Operationalizing the Science Practices Teacher Questionnaire User Guide

Eric R. Banilower
Meredith L. Hayes
A. Zehra Jaffri
Noell L. Egeland

August 2016

Horizon Research, Inc.
326 Cloister Court
Chapel Hill, NC 27514

Disclaimer

Operationalizing the Science Practices Teacher Questionnaire User Guide was prepared with support from the National Science Foundation under grant number DGE-1445543. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Suggested Citation

Banilower, E. R., Hayes, M. L., Jaffri, A. Z., & Egeland, N. L. (2016). *Operationalizing the science practices teacher questionnaire user guide*. Chapel Hill, NC: Horizon Research, Inc.

TABLE OF CONTENTS

Page

OVERVIEW

In 2013, the National Research Council (NRC) released the report *Monitoring Progress Toward Successful K–12 Education: A Nation Advancing?* The report calls for a national indicator system that could be used to improve science, technology, engineering, and mathematics (STEM) education in the United States, and sets forth 14 indicators related to students' access to quality learning, educators' capacity, and policy and funding initiatives. Horizon Research, Inc. (HRI) received a grant from the National Science Foundation (grant number 1445543) to develop a measure that could be used cost effectively at a large scale for Indicator 5—the extent to which classroom instruction aligns with the Framework for K–12 Science Education (National Research Council, 2012) and the subsequently released Next Generation Science Standards (Next Generation Science Standards [NGSS] Lead States, 2013).

This user guide describes the development process of this measure, which consists of two sets of closed-ended questionnaire items. The first set is intended to assess the extent to which teachers' objectives for science instruction include students gaining proficiency with the science practices. The second set is intended to assess how often teachers engage students in the practices as part of their science instruction. In addition, this user manual describes the measurement properties and appropriate uses of the questionnaire items.

DEVELOPMENT OF THE QUESTIONNAIRE ITEMS

An overview of the development process is shown in Figure 1. This section of the user guide describes the major steps used to develop the questionnaire items.

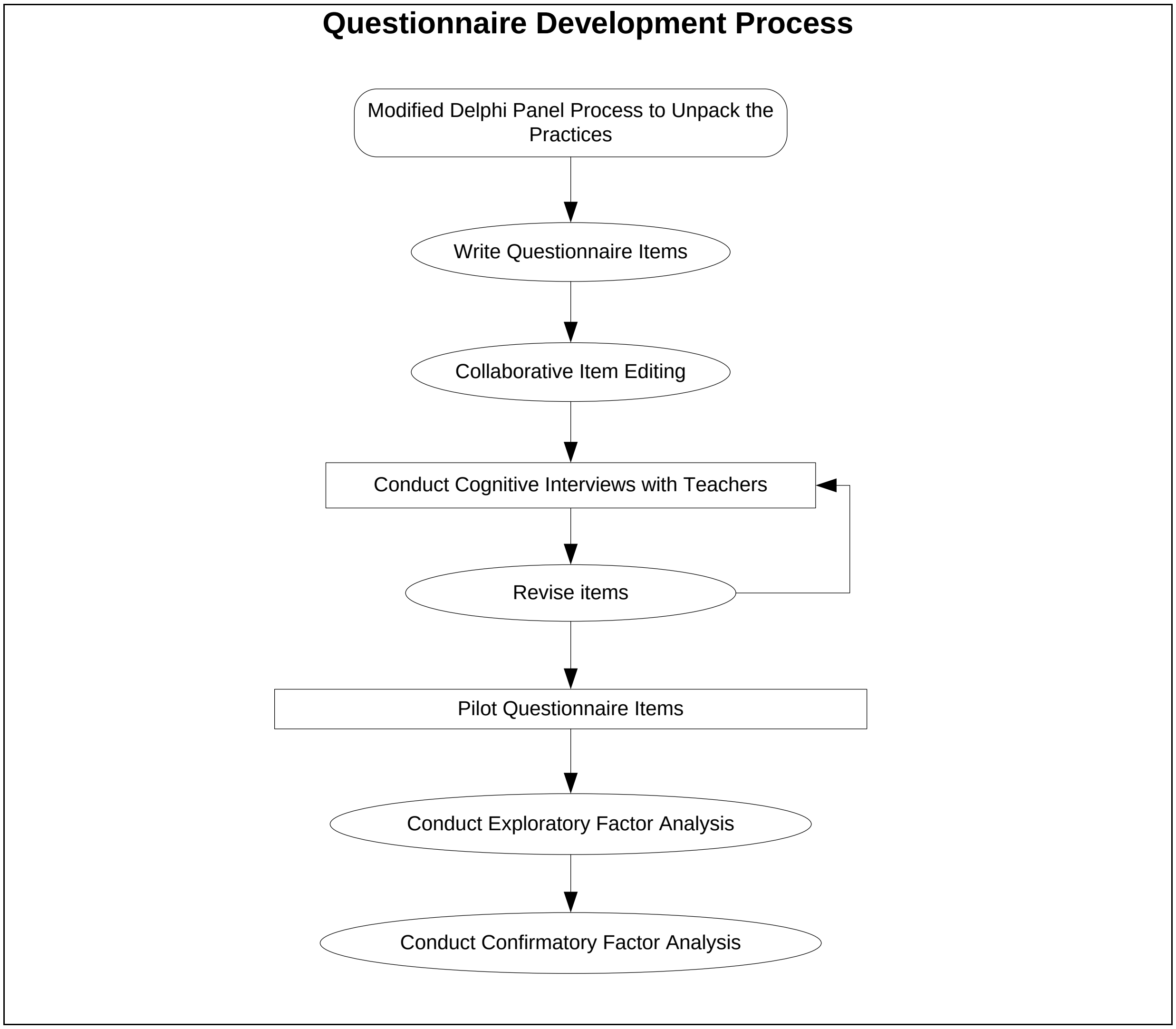

Figure 1

Unpacking the Practices

Two methods of data collection were used to examine the science practices. First, an extensive review of existing literature was conducted to identify and summarize current research and practice-based knowledge focused on engaging students with the science practices described in the *Framework for K–12 Science Education* and the NGSS. Second, an expert panel was convened to "unpack" the science practices, identifying key elements of the practices in different grade bands and areas of science.

