step individually, rather than for the process as a whole. We have organized the information in this section around the four steps.

Literature Analysis. Prior to analysis of the data, the research team examined existing coding schemes, primarily from the professional noticing literature, to develop an initial conceptualization for attentiveness unrelated to a specific mathematical topic. The primary focus of the literature review was to identify potential shifts between levels of attentiveness. These findings are summarized below:

- Teacher shifts from providing nonspecific descriptions of a student approach to (a) detailing what is mathematically correct and incorrect about a solution strategy, and/or (b) differentiating between the aspects of the student response that are generalizable and non-generalizable (Bartell, Webel, Bowen, & Dyson, 2013; Jacobs, Lamb, & Philipp, 2010; Sherin & van Es, 2005; Stockero, 2008)
- Teacher shifts from under- or over-generalizing student's understandings to describing the student's understanding related to both the particular context's quantitative reasoning demands and to generalizations beyond the context (Bartell et al., 2013; Jacobs et al., 2010), and
- Teacher shifts from accepting or affirming student responses, pressing towards an answer without probing thinking, or asking generic questions to (a) posing questions or prompts that make use of the student's reasoning to further probe or help students clarify their thinking, and/or (b) offering suggestions for next problems that press particular aspects of quantitative relationships (Jacobs et al., 2010; van den Kieboom, Magiera, & Moyer, 2014).

Using the research literature to identify shifts towards increasing levels of attentiveness to student thinking provided some initial concepts for consideration and supported an assessment design which targeted three of the four primary components of attentiveness (describing student approach, describing student understanding, and teacher response). The fourth component, which represents a teacher's ability to identify and articulate the important mathematical idea(s) a task is targeting (describing intent), affords differentiation between those whose analysis of student work is appropriately aligned to the intent of the task and the grade level for which the task is used. The next phase of the construct map development, structural coding, focuses on analysis of test takers' qualitative responses to the Attentiveness in QRI prompts. For illustration purposes and to accommodate space constraints, we focus on the analysis of test taker responses elicited by assessment item prompts which target the attentiveness component of *describing student understanding*.

Structural Coding. Structural coding identifies and applies conceptual phrases to segments of data typically collected via interview or constructed-response (Saldaña, 2015). This step of analysis involved an inductive, open-coding approach during which the researchers identified emergent themes for attentiveness in test takers' responses specific to the instrument prompt. Two researchers worked in multiple, iterative rounds of independent and collaborative coding to develop, apply, and revise the coding schemes.

Prior to examining any data, the researchers brainstormed features of the desired exemplar responses each prompt was intended to elicit and a potential range of responses that the shifts identified above might forecast. This served to focus their lens on potential indicators of attentiveness before analysis began.

In round one of data analysis, the researchers independently read through a subset of the teacher responses and used open coding methods to identify emergent themes. They then met to discuss noticed themes and to reach consensus on an initial coding scheme.

For round two, the researchers again independently coded, taking note of when challenges with the initial coding scheme arose. When roughly a third of these data had been independently coded, the researchers met to check for inter-coder agreement and revise the coding scheme as necessary. Responses with different codes were discussed among coders until agreement was reached on the meaning of the code, and how that code was evident in the response.

This iterative process of identifying, independently applying, and collaboratively revising a coding scheme for a particular prompt often entailed three to four rounds before full consensus was met across all responses. During this process, coding tables were developed to describe and summarize the codes at the individual prompt level (15 prompts, therefore 15 tables). These tables include the code name, code description, and one to three exemplar responses.

The example presented comes from responses to the "Describe Student C's understanding" prompt for the estimating total distance traveled assessment item (see Figure 1). Student C's work sample provided multiple aspects of understanding, both correct and incorrect, to which test takers could attend. For example, Student C appears to understand that calculating the total distance involves multiplying the amount of time traveled by an estimate for the average speed during that time. Responses that addressed this idea were labeled with the code *Distance Calculation*. Student C's selection of speeds (*Points*) to average does not include explicit consideration of the amount of time for which that speed was traveled. In addition, the student's calculation of the amount of time traveled is inaccurate. Responses that addressed these ideas were labeled with the codes *Duration of Speed* and *Total Time*, respectively. Table 1 presents an overview of the full set of code names, descriptions, and selected exemplars related to the Student C understanding prompt. This example is typical of the coding tables that were developed across the 15 prompts.

