Research in Science Education
https://doi.org/10.1007/s11165-018-9797-5

Opportunities to Participate (OtP) in Science: Examining Differences Longitudinally and Across Socioeconomically Diverse Schools

Christine L. Bae[1] · **Morgan DeBusk-Lane**[1] · **Kathryn N. Hayes**[2] · **Fa Zhang**[1]

Abstract

The purpose of this study was to develop and validate a survey of opportunities to participate (OtP) in science that will allow educators and researchers to closely approximate the types of learning opportunities students have in science classrooms. Additionally, we examined whether and how opportunity gaps in science learning may exist across schools with different socioeconomic levels. The OtP in science survey consists of four dimensions that include acquiring foundational knowledge, planning an investigation, conducting an investigation, and using evidence to communicate findings. A total of 1214 middle school students across 8 diverse school districts completed the survey. Tests of reliability, construct validity, measurement invariance, and external validity were conducted using data collected at the beginning and end of the school year. Results showed evidence that the OtP in science survey scores were internally reliable, invariant across school socioeconomic groups across and time points (i.e., lacking systematic biases in responses by group or time point), and externally valid. Given that scores from the survey were reliable and valid indicators of the four dimensions of interest, structural invariance tests were conducted to examine possible differences in OtP in science across schools from high, middle, and low socioeconomic backgrounds. Findings demonstrate specific ways opportunity gaps to learn science manifest in lower income schools. We discuss the implications of these gaps for science instruction, professional development, policy, and diverse students' interest and achievement in science, and propose several lines of future study.

Keywords Opportunities in science · Measurement invariance · Structural invariance · Survey · Middle school · Socioeconomic status

Introduction

Large science achievement gaps are present in United States (U.S.) schools, and the declining numbers of students who choose to study science and pursue a career in science, or science,

---

✉ Christine L. Bae
clbae@vcu.edu

Extended author information available on the last page of the article

technology, engineering, and mathematics (STEM) more broadly, is a national concern among educators, policy-makers, and researchers (Morgan et al. 2016; Quinn and Cooc 2015). Children with low science literacy are more likely as adults to misunderstand public policy related to topics such as climate change and genetic engineering, and generally, be at higher risk for unemployment (U.S. Department of Education 2000). Unfortunately, research shows that many students lose interest in science during the middle school grades (ages approximately between 11 to 13 years old) (Grolnick et al. 2007; Kahn and Kellert 2002; Tobin et al. 1999), and science achievement gaps in the U.S. tend to widen during this pivotal stage, trends that are also found internationally (Bybee and Kennedy 2005; Morgan et al. 2016). These achievement gaps have largely been explained by opportunity gaps: stark inequalities in access to quality science learning opportunities between students from lower versus higher socioeconomic backgrounds (Hayes and Trexler 2016; Morgan et al. 2016; Oakes 1990). Students from low socioeconomic backgrounds, who also often represent members of racial and ethnic minority groups, are at greater risk of attending under-resourced schools, where science more frequently consists of rote learning of facts rather than hands-on engagement in meaningful sense-making activities (Hayes and Trexler 2016; Lee and Buxton 2010). However, the research is clear that ongoing, authentic opportunities to participate in science learning is critical for sparking students' interest, achievement, and long-term persistence in STEM (Aschbacher et al. 2010; Minner et al. 2010; Rinke et al. 2013). The increasing diversity of students in U.S. schools, coupled with persistent achievement gaps along racial, ethnic, and socioeconomic lines makes the lack of access to quality science education in low-income schools a problem to be taken seriously.

To date, researchers have examined global indices of opportunities to learn science, such as the licensure of students' science teachers, the degree to which various science topics were taught during the academic year, and number of science courses offered and school conditions and resources (Byrnes and Miller 2007; Guiton and Oakes 1995; Lewis and Farkas 2017; Mo et al. 2013; Oakes 1990). Although these studies underscore the important role of opportunities for science learning in facilitating a broad range of desirable science outcomes, they do not represent the full range of opportunities to participate in science learning that exist in classrooms.

Existing Measures of Science Learning Opportunities

Existing measures of science learning opportunities largely gather information from teachers' perspectives (Hayes et al. 2016; OECD 2016). For example, on the international PISA and TIMSS science teacher questionnaires, items ask teachers to report on the frequency of science activities related to investigations (e.g., "Students write up laboratory reports," "Students make calculations using scientific formulas"; OECD 2016, and "Students observe natural phenomena," "Design or plan experiment"; TIMSS 2015) as well as more traditional learning activities (e.g., "Memorize facts and principles," TIMSS 2015; Gao 2014). However, such items do not exist in the student version of the questionnaires in TIMSS. Additionally, there is only a small set of questions that examine science classroom activities in the PISA student questionnaire, which focus specifically on inquiry-type activities (e.g., "Students are allowed to design their own experiments," "There is a class debate about investigations," OECD 2016). By in large, student questionnaires focus more on students' confidence in various science learning tasks (e.g., "I usually do well in science," "Science is harder for me than any other subject"), values (e.g., “I would like a job involves science,” “I need science to learn other school subjects”), intrinsic interest in science learning (e.g., “I generally have fun learning science topics,” “I like to conduct science experiments”), and out-of-school science learning opportunities (e.g., extra science lessons not provided by the school; Hayes and Trexler 2016; OECD 2016; TIMSS 2015). Thus, there is a need for a more comprehensive measure that assesses the full spectrum of science classroom activities that students report experiencing first-hand in in their classrooms.