The literature review identified and summarized current research and practice-based knowledge focused on engaging students with the science and engineering practices described in the *Framework for K–12 Science Education* and the NGSS. A search utilizing the ERIC and Google scholar databases was conducted to find articles on the practices published since the release of the *Framework* in 2011.

Key words and phrases, such as *science and engineering practices*, *NGSS*, *models*, *argumentation*, and *explanations* were used in conjunction, and the initial search resulted in 76 empirical and practitioner-oriented articles and conference papers. These resources were reviewed more carefully to determine if the practices are described generally or if they are elaborated upon; resources that include only general descriptions of the practices were removed from the literature pool. Articles in practitioner journals that provide stand-alone lesson plans without any discussion of the practices were also excluded from the literature review. The remaining 47 articles were coded by:

1. Science and/or Engineering
2. Practices Addressed
3. Grade levels (K–5, 6–8, 9–12)
4. Nature of the Journal (practitioner-oriented, researcher)

Following the coding process, all articles were analyzed for whether a purpose for the practice is described—either a purpose related to the enterprise of science or a purpose in science education. In addition, each article was examined for whether it provides explicit guidance on “key features” of a practice, i.e., what is important for students to experience when engaging in the practice.

The expert panel was composed of individuals with backgrounds and experiences in different grade levels and content areas. In addition, panelists had varied levels of involvement with the NGSS, including some panelists who served on the writing team.

The panel process involved four rounds of questions that were emailed to the panelists. Questions in Rounds One and Two addressed four practices: (1) constructing explanations; (2) engaging in argument from evidence; (3) obtaining, evaluating, and communicating information; and (4) using mathematics and computational thinking. Rounds Three and Four focused on the remaining practices: developing and using models, asking questions, planning and carrying out investigations, and analyzing and interpreting data.

For each science practice, panel members were asked to provide a detailed example of instruction engaging students in the practice as part of meeting an NGSS performance expectation or learning about a particular topic area. Panelists were then asked to identify the key elements of what students were doing in their description that are essential for helping students master that practice. In addition, panelists were asked to specify whether/how the key elements differed by grade band and science content area. The key elements offered by the panel, along with other key elements identified in the literature review, were compiled.

In the subsequent round, panelists were asked to provide feedback on the set of compiled key elements, noting whether they: (a) agreed that the element is essential for helping students master the practice; (b) agreed that the element is essential for helping students master the practice if modified (providing the edits needed); or (c) disagreed that the element is essential for helping students master the practice. Panelists also had the opportunity to add key elements that were missing from the compiled list and to specify at which grade bands students should experience each of the listed elements.

A fuller description of the unpacking process and the final set of key elements for each practice can be found in *What Does "Implementing The NGSS" Mean? Operationalizing the Science Practices for K–12 Classrooms* developed by this project and available on HRI's website at http://www.horizon-research.com/operationalizing-the-science-practices/.

Item Development

HRI staff drafted items individually then met to edit them collaboratively, resulting in an initial set of 42 items covering all eight practices. Nine items asked how much emphasis teachers place on practice-related learning objectives for their students. For example: students gaining proficiency in developing scientific models (physical, graphical, or mathematical representations of real-world phenomena) to help develop explanations, identify questions, make predictions, or communicate to others. The response options for these items were: "None," "Minimal emphasis," "Moderate emphasis," and "Heavy emphasis."

Thirty-three items focused on the frequency with which teachers have students engage in specific practice-related activities, such as determining what data would need to be collected in order to answer a scientific question. The response options for these items were: "Never," "Rarely (for example: a few times a year)," "Sometimes (for example: once or twice a month)," "Often (for example: once or twice a week)," and "All or almost all science lessons."

Teachers in grades K–12 were then recruited to participate in a first round of cognitive interviews (Desimone & Le Floch, 2004; Hamilton, Nussbaum, & Snow, 1997) around these items. Cognitive interviews help ensure item validity, as teachers' responses reveal whether they interpreted the items as intended. The cognitive interview process involved a teacher reading and thinking aloud as s/he answered the survey items. The interviewer then probed regarding particular item features (e.g., the clarity of the wording, how terms were interpreted). For example, teachers were presented with an item that asked how frequently students revise their explanations for real-world phenomena based on additional evidence, and then prompted by the interviewer to describe how they interpreted the term "explanations" in this context. Additional sample interview questions are shown in Figure 2.

Teacher Questionnaire Cognitive Interview Protocol

Prologue Script:

Thank you for participating in this pilot study of our teacher questionnaire. During this interview, I'd like for you to read each questionnaire item aloud as well as think aloud as you consider answering each item. I'd like to talk with you about all of the items on the survey, in particular items that you identify as confusing or difficult. Please be assured that there are no "right" or "wrong" answers. Do you have any questions before we begin?

Procedure:

Ask teacher to read aloud and "think aloud" as they read the questions and answer choices.