In addition to the identification of codes, responses to this prompt also indicated there were respondents who focused on (a) describing what Student C did understand, (b) describing what Student C did not understand, and (c) describing both what Student C did and did not understand. Recognizing this pattern was useful in the next step, code mapping to categories.

Code Mapping to Categories. The next phase of analysis involved code mapping (Saldaña, 2015) across the collective code tables generated for common prompt types (four prompt types: describing intent, describing student approach, describing student understanding, teacher response). This involved identifying and naming categories and sub-categories across common prompt types that would encompass the prompt-specific codes in each individual table. Again, the desired shifts in attentiveness described by the literature helped to inform this work. The example provided here describes the development of the hierarchical categories for the five coding tables related to the *describing student understanding* prompts.

These coding tables for the describing student understanding prompts were examined to identify categories and sub-categories that consistently threaded across the codes, regardless of the mathematical task and student work samples. Two primary categories were identified: *mathematical focus* and *perspective on understanding*. The mathematical focus category primarily related to the specificity of the claim made about student understanding in relation to the mathematical intent of the task. There were four subcategories; supported specific claim, supported non-specific claim, unsupported claim, and no claim. The perspective on understanding category related to whether the claim about student understanding focused on what the student understood and/or did not understand. The two subcategories were affordances <u>and</u> constraints, which featured a dual perspective and involved inference language related both to what the student did and did not understand, while the subcategory affordances <u>or</u> constraints involved a single perspective and included inference language related either to what the student did or did not understand.

Table 1. Coding table for the student C understanding prompt for the estimating total distance traveled assessment item.

Estimating total distance traveled from a time and speed graph: Describing student understanding		
Code Name	**Code Description - Candidate response...**	**Code Exemplar(s)**
Describes aspects of what the student does understand		
Points	Indicates that the student point selection is likely not random but they don't necessarily articulate the likely reasoning.	- I am unsure of how they chose there [their] 7 points to look at though.
Distance Calculation	Indicates the student knows or understands a velocity (or speed) should be multiplied by a time value to get distance.	- They understand that to find distance they need to multiply a time by a speed. - They have the basic idea right of multiplying time and velocity to get distance.
Average	References student understanding of average (related to calculation involving the 7 points)	- They understand how to average things in this sense of add them up and then divide by how many there are
Overgeneralize	Overgeneralizes student understanding in ways that are not supported by the evidence.	- Student C understands how the axes and their units relate to each other, this is shown by multiplying the mph by the hours to get miles.-
Generic	Fails to explicitly reference either the student approach or connect to the task intent	- It actually says a lot, this is a very interesting approach to this problem - How to read a graph
Describes aspects of what the student does NOT understand		
Duration of Speed	References student not taking into account the length of time a particular speed was traveled.	- They, however, missed some critical ideas about duration the vehicle was at certain speeds - First and foremost, they do not account for the "unevenness" of their sections, so averaging them does not give a very good idea of the true average speed
Total Time	References student mistake of using 12 (the last time marked on the x-axis) instead of 6 (the number of hours traveled) as the number of hours traveled.	-They misread the graph, as it starts at 6 so there are 6 total hours not 12.
Alternative Approach	References an alternative approach the student could have used to solve the problem.	- They used points that represented change in speed rather than average speeds.
Overgeneralize	Overgeneralizes student (lack of) understanding in ways that are not supported by the evidence.	- I think Student C was in a hurry to get this problem done because they didn't pay much attention to the starting point on the x-axis and made some simple mistakes. It also would appear that they do not understand what they are doing.
Formal	Identifies potential formal mathematical (mis)understandings related to the task, but the language used (e.g. area under the curve) is beyond the scope of the task.	- Student C does not understand the relationship between velocity and distance, or at least does not connect the ideas of area under a velocity curve and distance travelled.
Disparate	References (mis)understandings involving related ideas (e.g., e.g. acceleration, position, direction of movement, etc.) that are not the focus of the task or explicitly evidenced in the student's work.	- *no exemplars*
Incorrect Mathematics	Indicates an incorrect understanding of mathematics, whether in the task itself or in the student's work.	- This indicates that they don't know how to interpret values based on the lines/curves of the graph.