In this paper, we aim to contribute to the existing research by developing and validating an opportunities to participate (OtP) in science learning questionnaire for students. The questionnaire was created to measure a comprehensive range of science activities commonly observed in science classrooms. That is, the OtP in science questionnaire is a proximal measure designed to identify specific learning opportunities in science classrooms (beyond the more global indicators typically used, such as the number of science courses offered or teachers’ licensure). Use of more proximal measures, or measures sensitive to context-related features, is important to identify potentially small but significant differences that may not be detectable using more global indicators (Ruiz-Primo et al. 2002). Items were adapted and expanded from student measures of inquiry practices (Llewellyn 2005) as well as a teacher measure of science learning activities (Hayes et al. 2016). We first applied confirmatory factor analysis (CFA) to validate the factor structure of the measure, and tested the multi-group and longitudinal measurement invariance (MI) of the factor scores. Establishing MI is an important prerequisite for using scores from questionnaires to compare latent constructs as well as to examine the scores over time (Gregorich 2006). Specifically, we assessed multi-group and longitudinal invariance of the scores from the OtP in science questionnaire to test that the measurement structure of the questionnaire was equivalent across groups of students from high, middle, and low socioeconomic status (SES) schools (multi-group invariance) and over the academic year (longitudinal invariance; Meredith 1993; Vandenberg and Lance 2000). The multi-group and longitudinal MI tests ensure that any observed differences across groups and over time are due to meaningful differences in students’ OtP in science, as opposed to systematic biases in response patterns. Finally, multi-group structural invariance (SI) tests were conducted to examine possible differences in the latent factor means of the OtP in science learning dimensions across school SES groups. The four dimensions of the OtP in science questionnaire are reviewed next.

Dimensions of Opportunities to Participate in Science Learning

With the advent of the Next Generation Science Standards (NGSS) and similar standards movements across the United States (NGSS Lead States 2013; NRC 2012), there is renewed attention to ensuring that all students are provided with high-quality science education. Such standards movements outline ambitious goals that explicitly call for educators, researchers, and policymakers to make accessible to all students meaningful opportunities to participate in science learning that integrate disciplinary ideas with engagement in authentic scientific practices and cross-cutting concepts (NRC 2012).

Based on the well-established literature regarding science instruction, we developed the OtP in science questionnaire to capture a variety of classroom opportunities for middle school (grades 6 to 8) students to participate in science learning. We reviewed the literature to identify science learning activities that range from traditional direct instruction, which supports acquisition of basic knowledge, to more complex activities, which require the use of knowledge and skills to understand scientific phenomena (e.g., Chinn and Malhotra 2002; Hayes et al. 2016). Based on the review, the following four OtP in science dimensions were identified: (1) the acquiring foundational knowledge, (2) planning an investigation, (3) conducting an investigation, and (4) using evidence and communicating scientific ideas.

The first dimension of science practices, acquiring foundational knowledge, includes traditional practices, such as reading science text and learning science vocabulary. While traditional activities alone are not highly aligned with evidence-based approaches and the new standards for science education, elements of direct instruction can support students' engagement in more complex activities (Chi et al. 1994). A large body of literature shows that activities aimed to develop students' subject matter familiarity and fluency with core skills (e.g., reading science text) support their learning when interwoven with more complex science activities such as developing arguments supported by evidence (Anderson 1993; Baroody 2003; Rittle-Johnson et al. 2001). Thus, providing learning opportunities aimed to support students' acquisition of fundamental science content knowledge and skills (integrated with the practices of science) is important for supporting their mastery of more global competencies in science (Chinn and Malhotra 2002; Greiff et al. 2013; McGinn and Roth 1999). That said, such traditional approaches need to be carefully balanced with opportunities for students to engage in the practices of science in order to provide high-quality science learning experiences.

The second dimension of OtP in science, planning an investigation, relates to activities in which students are thinking critically about their questions, study design, and experimental procedures before conducting their scientific investigation. This dimension necessitates high involvement on the part of students, and is fairly uncommon in science classrooms, which often rely on labs that pre-designate variables, questions, and hypotheses (Hayes et al. 2016; Manz and Suárez 2018; Tekkumru-Kisa et al. 2018). In contrast to cookbook labs, the planning an investigation dimension involves active student sense-making with uncertain or open-ended investigations. That is, students are applying theory and evidence to generate novel research questions, identify key variables for investigation and consider bias, and determining appropriate procedures and measures to employ (Chinn and Malhotra 2002; Kolodner et al. 2003; Podschuweit and Bernholt 2017).

The third dimension of OtP in science, conducting an investigation, involves activities focused on inquiry practices of recording observations (Hmelo-Silver 2004; Krajcik et al. 1994), building, testing, and revising models of scientific phenomena (Passmore and Stewart 2002; Windschitl et al. 2008), and collecting and analyzing data for use as evidence in scientific explanations and arguments (e.g., McNeill and Krajcik 2008; Osborne et al. 2016). These activities are common in science classrooms that take an inquiry approach, where students collect data and engage in analysis and computational thinking. However, the level of deep student involvement in these activities across classrooms can vary greatly, from worksheet-based labs to predictive modeling of scientific phenomena (Driver et al. 2000; NRC 2012; Schwarz et al. 2009; Tekkumru-Kisa et al. 2018). The opportunities measured in the conducting an investigation dimension characterize this range of activities.

Finally, the fourth dimension of OtP in science, using evidence and communicating findings, involves activities associated with scientific discourse (Erduran et al. 2004; Forbes et al. 2013; Kuhn 2015; Lemke 2001). These can be broadly construed as science literacy practices, involving the ability to communicate scientific ideas through various written forms (e.g., texts, tables, diagrams) and through science talk in multiple social settings (e.g., dyads, class discussions, group presentations; Chinn and Malhotra 2002; Cromley et al. 2016;

Windschitl et al. 2012). Examples of opportunities to participate in this dimension of science include writing up findings, debating scientific ideas, and presenting conclusions (McNeill and Krajcik 2008; McNeill; Osborne et al. 2016). Being able to communicate, consider, and negotiate ideas within a community of learners is central to enculturating students into the practices of science; the ways of knowing and producing knowledge in science (Driver et al. 2000; Hogan et al. 1999; Lemke 2001).