SAMPLE Cognitive Interview Questions:

- Please identify any words/terms in the item that you did not understand.
- Please describe anything that made this item confusing or difficult to answer.
- We are interested in how you interpreted the response options for this question. I see that your response to this question was [answer]. Why did you select this response? Can you give me an example of what you do in the classroom that you considered as aligning with this question?
- We are concerned that teachers may not understand what we mean by "model." What would you count as a "model"?
- Were there things you weren't sure if you should count as "analyzing and interpreting data?" If so, what?

Figure 2

After a first round of 12 interviews, HRI staff met to discuss teachers' feedback. Some items were substantively revised and some new items were drafted as a result of the first round of cognitive interviews. The cognitive interviews revealed teachers had, overall, highly varied interpretations of the practices and the colloquial definitions of scientific terms caused teachers to interpret items in unintended ways. For example, several teachers, particularly in the elementary grades, indicated that they would not want to have students engage in argumentation because they wanted a positive classroom environment. In addition, middle and high school teachers tended to conflate argumentation with constructing explanations. Teachers also often used claims and explanations interchangeably. To provide further clarification, parenthetical definitions of these terms were added to relevant items.

In contrast, items related to the practices of conducting investigations, and collecting and analyzing data tended to have fewer interpretation issues, perhaps because teachers are more familiar with these practices. Still, there was one issue with items about planning and conducting investigations that was unique to this practice. There are separate key elements for this practice related to designing investigations, conducting investigations, and revising the design of investigations. During the cognitive interviews, teachers questioned whether it "counted" if students were only responsible for particular elements. For example, when reacting to items about implementation and revision, they were not sure whether to include instances when students implemented investigations they had not designed or revised procedures that they had not initially developed. For these reasons, we included parenthetical information in several items to make it clear that teachers should consider students' engagement with the element (e.g., revising the design of an investigation) regardless of whether the students developed the initial design.

Items about the practice of using mathematics and computational thinking were particularly problematic. Elementary grade teachers tended to interpret computational thinking as doing computation (i.e., addition, subtraction, multiplication, division). Secondary teachers tended to interpret the term as meaning using a computer, regardless of purpose (e.g., for word processing).

New and substantially revised items then went through a second round of cognitive interviews. This round of 13 interviews resulted in minor edits to some items; items that the interviews indicated had substantive issues were dropped from the pool. The final set of items to be piloted contained 44 items—10 about instructional objectives and 34 about instructional practices.

PILOT

Recruitment

The original goal of the project was to recruit 3,000 teachers of science (1,000 per grade band) to pilot the questionnaire items. This design would allow both exploratory and confirmatory factor analyses (EFA and CFA, respectively) to be conducted with sufficient sample sizes at each grade band. To incentivize participation, teachers were offered a $15 honorarium for completing the pilot questionnaire. Recruitment took place in the first part of 2016.

Three strategies were used to recruit teachers. First, HRI maintains an email list of teachers who expressed interest in learning about opportunities to participate in education research studies—1,539 teachers were contacted and 553 registered for the pilot. Second, HRI conducted an email marketing campaign with MCH Strategic Data. MCH emailed over 80,000 teachers with information about the study, including a link to the online registration form. This campaign resulted in an additional 876 teachers registering for the study. Third, the National Science Education Leadership Association (NSELA) distributed information about the study to its membership, asking them to share this information with teachers with which they work. This effort yielded an additional 413 registrants.

Pilot Administration

In total, 1,842 teachers registered to participate in the pilot, roughly evenly split among elementary, middle, and high school teachers. Registration data were screened to ensure that respondents were currently teaching science in a public, private, or Department of Defense school in the United States, resulting in the exclusion of one case.

The pilot was administered online in May of 2016. The instrument contained 10 items asking about instructional objectives and 34 items about instructional practices. Teachers responsible for teaching science to more than one class of students were instructed to respond about the first science class they taught that week. Of the 1,841 teachers sent the questionnaire, 1,455 responded.

Responses were examined for incomplete and duplicate submissions, as well as evidence that items were not thoughtfully answered—indicated by the presence of response sets (i.e., selecting the same response option for all items) or completing the questionnaire in a very short amount of time. Consequently, 229 cases were dropped resulting in a final sample size of 1,226 cases, including 314 elementary, 449 middle, and 463 high school teachers.

Descriptive data on the teachers in the study are shown in Table 1. Teachers of elementary classes spanned grades K–5, though grades 4 and 5 were more common. About three-quarters taught in self-contained classes and a similar percentage had at least some familiarity with the NGSS. The vast majority taught in public schools, with a large majority from the South and the West. Middle and high school teachers in the study tended to teach in public schools. They were also more evenly distributed across the country and tended to have more familiarity with the NGSS than the elementary teachers in the study. Although a majority of the middle and high school teachers taught biology/life science, chemistry, Earth/space science, and physics were each taught by a substantial proportion of teachers in the study.

Table 1
Characteristics of the Pilot Test Sample, by Grade Band

	Percent of Teachers		
	Elementary (N = 314)	**Middle (N = 449)**	**High (N = 463)**
Grades Taught in 2015–16†			
Kindergarten	14	--	--
1st grade	14	--	--
2nd grade	19	--	--
3rd grade	25	--	--
4th grade	34	--	--
5th grade	41	--	--
6th grade	--	28	--
7th grade	--	50	--
8th grade	--	52	--
9th grade	--	--	55
10th grade	--	--	76
11th grade	--	--	80
12th grade	--	--	80
Self-Contained?			
Yes	73	0	0
No	27	100	100
Topics Included in Curriculum†			
Biology/Life Science	92	76	60
Chemistry	25	58	50
Earth/Space Science	90	73	30
Engineering	36	33	11
Physics	48	63	36
Familiarity with NGSS			
Not at all familiar	27	10	10
Slightly familiar	33	28	33
Moderately familiar	28	43	40
Very familiar	12	19	17
School Type			
Public	96	97	91
Private	4	2	8
Other	1	1	2
Region			
Midwest	11	21	23
Northeast	6	21	19
South	35	29	26
West	47	29	31

† Percentages may add to more than 100 as teachers could select multiple categories.