The coding tables for the describing student understanding prompt were used to inform the development of the categories and subcategories related to the student understanding component of attentiveness. However, when assigning a subcategory to a response, we found that additional considerations were necessary to account for the mathematical intent of each task. For example, the code *duration of speed* was subcategorized as a supported specific claim because the student process for approximating the average speed – including the amount of time a particular speed was traveled – was not only supported, but it was considered critical to the intent of the task and a key aspect of the understanding demonstrated by Student C’s work sample. Similarly, the code *total time* tended to be subcategorized as supported non-specific because while it was evident the student did not account for the x-axis starting at six instead of zero, this was considered a less critical understanding in terms of the task’s mathematical intent. Additional assignments of the specific versus non-specific subcategory required a more nuanced examination and discussion of the task’s mathematical intent amongst coders, as a particular code from one coding table could be classified as specific while, were the student task for the assessment item to be changed, the same code could be classified as non-specific.

Table 2 presents the categories and sub-categories identified for the describing student understanding prompt and includes exemplar responses to the assessment item involving the estimating total distance traveled task (see Figure 1) for Student C. This example is typical of the category tables developed across the four prompt types and supports the work for the final step of construct map development – scoring and construct map diagram.

Scoring and construct map diagram. The next step in the process involved assigning numerical scores to the common categories and sub-categories given to each response. Assigning numerical scores to categorical labels has both benefits and drawbacks. One serious drawback is that when a numerical score is assigned to qualitative data, a significant amount of information is lost in terms of the depth and detail of the response. However, a benefit of assigning a numerical score to the category labels is the ability to see more general trends in respondents’ level of attentiveness. Sometimes the level of detail provided by coding and categorizing data can make it difficult to generate a more holistic picture of teachers’ attentiveness. By attending to both aspects in our construct map development process, our qualitative interpretations of the overall quantitative score present a more nuanced perspective of teachers’ attentiveness.

Table 2. Common categories and subcategories for describing student understanding prompts including exemplars from the estimating total distance traveled from a time and speed graph for Student C.

Category Name	Subcategory Name	Sub-category Description	Exemplars
Mathematical Focus	Supported specific claim	Makes an accurate inference about student understanding (or lack of understanding) supported by evidence that directly addresses the mathematical intent of the task and impacts student's ability to productively engage with the task.	They seem to understand that the question needs to be answered given the average speed of the truck over an interval of time, however they seem to be missing information about the increasing/decreasing speeds of the truck over different time intervals that would contribute to the overall distance traveled.
	Supported non-specific claim	Makes an accurate inference about student understanding (or lack of understanding) that is supported by evidence but does not address the mathematical intent of the task.	They understand how to find the average of a set of data.
	Unsupported claim	Makes an inference about student conceptual understanding (either what is understood or not understood) that is not supported by the evidence from student work. This includes overly formal and incorrect mathematics.	They do not understand that the area under the curve of velocity is the position.
	No claim	Describes the student's mathematics without making an inference about conceptual understanding or knowledge	*No exemplars for Student C*
Perspective on Understanding	Affordances AND Constraints	The teacher or candidate uses inference language about what the student understands AND does not understand.	Student C understands how the axes and their units relate to each other, this is shown by multiplying the mph by the hours to get miles. Though they also read the x-axis wrong by assuming it started at zero then just looked at the last point of the graph to get 12 hours.
	Affordances OR Constraints	The candidate uses inference language about what the student understands OR does not understand.	They didn't think about whether if traveling at different speeds for different amounts of time would matter (constraint perspective).