Taken together, the four dimensions of the OtP in science questionnaire provides a more comprehensive set of science classroom activities, compared to existing student measures that focus primarily on inquiry-based science activities (e.g., Llwellyn 2005). Scores from this questionnaire can therefore be used to inform more targeted approaches to address the longstanding inequities that have existed in science education, as well to determine how different types of classroom opportunities relate to science learning outcomes. To this end, we also examine potential differences in the various dimensions of OtP in science by school SES groups to inform specific avenues that programs and initiatives can target to create more equitable access to quality science education for our most vulnerable students. The relationship between school socioeconomic background and science learning opportunities are reviewed next.

School Socioeconomic Status and Gaps in Opportunities to Participate in Science

Socioeconomic status is a commonly used contextual variable that represents parental income, parental education, and parental occupation (Duncan et al. 1972; Sirin 2005). Although there is a well-established relationship between of ongoing participation in authentic science learning activities and long-term persistence and achievement in science, severe gaps in science education quality exist across U.S. schools with higher versus lower SES backgrounds (Morgan et al. 2016; Quinn and Cooc 2015). Further, despite the longstanding call to engage in equitable practices, race, gender, and economic status continue to be significant markers of the degree to which students have access to quality educational experiences, and in turn, achievement in science (Lee and Luykx 2005; Oakes 1990; President's Council of Advisors on Science and Technology [PCAST] 2010; Quinn and Cooc 2015). Specifically, it is well-documented that students from lower SES backgrounds often enter middle school with little to no formal science education, as time and resources are typically prioritized for math and language arts in the primary or elementary grades (ages between approximately 6 to 10 years old, Anderson and Shattuck 2012; Hayes and Trexler 2016). Once in middle school, students from lower SES schools are typically taught by middle school teachers who are less prepared to teach science (e.g., earned their Bachelor's degree in a subject outside of science) and are also not as likely to implement hands-on experimentation, group activities, and open-ended explorations of scientific phenomena (Hanushek and Rivkin 2006; Hartry et al. 2012; Oakes 1990). In addition to differences in quality of teachers and science instructional approaches, several gaps in opportunities for students to participate in authentic science learning have been documented at the classroom and school level via inequitable practices such as curricular tracking, lack of access to STEM programs, and insufficient resources related to lab facilities, equipment, and curricula that support participation in authentic science learning (e.g., Baker et al. 2002; Banilower et al. 2013; Hanushek and Rivkin 2006; Jacob 2007; Landkford et al. 2002; Mo et al. 2013;

Oakes 1990). The cumulative effect of this opportunity gap during students' formative primary and secondary years likely contributes to the significant drops in science interest and achievement documented among students underrepresented in STEM (McCoach et al. 2006; Saçkes et al. 2011).

Because of these systemic differences, the demographic makeup of a school has a demonstrated relationship with both student science learning opportunities and student achievement, above and beyond individual characteristics (Mo et al. 2013; Oakes 1990; Quinn and Cooc 2015). For example, findings from studies examining the effects of school SES on students' learning indicate that even after controlling for student-level characteristics, the effect of school SES variables on students' science achievement remain significant (e.g., Baker et al. 2002; Caldas and Bankston 1997; Sirin 2005). It has thus been argued that when examining the distribution of opportunities, the best way to assess individuals' opportunities is by examining the schools in which particular groups of students are clustered.

Therefore, in this study, we examined whether OtP in science differ as a function of school SES. The OtP in science questionnaire and findings regarding differences across school SES groups presented in this paper will support efforts to more fully understand the degree to which students are given opportunities to participate in activities identified in the literature to support science learning, as reported from students' perspectives.

Present Study

The aim of this study was to develop a questionnaire to capture a comprehensive range of opportunities to participate in science learning in middle school classrooms. The focus on developing a measure that assesses science classroom activities that students are directly engaged in adds to the existing literature that commonly rely on external, more global indices of science learning opportunities (e.g., teacher's years of experience, school conditions). Additionally, we extend upon the small number of existing student questionnaires, which focus more narrowly on inquiry activities (Llewellyn 2005) to assess a wider range of opportunities to engage in science learning. We first conducted a confirmatory factor analysis (CFA) to confirm the fit of the four-factor measurement model, followed by multi-group (tests of invariance based on SES group of school) and longitudinal (tests of invariance between time points) measurement invariance (MI) tests to examine if the questionnaire performed consistently across school SES groups and time points. In addition, the relationships among the four OtP in science dimensions and students' science engagement and self-efficacy were examined as tests of external validity. Establishing these psychometric properties of the measure is important to understand to what degree students have opportunities to participate in a range of science learning activities, including those advocated in many recent standards movements (e.g., The K12 Framework for Science Education, NRC 2012). Finally, multi-group structural invariance (SI) tests were conducted to examine possible differences in OtP in science among groups of students from schools with different socioeconomic status (SES).

The research questions that guided this study were as follows:

1. Is there evidence for the four-factor measurement model underlying the OtP in science questionnaire? (CFA)
2. Is the factor structure of the OtP in science questionnaire invariant across groups of students from schools with different SES? (multi-group MI)

3. Is the factor structure of the OtP in science questionnaire invariant across time? (longitudinal MI)
4. Do students' OtP in science differ as a function of school SES? (multi-group SI)

Regarding the psychometric properties of the OtP in science questionnaire, it was expected that the four-factor structure of the OtP in science questionnaire would fit the data well (CFA), and that students' responses would be equivalent across groups and time (multi-group and longitudinal MI), demonstrating evidence for the validity of the questionnaire ratings. Group differences in students' OtP in science were expected to exist by school SES groups (multi-group SI), based on past studies that show that even after controlling for individual student SES characteristics (e.g., family poverty, social status), school-level SES exerted a significant influence on the quality of learning experiences and students' academic achievement (e.g., Baker et al. 2002; Caldas and Bankston 1997). Specifically, we hypothesized that students from high SES schools would report significantly higher OtP in science compared to students from middle and low SES schools.