Teachers who were responsible for teaching science to multiple classes were asked to focus on a single class for the purpose of this study. They were instructed to select the first science class they taught that week as a way of reducing the potential bias of teachers selecting their best class. Table 2 shows the courses about which middle and high school teachers responded. Just over half of middle school teachers responded about a general or integrated science course. The remainder were evenly split among Earth, life, and physical science courses. At the high school level, about 4 in 10 teachers responded about a biology class, 3 in 10 about a chemistry class, and somewhat more than 1 in 10 about a physics class.

Table 2
Courses Represented in Pilot Sample, by Grade Band

	Percent of Teachers
Grades 6–8 Science (N = 449)	
General or Integrated Science	51
Life Science	17
Earth Science	16
Physical Science	16
Grades 9–12 Science (N = 463)	
Non-college Prep Biology	9
1st Year Biology	23
Advanced Biology	7
Non-college Prep Chemistry	2
1st Year Chemistry	22
Advanced Chemistry	3
Non-college Prep Physics	2
1st Year Physics	8
Advanced Physics	4
Non-college Prep Coordinated or Integrated Science including General and Physical Science	5
1st Year Coordinated or Integrated Science including General and Physical Science	2
Non-college Prep Earth Science	4
1st Year Earth Science	3
Advanced Earth Science	1
Non-college Prep Environmental Science/Ecology	2
1st Year Environmental Science/Ecology	2
Advanced Environmental Science/Ecology	2

MEASUREMENT PROPERTIES OF THE QUESTIONNAIRE

Because fewer teachers than hoped for participated in the study, the original analytic plan on conducting separate EFA and CFA for each grade band was not feasible. Instead, it was decided to conduct an EFA on a random half of the secondary teachers, and use the other half along with the elementary teachers for the CFA.

EFA

The EFA was run using SPSS v22 on a random half of the middle and high school teachers (N = 445). The two types of items (NGSS-aligned objectives and instructional practices) were analyzed separately. For both types of items, a principal components extraction method was first used to investigate the number of factors present in the data. Then, a principal axis factoring extraction method with a direct oblimin rotation, which allows factors to correlate, was used with fixed numbers of factors.

NGSS-Aligned Instructional Objectives

For the 10 NGSS-aligned objectives items, one-, two-, and three-factor solutions were explored, with a one-factor solution being the most plausible. Two items that focused on students understanding the nature of science had substantially lower factor loadings. Because of the lower loadings, and the fact that these items were not explicitly tied to one of the practices, these two items were dropped from further analysis.

NGSS-Aligned Instructional Practices

For the 34 instructional practices items, one- through nine-factor solutions were explored. An analysis of the solutions suggested that an eight-factor solution, with one factor per practice, may be present in the data. However, factor correlations were high, suggesting that these survey items may all be measuring a single underlying construct.

CFA

Using the solutions suggested by the EFA, CFA was performed on each set of items in MPLUS v7.3 using the robust weighted least squares (WLSMV) estimator. Model fit statistics were used to examine the adequacy of the model. Typically, a number of fit indices are examined to judge model fit, using a somewhat holistic approach (Schumacker & Lomax, 1996). For this analysis, the fit indices available in MPLUS were used: the Chi-Square Goodness of Fit test, the CFI, the TLI, and the RMSEA.[1] A significant Chi-Square test indicates that the model is not an adequate fit of the data; however, this test is very sensitive to sample size, and with the relatively large samples used in our study, is not a good measure of fit (Tabachnick & Fidell, 2007). The research community has debated the best criteria for judging fit on each of the remaining indices. We elected to use the traditional criteria, where a good fit is defined as: CFI > 0.9, TLI > 0.9, and RMSEA < 0.08 (Browne & Cudeck, 1993). In addition, model modification indices were examined as indicators of possible model misspecification. In general, the only modifications made to the models were allowing pairs of items within the same factor or that asked about similar aspects of the practices (e.g., items for the practices of Analyzing and Interpreting Data and Using Mathematics and Computational Thinking had similarities) to correlate.

This analysis used the remaining half of the middle and high school teachers and all of elementary teachers. To test whether the factor structure was the same across grade levels, a multiple-group CFA procedure was followed. This procedure involved conducting an initial CFA for each grade band separately, followed by a multiple-group CFA.

[1] These fit indices are typically referred to in abbreviated form. The formal names of the fit indices are: CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation. For more information about each fit index, see Tabachnick & Fidell, (2007).

NGSS-Aligned Instructional Objectives

The CFA for the NGSS-aligned objectives items was based on a sample of 314 elementary school teachers, 222 middle school teachers, and 234 high school teachers who had complete data for the eight items retained for this analysis. The separate grade-band solutions each indicated a good model fit; subsequently, the multi-group model was run. Table 3 shows the fit indices for the multi-group, one-factor solution. Although the chi-square test is significant, the CFI and TLI are both above 0.97 and the RMSEA is below 0.06, indicating good model fit.