Following the development of the common categories and sub-categories, research project personnel discussed the process of converting these to numerical scores. Within the mathematical focus category, responses categorized as supported specific claim were assigned two points, while responses categorized as supported non-specific claim were assigned one point. Responses categorized as unsupported or no-claim were assigned zero points. The point levels assigned to these sub-categories were based both on our literature review and their likely impact on classroom instruction. For example, an unsupported claim made by a teacher could result in the teacher misdirecting the student, or at the very least, offering a suggested next step that would potentially be confusing for the student. Responses categorized as either an unsupported claim or as a no-claim were determined to be equally less beneficial and potentially an obstruction to supporting student understanding. Within the perspective on understanding category, responses categorized as affordances and constraints were assigned one point. Responses categorized as affordance or constraints were assigned zero points. Similar to the mathematical focus category, the point levels assigned to the perspective on understanding sub-categories were primarily based on their likely impact on classroom instruction. In this case, responses categorized with both affordance and constraint were determined to be potentially more beneficial to a student's understanding than responses categorized as either an affordance or constraint. Teachers who recognize both the productive and unproductive ideas in student work samples were seen as likely to build upon the student thinking in a way that would also target the mathematical intent of the task. While teachers who only describe either the affordance or constraints were seen as likely to praise student work or correct the student misunderstanding, respectively.

Figure 2 presents the construct map for the describing student understanding component of attentiveness, based on the identified categories and sub-categories, and includes the point assignments to each subcategory. The construct map diagram presented differs from Wilson's (2005) version of a construct map, as typically depicted in a table similar to Table 2, but the two construct maps are similar in that they present increasingly sophisticated ways of reasoning within a construct and have the potential to assist with meaningful score interpretation.

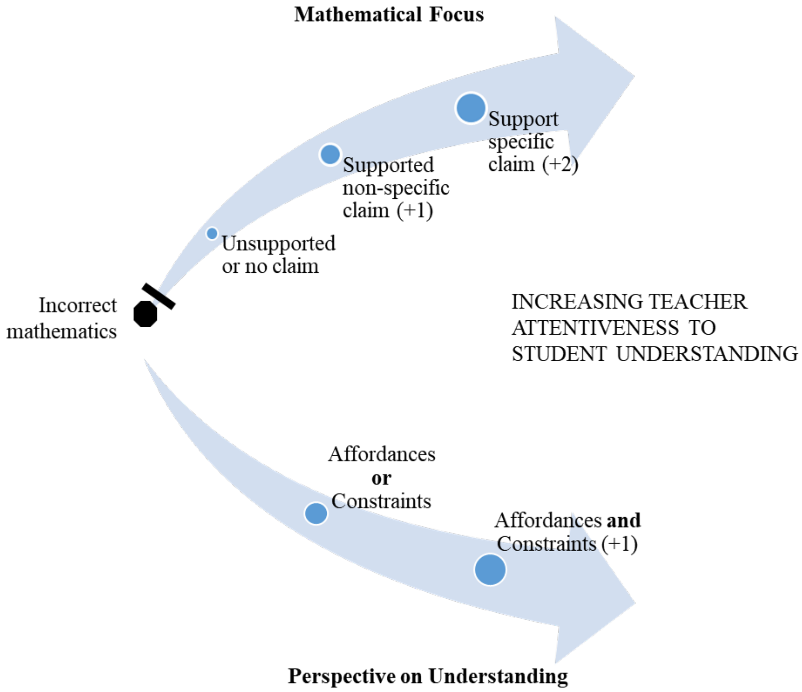


Figure 2. Describing student understanding construct map diagram for teacher attentiveness.