Methods

Participants and Procedures

A total of 1214 students across 8 diverse school districts from the western region of the United States participated in this study. The student sample included grades 6 ($n = 374$), 7 ($n = 439$), and 8 ($n = 400$), male (46.3%) and female (53.7%). Approximately 50.35% of the students qualified for Free and Reduced Lunch (FRL),[1] and 17.94% were identified as English Language Learners (students who speak English as a second language and/or are non-native English speakers). As done in past studies (e.g., Harwell and Lebeau 2010; McKenna et al. 2012), the distribution of school %FRL was used to create tertile SES groups for tests of measurement and structural invariance (see Table 1 for demographics by high, middle, and low SES group). The student questionnaires and demographics questionnaire were administered during the regularly scheduled class time by the teacher during the beginning (time 1[T1]) and end (time 2[T2]) of the 2014–15 academic year.

Measures

Opportunities to Participate in Science (Otp) Questionnaire The opportunities to participate (OtP) in science questionnaire consists of 23 items, which are rated on a 5-point Likert scale ranging from 1 (*Not at all*), 2 (*A little*), 3 (*Some*), 4 (*A lot*), and 5 (*All of the time*) to indicate the extent to which students had opportunities to participate in these activities in their science classrooms. The items were organized into four dimensions that include (a) acquiring foundational knowledge (three items, e.g., "Read from a science book or other handouts in class"), (b) planning an investigation (four items, e.g., "Choose variables to in investigate"), (c) conducting an investigation (nine items, e.g., "Write or draw observations"), and (d) using evidence and communicating scientific ideas (seven items, e.g., "Present data and conclusions to the class").

---

[1] The FRL is a federally assisted meal program in public and nonprofit private schools that provide low-cost or free lunches every school day to children from families who meet the income eligibility guidelines.

Table 1 Group sample sizes and demographics by school SES tertile group

School SES	Student *n*	School *n*	School %FRL								
			M	*SD*	Min	Max	White	Black	Hispanic	Asian	Other
High	420	9	27.24	11.38	0.00	40.50	21.00	33.10	34.04	7.80	1.42
Middle	525	6	60.29	7.05	50.50	67.50	21.10	23.19	47.72	6.08	1.90
Low	255	7	75.62	3.36	69.30	79.90	12.50	14.02	64.02	7.20	2.27

Science Engagement and Self-Efficacy in Science Self-report engagement and self-efficacy in science questionnaires were administered via paper-and-pencil (Lee et al. 2016). The items were measured on a 5-point Likert scale, ranging from 1 (*Not true at all*) to 5 (*Very true*). The three science engagement subscales included behavioral (three items, T1: $\alpha = .74$, T2: $\alpha = .73$, e.g., "I follow the rules in my science class"), affective (three items, T1: $\alpha = .76$, T2: $\alpha = .78$, e.g., "I feel excited about the learning activities in my science class", and cognitive (three items, T1: $\alpha = .73$, T2: $\alpha = .74$, "During science class, I ask questions and offer suggestion") engagement. Science self-efficacy was assessed using three items (T1: $\alpha = .79$, T2: $\alpha = .81$) that asked students about their self-efficacy in science ("Even if the science classwork is hard, I can do it").

Analyses

Confirmatory Factor Analysis and Reliability

To establish the construct validity of the scale, the fit of the four-factor (representing the four dimension of OtP in science) model was tested at T1 and T2 using confirmatory factor analysis (CFA), and evaluated based on the following recommended cut-off criteria: RMSEA ≤ 0.06, CFI/TLI ≥ 0.90, and SRMR ≤ 0.06 (Hu and Bentler 1998, 1999). Although a probability value of $\alpha = .05$ for the chi-square ($\chi 2$) test statistic is also reported, because χ^2 is sensitive to sample size and model complexity, the GOF indices were used to determine model fit (Kline 2015). Each item was specified to load onto one of the four factors representing an OtP in science dimension. That is, each item was associated with only one of the four OtP dimensions.

In order to determine internal reliability, robust omega coefficients and bootstrapped confidence intervals were calculated for each of the four subscales (Zhang and Yuan 2016). We preferred omega over the traditional Cronbach's alpha since omega makes fewer and more realistic methodological assumptions, so problems associated with inflation and attenuation of internal consistency estimation are far less likely (Dunn et al. 2014). The measurement consistency of the OtP in science questionnaire was examined by calculating the product-moment correlations between the beginning and end-of-year OtP in science dimension scores. A Pearson correlation of .70 or higher is considered evidence for test-retest reliability (Brown et al. 2004; Stemler 2004).

Multi-Group Measurement and Longitudinal Measurement Invariance

Tests of measurement invariance (MI) are used to determine whether the measurement structure is stable across groups or across measurement occasions (Vanderberg and Lance

2000). The argument for testing MI follows that subgroups within a population are often heterogeneous regarding the parameter values of a model, and thus assuming homogeneity of the population is not appropriate (Muthén 1989). This issue is particularly important in educational research with convenience samples, because groups may differ from one another or from the overall population in regard to measurement parameters (Steinmetz et al. 2009). Therefore, tests of MI using the multi-group and longitudinal technique (Byrne et al. 1989; Little 1993) were conducted to examine whether the measurement structure of the OtP in science questionnaire were equivalent across groups of students from high, middle, and low SES schools as well as between the beginning and end of the academic year (Meredith 1993; Vandenberg and Lance 2000). We assessed multi-group SES MI at both time points, and also conducted longitudinal MI for the entire sample.

Specifically, we conducted tests of (1) configural invariance, which tests whether the same factor structure exists between groups; (2) metric (or weak) invariance, which tests whether all groups respond equally to the scale items (i.e., equivalent factor loadings across groups); (3) scalar (or strong) invariance, which tests whether each group has invariant starting points on the scale (i.e., equivalent item intercepts across groups invariance); and (4) error (or strict) invariance, which assesses whether each group has equal amount of error (Meredith 1993; Vandenberg and Lance 2000). The MI tests were conducted in Mplus 8.1 (Muthén and Muthén 2011). Robust maximum likelihood estimation was used to account for missing data (Muthén and Muthén 2011), and all Likert-scale variables were treated as continuous (Rhemtulla et al. 2012). Nested models were hierarchically ordered with parameter equality constraints added at each step of the MI test (Jöreskog 1993; Meredith 1993). The same cut-points for reasonable fit used in the CFA were applied to evaluate the configural model (Vandenberg and Lance 2000). Evidence for MI is found if imposing additional constraints to the model results in little to no change in the goodness of fit statistics. To compare the models, we applied the following recommended cut-off criteria: changes in CFI greater than or equal to .01, supplemented by a change in RMSEA greater than or equal to .015 or a change in SRMR greater than or equal to .03 (metric invariance) and greater than or equal to .01 (for scalar invariance) (Chen 2007; Cheung and Rensvold 1999, 2002; Hu and Bentler 1999).