Table 3
CFA Model Fit Indices: NGSS-Aligned Instructional Objectives

	Criterion	Value
Chi-Square Goodness of Fit test	Not statistically significant	$\chi^2(87, N = 770) = 159.371$, $p < .01$
CFI	> 0.9	0.976
TLI	> 0.9	0.977
RMSEA	< 0.08	0.057

Table 4 shows the factor loadings, by grade band, from the multi-group analysis. All loadings are greater than 0.5, another indicator of the appropriateness of a one-factor solution.

Table 4
CFA Factor Loadings, by Grade Band: NGSS-Aligned Instructional Objectives

	Factor Loadings		
	Elementary (N = 314)	Middle (N = 222)	High (N = 234)
Gaining proficiency in developing scientific questions (meaning questions that require an answer supported by evidence gathered through systematic investigation) about real-world phenomena (events that occur in the natural world)	0.559	0.520	0.533
Gaining proficiency in developing scientific models (physical, graphical, or mathematical representations of real-world phenomena) to help develop explanations, identify questions, make predictions, or communicate to others	0.657	0.539	0.554
Gaining proficiency in planning and carrying out science investigations that will provide valid and reliable evidence to describe, support a claim about, or test a model of a real-world phenomenon	0.601	0.522	0.504
Gaining proficiency with multiple grade-appropriate methods of analyzing data (for example: tabulation, statistical analysis, graphic representation) in order to reveal patterns and relationships that facilitate the use of data as evidence	0.672	0.624	0.581
Gaining proficiency in developing scientific explanations—claims about real-world phenomena (for example: what causes it, relationships among variables) supported by evidence (for example: data, science principles), and reasoning about how the evidence supports the claim	0.787	0.771	0.732
Gaining proficiency in argumentation—the social process used by scientists to determine the best explanation for a real-world phenomenon, which involves comparing and evaluating multiple ideas or pieces of evidence, noting their strengths and weaknesses, and communicating findings in a clear and logical manner	0.750	0.676	0.639
Gaining proficiency with writing text that provides scientific information in a clear and logical manner	0.565	0.546	0.526
Gaining proficiency with critically evaluating information from multiple sources (regardless of whether it is their own or others' work)—assessing credibility, identifying sources of error or methodological flaws, distinguishing observations from inferences	0.630	0.561	0.524

Cronbach's alpha for these items are shown in Table 5, both overall and by grade band. The alpha for each group is above 0.70.

Table 5
Cronbach's Coefficient Alpha, by Grade Band: NGSS-Aligned Instructional Objectives

	N	Cronbach's Alpha
Overall	770	0.820
Elementary	314	0.839
Middle	222	0.790
High	234	0.772

NGSS-Aligned Instructional Practices

The CFA for the 34 instructional practices items involved 755 teachers: 310 at the elementary, 218 at the middle, and 227 at the high school level. Based on the EFA, an eight-factor model was examined with factors allowed to correlate. As with the objectives items, the model was first run separately for each grade band and then as a multi-group CFA. As can be seen in Table 6, the fit indices provide evidence of the appropriateness of the eight-factor solution.

Table 6
CFA Model Fit Indices: NGSS-Aligned Instructional Practices

	Criterion	Value
Chi-Square Goodness of Fit test	Not statistically significant	$\chi^2(1685, N = 755) = 2719.201, p < .01$
CFI	> 0.9	0.977
TLI	> 0.9	0.977
RMSEA	< 0.08	0.049

Factor loadings for each practice, by grade band, are shown in Tables 7–14. All factor loadings are greater than 0.5, with the vast majority greater than 0.7.

Table 7
Factor Loadings, by Grade Band: Asking Questions

	Factor Loadings		
	Elementary (N = 310)	Middle (N = 218)	High (N = 227)
Determine whether or not a question is "scientific" (meaning it requires an answer supported by evidence gathered through systematic investigation)	0.773	0.796	0.690
Generate scientific questions based on their curiosity, prior knowledge, careful observation of real-world phenomena, scientific models, or preliminary data from an investigation	0.715	0.723	0.710
Revise scientific questions based on their curiosity, prior knowledge, careful observation of real-world phenomena, scientific models, or preliminary data from an investigation	0.785	0.807	0.771

Table 8
Factor Loadings, by Grade Band: Developing and Using Models

	Factor Loadings		
	Elementary (N = 310)	**Middle (N = 218)**	**High (N = 227)**
Develop scientific models—physical, graphical, or mathematical representations of real-world phenomena—based on data and reasoning	0.816	0.769	0.714
As part of developing a model, identify the relevant components of the real-world phenomenon being represented, the relationships among the components, and/or potential causal mechanisms involved	0.833	0.762	0.797
Discuss how scientific models and explanations are based on the best available evidence and may be revised as new evidence is developed	0.869	0.766	0.730
Revise scientific models based on additional evidence	0.870	0.868	0.865
Identify the strengths and limitations of a scientific model—in terms of accuracy, clarity, generalizability, accessibility to others, strength of evidence supporting it—regardless of who created the model	0.869	0.824	0.815
Compare multiple scientific models for a phenomenon in terms of their relative strengths and limitations	0.872	0.844	0.863

Table 9
Factor Loadings, by Grade Band: Planning and Carrying out Investigations

	Factor Loadings		
	Elementary (N = 310)	**Middle (N = 218)**	**High (N = 227)**
Determine what data would need to be collected in order to answer a scientific question	0.833	0.833	0.849
Develop procedures for a scientific investigation to answer a scientific question (regardless of who generated the question)	0.841	0.709	0.820
Revise procedures for a scientific investigation to answer a scientific question (regardless of who developed the original procedures)	0.827	0.743	0.901
Conduct a scientific investigation (regardless of who developed the procedures)	0.682	0.611	0.630