The construct map diagram in figure 2 illustrates hypothesized relationships and, in conjunction with the category and subcategory descriptions from table 2, provides the describing student understanding construct map for teacher attentiveness. It depicts the categories of mathematical focus and perspective on understanding as separate components of teachers' attentiveness to student understanding. However, it is important to note that we do not anticipate they will function as separate dimensions in our measurement model because they are so intertwined and the nature of the relationships between these categories needs to be further examined. We represent them separately in the construct map diagram because the categories and subcategories are worth capturing for score interpretation and use in instructional applications. It is also worth noting that subcategories within a category are not necessarily aligned with the subcategory in an adjacent category. Instead, the subcategories are aligned with the arrow representing increasing/decreasing attentiveness. For example, some responses were categorized as affordances <u>and</u> constraints, the highest subcategory in the perspective on understanding category, while also being categorized as supported non-specific in terms of the mathematical focus. Other responses, due to the test taker's use of incorrect mathematics, did not score on the mathematical focus component at all, though they still provided evidence of a perspective on understanding. In terms of a teacher's ability to respond productively to students' thinking, a perspective that considers both what students understand and do not understand could potentially lead to dialogue with students where the teacher's incorrect mathematics is rectified.

Discussion

Developing a construct map for attentiveness serves multiple purposes and has numerous applications in terms of assessment development, validation, and score interpretations and uses related to instruction. We address three major purposes and their applications: score interpretation, construct representation, and further assessment development.

Score Interpretation

The construct map development process described in this chapter can be viewed as a response to recent calls for assessment developers to provide for substantive qualitative descriptions in relation to quantitative score continuum (Shepard 2017), as the development of construct maps is an important first step in this work. The identified categories and subcategories for item responses and their relationships can be used to generate initial statements related to respondents' qualitatively different levels of attentiveness. For example, we anticipate scores in the upper third of the scoring range will be associated with respondents who consistently attend to the mathematics concepts evidenced in student work that involves articulation both of what the student understands and does not understand. This is based on scores in the upper third of the scoring range primarily being associated with responses from the subcategories of supported specific claim and affordances and constraints. Similarly, we anticipate scores from the middle third of the scoring range will be associated with respondents who explicitly attend to mathematics concepts evidenced in student work (subcategory supported specific claim) but only articulate what the student understands or does not understand (subcategory affordances or constraints), or they may have less specific attention to the mathematical intent of the task (subcategory supported non-specific) but focus on articulating what the student understands and does not understand (subcategory affordances or constraints). By articulating the likely combinations of response sub-categories and point options illustrated by the construct map diagram, qualitative descriptors of respondents' scores are readily generated.. These descriptors can then be further examined in relation to empirical data and refined to develop the qualitative ordering of respondents within the construct map. This work supports valid inferences regarding teachers' level of attentiveness based on their quantitative scores. Meaningful score interpretations and valid inferences also contribute to research being done in both formative assessment and professional noticing, as examining the relationships between attentiveness scores and the quality of noticing and assessment during mathematics instruction could reveal unexplored connections worthy of further study.

Construct Representation

The attentiveness construct draws from multiple aspects of the research literature, such as mathematical knowledge for teaching, progressive formalization, and professional noticing. The generation of construct maps that summarize and generalize the range of reasoning captured by our constructed response items has two important applications. First, it helps to describe what is being assessed in our operationalized version of the construct, assisting with construction representation validation processes. As previously described, operationalizing a construct necessarily restricts the boundaries of that construct beyond what might be included in a theoretical description. Our construct map describes which aspects are and are not included within the operationalization. For example, the *describing student understanding* construct map clarifies the aspects of understanding teachers focus on in their response – mathematical focus and perspective on understanding. Providing a summary description of this work to assessment users can help them better understand the construct focus of the assessment to determine if it is appropriate for their needs. A second potential application involves the use of the construct map by mathematics teacher educators in instructional or professional development settings. Because the construct map illustrates and describes the components of attentiveness, it has the potential to support teacher learning and understanding of the construct itself and to frame and inform activities designed to support growth and development of attentiveness in educators.