External Validity

We tested for external validity using the four composite factors scores from the OtP in science questionnaire. Correlational analyses were conducted to examine the relationships between the OtP in science factor scores and students' mean ratings on scales of science engagement and self-efficacy in science. Evidence of external structure validity would be indicated by moderate, positive correlations between the OtP dimensions and the three indicators of science engagement and students' self-efficacy in science (e.g., Fredricks et al. 2016; Lee et al. 2016; Owens et al. 2017; Pajares et al. 2000). That is, we would expect there to be a relatively positive relationship between OtP in science and students' engagement and self-efficacy related to science learning. On the other hand, we would expect small, negative correlations between the four dimensions of OtP in science and school %FRL. These patterns of correlations are indicative of convergent external structure validity (Messick 1989).

Multi-Group and Longitudinal Structural Invariance

Using the most constrained model from the MI tests, we assessed structural invariance (SI) across SES groups and longitudinally within a multi-group CFA framework (Little 2013). These analyses involve testing increasingly constrained models to assess the equivalence of latent structural components of factor variances, covariances, and means (Little 2013). The cut-off criteria for comparing model fit in the MI tests were used here (Chen 2007). Although similar in many ways to MI testing, SI tests are used to assess differences between groups or time in regard to the measured latent constructs and thus, can be invariant. That is, after MI is established, tests of SI are concerned with the substantive difference in latent factors between groups or time points (Little 2013). If latent means proved to be invariant, Wald chi-square difference testing is used to further examine such differences and to establish statistical difference using the MODEL CONSTRAINT function inherent to Mplus (Muthén and Muthén 2017).

Results

The Four-Factor Structure and Reliability of the OtP in Science Questionnaire

The goodness of fit (GOF) indices from the CFA at T1 and T2 showed that the four-factor model adequately fit the data (Table 2). However, based on a review of the modification indices (statistics that suggest post hoc model parameter adjustments to improve model fit), and attention to similarity in item wording (Marsh et al. 2004a, b), possible method effects were accounted for by allowing residuals of items 8 and 9 to correlate. Specifically, residuals of these two items were allowed to correlate because there was a high level of similarity in the wording and/or content of the items (i.e., both items 8 and 9 refer to drawing while conducting scientific investigations) (Kline 2015; Marsh et al. 2004a, b). These modifications resulted in significantly improved model fit at both time points. Thus, this model was retained for future MI tests. Standardized factor loadings ranged from .423 to .825 across the four factors and both time points, meeting the criteria of a minimum .40 factor loading for retaining items (Marsh et al. 2004a, b). The correlations among the four factors ranged between .624 and .888, indicating that the factors represent related but distinct constructs of science practices.

We also found evidence of the reliability of the factor scores; omega coefficients ranged from .67 to .91 for scores at T2 (Table 3). Correlations between factor scores at T1 and T2 ranged from .72 to .83 demonstrating evidence for test-retest reliability.

Multi-Group MI Test by School SES Groups and Longitudinal MI Tests

Multi-Group MI

Results from the multi-group MI tests of the OtP in science practices questionnaire by school SES groups are presented in Table 4. Results showed evidence of marginal fit for the configural (baseline) model, thus this model was used for subsequent MI tests. Metric and scalar invariance held across the four dimensions of the OtP in science questionnaire across groups. Thus MI was established for the equivalence of the OtP in science factor structure, factor loadings, and item intercepts across school SES groups, allowing us to compare and make valid inferences about the differences between latent factor means (Byrne et al. 1989). In other words, the latent means

Table 2 Fit statistics for four-factor OtP in science CFA model at T1 and T2

	χ^2	*df*	*p* value	RMSEA	CFI	TLI	SRMR	Range of stdyx. factor loadings
T1 4-factor model	1875.70	224	< .001	.057	.900	.887	.041	.423–.775
T1 4-factor model + CR	1574.68	223	< .001	.052	.918	.907	.039	.532–.783
T2 4-factor model	1674.764	224	< .001	.060	.905	.893	.043	.563–.825
T2 4-factor model + CR	1464.845	223	< .001	.055	.919	.908	.042	.563–.824

T1 time 1, *T2* time 2, *CR* correlated residuals for parallel items 8 and 9, *stdyx* standardized values

from the four dimensions of the OtP questionnaire can be used to compare groups. Evidence was not found for residual (strict) invariance, indicating that the explained variance for every item is not equivalent across groups. However, even if the error variances are not equivalent, because residual invariance imposes strict constraints that are typically difficult to meet, it has been suggested that in practice, groups can still be compared on the latent variables allowing variables to be measured with different amount of error between groups (Vadenberg and Lance 2000).