Table 10
Factor Loadings, by Grade Band: Analyzing and Interpreting Data

	Factor Loadings		
	Elementary (N = 310)	**Middle (N = 218)**	**High (N = 227)**
Organize and/or represent data using tables, charts, or graphs in order to facilitate analysis of the data	0.721	0.755	0.685
Compare data from multiple trials or across student groups for consistency in order to identify potential sources of error or inconsistencies in the data	0.838	0.728	0.797
Analyze data using grade-appropriate mathematical or statistical methods in order to identify patterns, trends, or relationships	0.825	0.886	0.707
Consider how missing data or measurement error can affect the interpretation of data	0.858	0.837	0.827

Table 11
Factor Loadings, by Grade Band: Using Mathematics and Computational Thinking

	Factor Loadings		
	Elementary (N = 310)	**Middle (N = 218)**	**High (N = 227)**
Break complex problems down into smaller parts and consider each separately in order to answer a question	0.803	0.736	0.803
Use mathematics to develop and then use scientific models in order to explore a real-world phenomenon (for example: determine a mathematical relationship to represent a phenomenon)	0.900	0.895	0.871
Select and use grade-appropriate mathematical and/or statistical techniques to analyze data (for example: determining the best measure of central tendency, examining variation in data, or developing a fit line)	0.839	0.853	0.862
Use mathematical and/or computational models to generate data to support a scientific claim	0.911	0.897	0.940

Table 12
Factor Loadings, by Grade Band: Constructing Explanations

	Factor Loadings		
	Elementary (N = 310)	**Middle (N = 218)**	**High (N = 227)**
Make and support claims (proposed answers to scientific questions) with evidence	0.814	0.771	0.805
Use multiple sources of evidence (for example: different investigations, scientific literature) to develop an explanation	0.839	0.875	0.848
Revise their explanations (claims supported by evidence and reasoning) for real-world phenomena based on additional evidence	0.893	0.776	0.903

Table 13
Factor Loadings, by Grade Band: Engaging in Argument from Evidence

	Factor Loadings		
	Elementary (N = 310)	**Middle (N = 218)**	**High (N = 227)**
Determine what details about an investigation (for example: its design, implementation, and results) might persuade a targeted audience about a scientific claim (regardless of who made the claim)	0.854	0.774	0.843
Use data and reasoning to defend, verbally or in writing, a claim or refute alternative scientific claims about a real-world phenomenon (regardless of who made the claims)	0.819	0.833	0.792
Evaluate the strengths and weaknesses of competing scientific explanations (claims supported by evidence) for a real-world phenomenon	0.879	0.840	0.866
Construct a persuasive case, verbally or in writing, for the best scientific model or explanation for a real-world phenomenon	0.829	0.839	0.812
Pose questions that elicit relevant details about the important aspects of a scientific argument (for example: the claims/models/explanations, research design, implementation, data analysis)	0.798	0.856	0.816
Respond to questions that elicit relevant details about the important aspects of a scientific argument (for example: the claims/models/explanations, research design, implementation, data analysis)	0.793	0.790	0.725

Table 14
Factor Loadings, by Grade Band: Obtaining, Evaluating, and Communicating Information

	Factor Loadings		
	Elementary (N = 310)	**Middle (N = 218)**	**High (N = 227)**
Read and gather information from scientific resources (for example: books, articles, tables, graphs, models) other than their textbook	0.651	0.703	0.576
Evaluate the credibility of scientific information—for example: its reliability, validity, consistency, logical coherence, lack of bias, or methodological strengths and weaknesses (regardless of whether it is from their own or others' work)	0.880	0.831	0.829
Summarize patterns, similarities, and differences in scientific information obtained from multiple sources (regardless of whether it is from their own or others' work)	0.809	0.816	0.798
Produce scientific text, or make oral presentations, in order to clearly communicate the results of their work (for example: investigations, literature research)	0.771	0.627	0.676

Tables 15–17 show the correlations among the factors for each grade band. Most of the correlations are sizeable, indicating the interrelatedness of the practices.

Table 15
Correlations among Factors: Elementary (N = 310)

	Asking Questions	Developing and Using Models	Planning and Carrying Out Investigations	Analyzing and Interpreting Data	Using Mathematics and Computational Thinking	Constructing Explanations	Engaging in Argument from Evidence	Obtaining, Evaluating, and Communicating Information
Asking Questions	1.00							
Developing and Using Models	0.81	1.00						
Planning and Carrying Out Investigations	0.85	0.81	1.00					
Analyzing and Interpreting Data	0.76	0.86	0.82	1.00				
Using Mathematics and Computational Thinking	0.71	0.80	0.69	0.76	1.00			
Constructing Explanations	0.78	0.86	0.79	0.80	0.73	1.00		
Engaging in Argument from Evidence	0.79	0.89	0.80	0.79	0.84	0.83	1.00	
Obtaining, Evaluating, and Communicating Information	0.68	0.74	0.64	0.64	0.73	0.70	0.87	1.00

Table 16
Correlations among Factors: Middle (N = 218)