Further Assessment Development

Lastly, regarding further assessment development, the process of categorizing and scoring constructed-responses can be time-consuming and expensive when applied to a large data set. One potential solution is further item development through the construction of selected-response items (for an example of this application see Carney, Cavey, and Hughes, 2017). The coding and category tables describe the range of reasoning that occurs in response to a particular assessment task, student work sample, and related prompt. Describing the range of reasoning elicited by constructed response items makes it possible - as assessment developers – to identify authentic responses for use in the development of selected-response items that represent this range of reasoning found in the constructed-response version. This change in item type provides for an more efficient scoring process. By incorporating authentic responses generated by teachers or teacher candidates, we can be more confident that our selected-response items have the potential to (a) generate a full range of scores along the attentiveness continuum, (b) validly measure attentiveness in meaningful ways, and (c) inform future refinement of our construct operationalization. The codes and construct maps will also provide a point of comparison for data from cognitive interviews on the response process for the selected-response items to assist in determining how the change in item format potentially influences and/or further bounds the construct.

The purposes identified here are just a few examples of the usefulness of developing initial construct maps for applications related to assessment development, scoring, and validation, in addition to further developing our understanding of the operationalized version of the construct.

Limitations

It is important to realize that any time a construct is measured, the potential reasoning space encompassed by the theoretical construct becomes limited by the operationalization. The tasks, student work samples, and prompts all serve to provide a frame within which teachers can demonstrate their level of attentiveness. But we must realize these selections also limit the ways in which this demonstration occurs. Therefore, our construct map for attentiveness to student understanding is limited to the assessment from which is was created. It is possible that additional categories or subcategories of responses would emerge, had we operationalized the construct in a different manner.

Current & Future Work

In this section, we briefly describe our current and future work related to the Attentiveness in QRI and the ways in which the development of a construct map for attentiveness will inform other work related to the VCAST project, thereby illustrating some of our points in the discussion.

The construct map for attentiveness was recently used to inform modifications to the Attentiveness in QRI. In particular, the process of developing the construct map resulted in greater awareness of our intent for test takers to consciously attend to the context of item prompts. Explicit articulation of this intent occurred while code mapping to categories, as the code-mapping process revealed key similarities in the quality of responses across common prompt types. That is, for all four prompt types (describing intent, describing student approach, describing student understanding, teacher response), we realized we consistently assigned higher scores to responses that were informed by the mathematical level at which the task was used,. This insight resulted in revising item prompts in an attempt to make our intention more clear to test takers. For example, the prompt "What mathematical ideas is this task targeting?" would be adjusted to "What Algebra I mathematical ideas is this task targeting?" for a constructed-response prompt (describe intent) when a task designed for Algebra I students is used.

In addition, the revised version of the Attentiveness in QRI is now comprised of selected-response items and addresses a broader range of quantitative reasoning tasks. The process of developing the construct map contributed to our ability to construct authentic responses. The exemplars included in our code tables were particularly helpful in the identification of responses as it allowed us to confidently choose appropriate level responses for prompts. To further develop the Attentiveness in QRI, we plan to use Rasch modeling to study how the items are functioning psychometrically. Beyond testing for item functionality, we plan to draft substantive qualitative descriptions along the quantitative score continuum for attentiveness.

Finally, we intend to apply the construct map to assess the attentiveness intervention developed for the VCAST project. As part of the intervention, teacher candidates work through an online learning module which includes prompts directly related to operationalized components of the attentiveness construct (e.g. describing student understanding). Applying the construct map in this way will potentially lead to our ability to fine-tune our operationalization of attentiveness and thereby influence both the attentiveness instructional interventions and the Attentiveness in QRI.

Conclusion

The goal of this chapter was to illustrate how our instrument development process aimed at assessing teachers' attentiveness to secondary students' quantitative reasoning might provide insight to other researchers as they engage in their own assessment development efforts. This chapter builds upon Wilson's (2004) explication of construct maps by providing an example embedded within the assessment of a particularly complex construct. The focus includes, but also goes beyond, how construct maps can be used for assessment development. Generating construct maps and their associated category/code descriptors through analysis of authentic qualitative responses to constructed-response items has the potential to provide clarity on how the resultant operationalization bounds the construct, further develop understanding of theory, and provide a basis for the development of substantive qualitative interpretations of assessment scores along a quantitative continuum.