Table 3 OtP in science survey items and reliability coefficients at T1 and T2

Factor	OtP in science item	Omega at T1 (95% CI)	Omega at T2 (95% CI)
Acquiring foundational knowledge	1. Fill out science worksheets	0.65 (0.62, 0.68)	0.67 (0.64, 0.70)
	2. Go over science vocabulary		
	3. Read from a science book or handout		
Planning an investigation	4. Write hypothesis or predictions	0.89 (0.88, 0.89)	0.91 (0.90, 0.91)
	5. Choose variables to investigate		
	6. Come up with and complete your own scientific investigations		
	7. Come up with questions to investigate		
Conducting an investigation	8. Write or draw observations	0.81 (0.80, 0.83)	0.85 (0.84, 0.87)
	9. Draw and label diagrams		
	10. Follow the steps of an investigation or experiment		
	11. Collect different types of data		
	12. Make measurements		
	13. Make charts or graphs out of data		
	14. Describe patterns among variables		
	15. Analyze data using math		
	16. Make conclusions based on data		
Using evidence and communicating scientific ideas	17. Write findings in a lab report or notebook	0.82 (9.81, 0.84)	0.83 (0.82, 0.85)
	18. Present data and conclusions to the class		
	19. Have a whole class discussion		
	20. Talk with group members about the investigation		
	21. Support your claim with evidence		
	22. Debate a scientific idea with classmates		
	23. Discussing real-world aspects of science		

Longitudinal MI

Results from the longitudinal MI tests of the OtP in science questionnaire are also presented in Table 4. Results showed evidence of fit for the configural (baseline) model, thus this model was used to subsequent MI tests. The metric, scalar, and residual held across the four dimensions of the science practices questionnaire at both time points (beginning and end of the academic year). In other words, we found strong evidence that students respond in a similar fashion (i.e., without bias over time) to the items on the OtP in science questionnaire.

External Validity

Correlational analyses among the four OtP in science factors and students' science engagement and self-efficacy in science were significantly moderate and positive as expected, providing evidence for convergent external validity of the OtP in science scores (Table 5). At T1 and T2, correlations between the OtP in science dimensions and students' engagement in science ranged between $r = .309$ to .490, and between .330 and .475, respectively. Similarly, the correlations between the OtP in science dimensions and students' self-efficacy in science ranged between $r = .332$ and .402, and between .348 and .401, at T1 and T2 respectively. Also as expected, correlations among the four OtP in science dimensions and school % FRL were significant and negative, albeit small in magnitude (ranging between $r = -.179$ to $-.098$ and $-.163$ to $-.297$, at T1 and T2 respectively).

Multi-Group Structural Invariance

Having established strong MI across school SES groups, structural invariance (SI) (invariance of factor variance, covariance, and latent means) tests were conducted to assess the equality of latent constructs between SES tertile groups (i.e., whether school SES groups differed on the OtP in science dimensions). Using the strong invariance model from the MI tests, results at both time points showed invariant factor variance and covariance (Table 6). Although the

Table 4 Model fit statistics for models representing different degrees of measurement invariance (MI) across school SES groups (multi-group MI) and over the academic year (longitudinal MI)

Model	χ^2	*df*	CFI	ΔCFI	RMSEA	ΔRMSEA	SRMR	ΔSRMR
Multi-group MI tests at T1								
1. Configural	1511.900	669	0.896		0.059		0.055	
2. Metric (weak)	1574.214	707	0.893	−0.003	0.058	−0.001	0.061	0.006
3. Scalar (strong)	1671.926	745	0.885	−0.008	0.058	0.000	0.064	0.003
4. Residual (strict)	1829.203	787	0.871	0.014	0.060	0.002	0.080	0.016
Multi-group MI tests at T2								
1. Configural	1769.070	669	0.890		0.064		0.055	
2. Metric (weak)	1872.206	707	0.883	−0.007	0.064	0.000	0.065	0.010
3. Scalar (strong)	1988.889	745	0.876	−0.007	0.065	0.001	0.069	0.004
4. Residual (strict)	2082.096	787	0.870	0.006	0.064	−0.001	0.080	0.011
Longitudinal MI tests								
1. Configural	3242.866	959	0.888		0.044		0.043	
2. Metric (weak)	3299.599	978	0.886	−0.002	0.044	0.000	0.045	0.002
3. Scalar (strong)	3419.679	996	0.881	−0.005	0.045	0.001	0.046	0.001
4. Residual (strict)	3526.259	1017	0.877	−0.004	0.045	0.000	0.048	0.002

All models include correlated residuals for parallel items 8 and 9 change in CFI was greater than 0.01 between the factor variance and factor covariance models, drop in model fit was not substantiated by drops in RMSEA or SRMR values. Constraining factor means to equality (each set of means set to zero) across school SES groups resulted in a significant reduction in model fit. This indicates that the dimensions of OtP in science, on average, differ across school SES groups. Thus, a closer examination of mean differences by school SES group was warranted.

The descriptive statistics for each of the four dimensions of the OtP in science across the school SES groups are presented in Table 7. As expected, the high SES group report higher OtP in science across all four dimensions compared to the middle SES group, and similarly, the middle SES group reported higher OtP in science compared to the low SES group (Fig. 1). To test whether these differences were significant, two of the three groups were allowed to be freely estimated (Little 2013; Muthén and Muthén 2017). The latent factor means (representing the OtP in science dimension) of the middle and low SES groups were all statistically lower than the high SES group. Significant statistical differences were also found between the middle and low SES group on the OtP in science dimension scores. At T1, the middle and low SES groups were significantly different on all OtP in science dimensions except for the using evidence and communicating scientific ideas dimension. At T2, the middle and low SES groups were significantly different on all OtP in science dimensions except for the acquiring foundational knowledge dimension.

Discussion

In this study, we conducted a series of rigorous validation tests of the OtP in science questionnaire, developed to assess a comprehensive measure of classroom opportunities for students to participate in science learning. Our results provide evidence for the multidimensionality (i.e., four-factor structure) of the OtP in science questionnaire that include the following four dimensions: acquiring foundational knowledge, planning an investigation, conducting an experiment, and using evidence and communicating findings. Further, findings showed that the equivalence of the scores from the OtP in science questionnaire demonstrated strong (metric) invariance across groups of students from high, middle, and low SES schools,

Table 5 Correlations among OtP in science factors, engagement, self-efficacy, and school %FRL