	Asking Questions	Developing and Using Models	Planning and Carrying Out Investigations	Analyzing and Interpreting Data	Using Mathematics and Computational Thinking	Constructing Explanations	Engaging in Argument from Evidence	Obtaining, Evaluating, and Communicating Information
Asking Questions	1.00							
Developing and Using Models	0.77	1.00						
Planning and Carrying Out Investigations	0.87	0.71	1.00					
Analyzing and Interpreting Data	0.60	0.65	0.78	1.00				
Using Mathematics and Computational Thinking	0.61	0.74	0.67	0.72	1.00			
Constructing Explanations	0.73	0.72	0.74	0.70	0.62	1.00		
Engaging in Argument from Evidence	0.81	0.79	0.72	0.64	0.73	0.86	1.00	
Obtaining, Evaluating, and Communicating Information	0.76	0.71	0.65	0.54	0.62	0.80	0.87	1.00

Table 17
Correlations among Factors: High (N = 227)

	Asking Questions	Developing and Using Models	Planning and Carrying Out Investigations	Analyzing and Interpreting Data	Using Mathematics and Computational Thinking	Constructing Explanations	Engaging in Argument from Evidence	Obtaining, Evaluating, and Communicating Information
Asking Questions	1.00							
Developing and Using Models	0.73	1.00						
Planning and Carrying Out Investigations	0.77	0.73	1.00					
Analyzing and Interpreting Data	0.53	0.78	0.78	1.00				
Using Mathematics and Computational Thinking	0.43	0.64	0.60	0.75	1.00			
Constructing Explanations	0.87	0.87	0.75	0.74	0.51	1.00		
Engaging in Argument from Evidence	0.81	0.84	0.68	0.61	0.54	0.84	1.00	
Obtaining, Evaluating, and Communicating Information	0.75	0.72	0.64	0.55	0.38	0.86	0.84	1.00

Cronbach's alpha for the eight factors are shown in Table 18, both overall and by grade band. Overall, the alphas are quite high; most are above 0.8 and the lowest being 0.767.

Table 18
Cronbach's Coefficient Alpha, by Grade Band: NGSS-Aligned Instructional Practices

	Overall (N = 755)	Elementary (N = 310)	Middle (N = 218)	High (N = 227)
Asking Questions	0.800	0.808	0.820	0.770
Developing and Using Models	0.913	0.923	0.896	0.895
Planning and Carrying Out Investigations	0.839	0.855	0.810	0.833
Analyzing and Interpreting Data	0.863	0.865	0.859	0.817
Using Mathematics and Computational Thinking	0.877	0.870	0.849	0.877
Constructing Explanations	0.843	0.850	0.810	0.847
Engaging in Argument from Evidence	0.909	0.909	0.907	0.897
Obtaining, Evaluating, and Communicating Information	0.796	0.822	0.777	0.767

A copy of the final set of questionnaire items can be found in the appendix.

USING THE QUESTIONNAIRE

Although HRI holds a copyright on all instruments it develops, HRI grants permission for use at no cost for non-commercial purposes. In any writing in which data from this questionnaire are included, the following citation must be used:

> The questionnaire was developed by the Operationalizing the Science and Engineering Practices project at Horizon Research, Inc., funded by the National Science Foundation under grant number 1445543. Any opinions, findings, and conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation or Horizon Research, Inc.

References

Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage Focus Editions*, *154*, 136–136.

Desimone, L. M. & Le Floch, K. C. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational Evaluation and Policy Analysis*, *26*(1), 1–22. doi:10.3102/01623737026001001

Hamilton, L. S., Nussbam, E. M., & Snow, R. E. (1997). Interview procedures for validating science assessments. *Applied Measurement in Education*, *10*, 181–200.

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

National Research Council. (2013). *Monitoring progress toward successful K-12 STEM education: A nation advancing?* Committee on the Evaluation Framework for Successful K–12 STEM Education. Board on Science Education and Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

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

Schumacker, R. E. & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: L. Erlbaum Associates.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics*. Boston: Pearson/Allyn & Bacon.

APPENDIX

Operationalizing the Science Practices
Teacher Questionnaire

Directions:
We would like to learn more about what science instruction looks like in your classroom. We understand that some teachers teach science to multiple classes. When responding to each item, we would like you to think about **one** science class that you are currently teaching—the class that met first this week. Please fill in only one answer for each statement.

Please be assured that there are no "right" or "wrong" answers. The questionnaire is intended for use with teachers of all grade levels and who teach different science content. Your responses are important whether or not you address these practices in your science instruction**.** Please complete the entire questionnaire even if you do not include these practices in your instruction.

1. Think about your plans for this class for the entire [course/year[1]]. By the end of the [course/year], how much emphasis will each of the following student objectives receive?

	None	Minimal Emphasis	Moderate Emphasis	Heavy Emphasis
a. Gaining proficiency in developing scientific questions (meaning questions that require an answer supported by evidence gathered through systematic investigation) about real-world phenomena (events that occur in the natural world)	○	○	○	○
b. Gaining proficiency in developing scientific models (physical, graphical, or mathematical representations of real-world phenomena) to help develop explanations, identify questions, make predictions, or communicate to others	○	○	○	○
c. Gaining proficiency in planning and carrying out science investigations that will provide valid and reliable evidence to describe, support a claim about, or test a model of a real-world phenomenon	○	○	○	○
d. Gaining proficiency with multiple grade-appropriate methods of analyzing data (for example: tabulation, statistical analysis, graphic representation) in order to reveal patterns and relationships that facilitate the use of data as evidence	○	○	○	○
e. Gaining proficiency in developing scientific explanations—claims about real-world phenomena (for example: what causes it, relationships among variables) supported by evidence (for example: data, science principles), and reasoning about how the evidence supports the claim	○	○	○	○
f. Developing an understanding of the various methods used in science to investigate the natural world	○	○	○	○
g. Gaining proficiency in argumentation—the social process used by scientists to determine the best explanation for a real-world phenomenon, which involves comparing and evaluating multiple ideas or pieces of evidence, noting their strengths and weaknesses, and communicating findings in a clear and logical manner	○	○	○	○
h. Gaining proficiency with writing text that provides scientific information in a clear and logical manner	○	○	○	○
i. Gaining proficiency with critically evaluating information from multiple sources (regardless of whether it is from their own or others' work)—assessing credibility, identifying sources of error or methodological flaws, distinguishing observations from inferences	○	○	○	○

[1] "Course" was shown to teachers of non-self-contained classes; "year" was shown to teachers of self-contained classes.