References

Ball, D. L., Thames, M. H., & Phelps, G. C. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389-407.

Bartell, T. G., Webel, C., Bowen, B., & Dyson, N. (2013). Prospective teacher learning: Recognizing evidence of conceptual understanding. *Journal of Mathematics Teacher Education, 16*(1), 57-79.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31.

Carney, M., Cavey, L., & Hughes, G. (2017). Assessing teacher attentiveness: Validity claims and evidence. *Elementary School Journal, 118*(2), 281-309.

Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht, The Netherlands: Reidel.

Freudenthal, H. (1991). *Revisiting mathematics education*. Dordrecht, The Netherlands: Kluwer.

Gravemeijer, K., Cobb, P., Bowers, J., & Whitenack, J. (2000). Symbolizing, modeling, and instructional design. In P. Cobb, E. Yackel, & K. McClain (Eds.), *Symbolizing and communicating in mathematics classrooms: Perspectives on discourse, tools, and instructional design* (pp. 225-274). Mahwah, NJ: Routledge.

Gravemeijer, K., & van Galen, F. (2003). Facts and algorithms as products of students' own mathematical activity. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 114-122). Reston, VA: National Council of Teachers of Mathematics.

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: Theory and Practice, 15*(2), 273-289.

Jacobs, V. R., Lamb, L. L., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education, 41*(2), 169-202.

Lampert, M. (2009). Learning teaching in, from, and for practice: What do we mean? *Journal of Teacher Education, 61*(1-2), 21-34.

Lampert, M., Franke, M. L., Kazemi, E., Ghousseini, H., Turrou, A. C., Beasley, H., . . . Crowe, K. (2013). Keeping it complex: Using rehearsals to support novice teacher learning of ambitious teaching. *Journal of Teacher Education, 64*(3), 226-243.

Learning Mathematics For Teaching. (2005). Mathematical knowledge for teaching measures. Ann Arbor, MI.

Lobato, J., Ellis, A. B., & Charles, R. I. (2010). *Developing essential understanding of ratios, proportions, and proportional reasoning for teaching mathematics in grades 6-8*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all*. Reston, VA :National Council of Teachers of Mathematics.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington DC: National Academy Press.

Saldaña, J. (2015). *The coding manual for qualitative researchers*: Thousand Oaks, CA: Sage.

Shepard, L. A. (2018). Learning progressions as tools for assessment and learning. *Applied Measurement in Education, 31*(2), 165-174.

Sherin, M. G., & van Es, E. A. (2005). Using video to support teachers' ability to notice classroom interactions. *Journal of Technology and Teacher Education, 13*(3), 475-491.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-22.

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking & Learning, 10*(4), 313-340.

Stein, M. K., Smith, M. S., Henningsen, M., & Silver, E. A. (2000). *Implementing standards-based mathematics instruction: A casebook for professional development*: New York, NY: Teachers College Press.

Stockero, S. L. (2008). Using a video-based curriculum to develop a reflective stance in prospective mathematics teachers. *Journal of Mathematics Teacher Education, 11*(5), 373-394.

High-Leverage Practices. (2018). TeachingWorks. Retrieved from http://www.teachingworks.org/work-of-teaching/high-leverage-practices

Treffers, A. (1987). *Three dimensions: A model of goal and theory description in mathematics instruction – The Wiskobas Project*. Dordrecht, The Netherlands: Reidel.

van den Kieboom, L. A., Magiera, M. T., & Moyer, J. C. (2014). Exploring the relationship between K-8 prospective teachers' algebraic thinking proficiency and the questions they pose during diagnostic algebraic thinking interviews. *Journal of Mathematics Teacher Education, 17*(5), 429-461.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, N.J.: Lawrence Erlbaum Associates.