Variable	1	2	3	4	5	6	7	8	9
1. Found	1	.482	.555	.492	.393	.356	.352	.357	−.163
2. Plan	.457	1	.898	.652	.330	.394	.471	.348	−.248
3. Conduct	.530	.889	1	.724	.406	.440	.469	.398	−.297
4. Comm	.461	.641	.714	1	.367	.446	.475	.401	−.220
5. Beh Engage	.403	.309	.405	.356	1	.491	.372	.540	−.233
6. Affect Engage	.357	.355	.414	.377	.505	1	.520	.545	−.136
7. Cog Engage	.363	.468	.490	.465	.384	.489	1	.409	−.137
8. Self-Efficacy	.362	.332	.402	.395	.547	.521	.440	1	−.110
9. %FRL	−.098	−.143	−.179	−.103	−.179	−.096	−.090	−.075	1

Lower diagonal represent T1 and upper diagonal represents T2 correlations; OTP in science variables include Found = acquiring foundational knowledge, Plan = planning an investigation, Conduct = conducting an investigation, Comm = using evidence and communicating scientific ideas. Science engagement variables include: Beh_Engage = behavioral engagement, Affect_Engage = affective engagement, Cog_Engage = cognitive engagement. %FRL = % free reduced lunch of school. All correlations were significant at $p < .01$

Table 6 Model fit statistics for models representing different degrees of structural invariance (SI) across school SES groups

Model	χ^2	*df*	CFI	ΔCFI	RMSEA	ΔRMSEA	SRMR	ΔSRMR
Multi-group SI tests at T1								
3. Scalar (strong)	1671.926	745	0.885		0.058		0.064	
5. Factor variance	1680.738	753	0.885	0.000	0.058	0.000	0.070	0.006
6. Factor covariance	1835.784	770	0.868	−0.017	0.061	0.003	0.075	0.005
7. Factor mean	1929.941	778	0.858	−0.010	0.064	0.003	0.090	0.015
Multi-group SI tests at T2								
3. Scalar (strong)	1988.889	745	0.876		0.065		0.069	
5. Factor variance	2013.618	753	0.874	−0.002	0.065	0.000	0.081	0.012
6. Factor covariance	2152.663	770	0.862	−0.012	0.067	0.002	0.085	0.004
7. Factor mean	2264.220	778	0.851	−0.011	0.069	0.002	0.108	0.023

All models include correlated residuals for parallel items 8 and 9

and strict (scalar) invariance over the school year. In summary, results show that the scores from the OtP in science questionnaire support the theorized four-factor model structure, and

Table 7 Descriptive statistics by school SES group for the four dimensions of OtP in science

Dimension of science practices at T1	M (SD)		
	High SES	Middle SES	Low SES
Acquiring foundational knowledge	3.81 (.78)	3.37 (.79)	3.50 (.83)
Planning an investigation	3.39 (.88)	3.07 (.86)	2.91 (.93
Conducting an investigation	3.53 (.76)	3.26 (.72)	3.08 (.82)
Using evidence and communicating scientific ideas	3.56 (.78)	3.24 (.75)	3.22 (.81)
Dimension of science practices At T2			
Acquiring foundational knowledge	3.83 (.76)	3.40 (.81)	3.45 (.88)
Planning an investigation	3.37 (.97)	3.07 (.84)	2.90 (.97)
Conducting an investigation	3.59 (.80)	3.28 (.72)	3.05 (.86)
Using evidence and communicating scientific ideas	3.69 (.77)	3.34 (.75)	3.43 (.80)
Engagement and self-efficacy at T1			
Behavioral engagement	4.29 (.76)	4.00 (.72)	3.93 (.71)
Affective engagement	4.15 (.76)	3.92 (.81)	3.85 (.85)
Cognitive engagement	3.17 (.87)	2.87 (.87)	2.85 (.83)
Self-efficacy	4.14 (.76)	3.93 (.80)	3.96 (.74)
Engagement and self-efficacy at T2			
Behavioral engagement	4.19 (.69)	3.90 (.70)	3.73 (.72)
Affective engagement	4.03 (.82)	3.73 (.86)	3.81 (86)
Cognitive engagement	3.16 (.91)	2.79 (.86)	2.93 (.89)
Self-efficacy	4.16 (.78)	3.87 (.81)	3.97 (.81)
Latent factor scores at T1	Latent M		
	High SES	Middle SES	Low SES
Acquiring foundational knowledge	0	−0.75	−0.53
Planning an investigation	0	−0.43	−0.64
Conducting an investigation	0	−0.341	−0.60
Using evidence and communicating scientific ideas	0	−0.46	−0.52*
Latent Factor Scores at T2			
Acquiring Foundational Knowledge	0	−0.71	−0.64*
Planning an Investigation	0	−0.37	−0.59
Conducting an Investigation	0	−0.43	−0.80
Using Evidence and Communicating Scientific Ideas	0	−0.48	−0.68

Latent mean scores for middle and low SES are mean differences to zero (with High SES as the referent group). All High SES latent factors were fixed to zero for model identification. * = statistically non-significant difference between middle and low SES groups. All middle and low SES factor mean scores are significantly different than zero