2. How often do **students** do each of the following during science instruction in this class?

	Never	Rarely (for example: a few times a year)	Sometimes (for example: once or twice a month)	Often (for example: once or twice a week)	All or almost all science lessons
a. Determine whether or not a question is “scientific” (meaning it requires an answer supported by evidence gathered through systematic investigation)	○	○	○	○	○
b. Generate scientific questions based on their curiosity, prior knowledge, careful observation of real-world phenomena, scientific models, or preliminary data from an investigation	○	○	○	○	○
c. Revise scientific questions based on their curiosity, prior knowledge, careful observation of real-world phenomena, scientific models, or preliminary data from an investigation	○	○	○	○	○
d. Determine what data would need to be collected in order to answer a scientific question	○	○	○	○	○
e. Develop procedures for a scientific investigation to answer a scientific question (regardless of who generated the question)	○	○	○	○	○
f. Revise procedures for a scientific investigation to answer a scientific question (regardless of who developed the original procedures)	○	○	○	○	○
g. Conduct a scientific investigation (regardless of who developed the procedures)	○	○	○	○	○
h. Organize and/or represent data using tables, charts, or graphs in order to facilitate analysis of the data	○	○	○	○	○
i. Compare data from multiple trials or across student groups for consistency in order to identify potential sources of error or inconsistencies in the data	○	○	○	○	○
j. Analyze data using grade-appropriate mathematical or statistical methods in order to identify patterns, trends, or relationships	○	○	○	○	○
k. Consider how missing data or measurement error can affect the interpretation of data	○	○	○	○	○
l. Make and support claims (proposed answers to scientific questions) with evidence	○	○	○	○	○
m. Use multiple sources of evidence (for example: different investigations, scientific literature) to develop an explanation	○	○	○	○	○
n. Revise their explanations (claims supported by evidence and reasoning) for real-world phenomena based on additional evidence	○	○	○	○	○
o. Develop scientific models—physical, graphical, or mathematical representations of real-world phenomena—based on data and reasoning	○	○	○	○	○
p. As part of developing a model, identify the relevant components of the real-world phenomenon being represented, the relationships among the components, and/or potential causal mechanisms involved	○	○	○	○	○
q. Discuss how scientific models and explanations are based on the best available evidence and may be revised as new evidence is developed	○	○	○	○	○
r. Revise scientific models based on additional evidence	○	○	○	○	○
s. Identify the strengths and limitations of a scientific model—in terms of accuracy, clarity, generalizability, accessibility to others, strength of evidence supporting it—regardless of who created the model	○	○	○	○	○
t. Compare multiple scientific models for a phenomenon in terms of their relative strengths and limitations	○	○	○	○	○

2. (continued) How often do students do each of the following during science instruction in this class?

	Never	Rarely (for example: a few times a year)	Sometimes (for example: once or twice a month)	Often (for example: once or twice a week)	All or almost all science lessons
u. Break complex problems down into smaller parts and consider each separately in order to answer a question	○	○	○	○	○
v. Use mathematics to develop and then use scientific models in order to explore a real-world phenomenon (for example: determine a mathematical relationship to represent a phenomenon)	○	○	○	○	○
w. Select and use grade-appropriate mathematical and/or statistical techniques to analyze data (for example: determining the best measure of central tendency, examining variation in data, or developing a fit line)	○	○	○	○	○
x. Use mathematical and/or computational models to generate data to support a scientific claim	○	○	○	○	○
y. Determine what details about an investigation (for example: its design, implementation, and results) might persuade a targeted audience about a scientific claim (regardless of who made the claim)	○	○	○	○	○
z. Use data and reasoning to defend, verbally or in writing, a claim or refute alternative scientific claims about a real-world phenomenon (regardless of who made the claims)	○	○	○	○	○
aa. Evaluate the strengths and weaknesses of competing scientific explanations (claims supported by evidence) for a real-world phenomenon	○	○	○	○	○
bb. Construct a persuasive case, verbally or in writing, for the best scientific model or explanation for a real-world phenomenon	○	○	○	○	○
cc. Pose questions that elicit relevant details about the important aspects of a scientific argument (for example: the claims/models/explanations, research design, implementation, data analysis)	○	○	○	○	○
dd. Respond to questions that elicit relevant details about the important aspects of a scientific argument (for example: the claims/models/explanations, research design, implementation, data analysis)	○	○	○	○	○
ee. Read and gather information from scientific resources (for example: books, articles, tables, graphs, models) other than their textbook	○	○	○	○	○
ff. Evaluate the credibility of scientific information—for example: its reliability, validity, consistency, logical coherence, lack of bias, or methodological strengths and weaknesses (regardless of whether it is from their own or others' work)	○	○	○	○	○
gg. Summarize patterns, similarities, and differences in scientific information obtained from multiple sources (regardless of whether it is from their own or others' work)	○	○	○	○	○
hh. Produce scientific text, or make oral presentations, in order to clearly communicate the results of their work (for example: investigations, literature research)	○	○	○	○	○