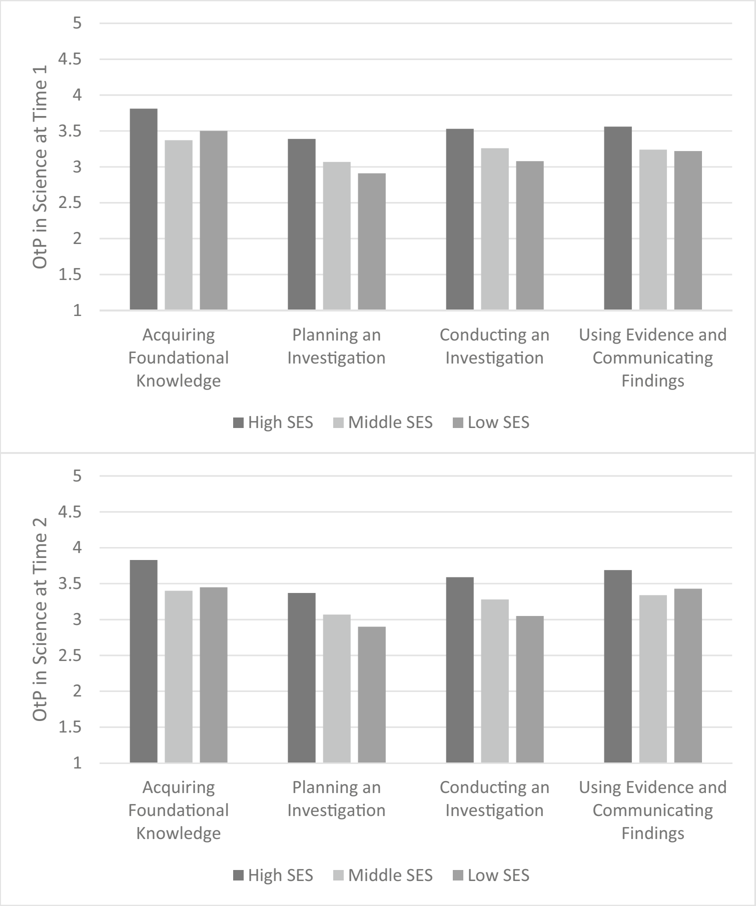

Fig. 1 Differences in the four OtP in science dimensions across school SES groups at time 1 (beginning of the school year) and time 2 (end of the school year)

the four dimensions of OtP in science can be reliability measured across students from schools with diverse socioeconomic backgrounds as well as over time. Establishing multi-group and longitudinal MI is an important prerequisite for making group comparisons using the factor mean scores of the questionnaire (Chen 2007; Cheung and Rensvold 2002; Meredith 1993). Evidence for the external validity of the scores from the OtP in science questionnaire was also found, based on positive, moderate correlations between the four OtP in science dimensions and indicators of students' engagement and self-efficacy in science.

Given evidence for the robust psychometric properties of the OtP in science questionnaire, we turned our attention to questions of substantive interest, that is, whether OtP in science differed in significant ways among students who attended schools of different SES levels. Findings from multi-group SI tests indicated that significant differences existed by school SES group across all four OtP in science dimensions. As expected, we found that students in the low SES school group reported significantly fewer opportunities to participate in science across all four dimensions, compared to their peers in the high and middle SES school groups. These patterns of differences in OtP in science were also observed when comparing the middle and high SES groups in our analyses. Finally, it is worth noting that we found that these differences persisted over the school year, as students in the low SES schools continued to report the lowest opportunities across all four OtP in science dimensions at the end of the school year.

Taken together, these findings indicate that schools in which approximately half or more of students qualify for FRL (identified as middle and low SES schools with %FRL ranging from 40.50 to 79.90 in our study) have fewer opportunities to participate in science, ranging from activities focused on acquiring foundational knowledge (e.g., learning scientific vocabulary) to more complex practices related to scientific investigation (e.g., developing and labeling conceptual models of scientific phenomena) and discourse (e.g., presenting conclusions based on evidence). It is also perhaps not surprising, but a great concern that students in the low and middle SES school groups represent higher percentages of minority (Hispanic and Black) students, who are historically underserved and underrepresented in STEM fields (Morgan et al. 2016). Our findings align with existing literature demonstrating stark and possibly cumulative opportunity gaps along both socioeconomic and racial lines (e.g., Morgan et al. 2016; Quinn et al. 2016; Quinn and Cooc 2015), but provide more detailed information regarding how gaps in students' opportunities to participate in science learning manifest within middle school classrooms.

Notably, our results showed that students in middle and low SES schools reported the lowest ratings on the planning an experiment dimension of science practices. This dimension represents important sense-making activities that support students' participation in the practice of science, particularly in terms of having choice and making decisions about what to investigate (e.g., developing questions, selecting variables) and developing informed predictions about anticipated outcomes. The lack of these opportunities may be an indication that in lower SES schools, students are less likely to encounter inquiry-based approaches to science learning, in which they are expected to navigate open-ended tasks and make decisions when faced with uncertainty in scientific investigations. Rather, based on existing research, it is more likely that students attending schools with lower socioeconomic status are receiving science in "*final form*," as a set of irrefutable facts (Duschl et al. 2007; Lee and Luykz 2007), or are marched through predetermined steps of an investigation towards specific results (Manz and Suárez 2018). The findings related to the acquiring foundational knowledge dimension also support this interpretation. At both the beginning and end of the school year, students in the low SES group reported the highest participation in these traditional science learning activities (e.g., completing worksheets, reading science text) compared to the other three OtP dimensions that are more representative of complex science practices. Although additional research is needed to better inform targeted interventions addressing gaps in science learning opportunities, the trends identified in this study provide a starting point to contribute to this larger effort.

Implications for Practice

Our findings regarding the psychometric properties of the OtP in science questionnaire, as well as the differences in the OtP in science dimensions identified across schools with different SES have several important implications for practice. First, we present evidence that the OtP in science questionnaire can be used by researchers and educators to reliably assess the various opportunities to learn science across middle schools of varying socioeconomic backgrounds, and over time. Such instruments are important to provide science educators and scholars the ability to identify specific gaps in opportunities for students to learn science in diverse settings, and how these opportunities may change over time. A closer approximation of opportunity gaps will support greater accuracy in developing science education policy, professional development programs, and interventions aimed to increase equitable access to science learning for all students. This is especially important for low SES schools attended by students from minority groups who are often underserved and underrepresented in STEM fields. The OtP in science questionnaire can also be used to analyze instruction, curriculum, policy, and/or professional development programs to systematically document the degree to which different types of science learning opportunities are present in students' science education.

It is also worth noting that this study was conducted among diverse middle schools. Middle school is a pivotal stage in students' academic development. In students' secondary years, curriculum becomes more clearly differentiated by subject and students begin to make concrete decisions about their future lines of study and career options (Wang et al. 2016). Unfortunately, it is also during this time when significant drops in students' science motivation, interest, and achievement have been documented (Anderman et al. 1999; Britner and Pajares 2006; Shim et al. 2008, 2008; Tobin et al. 1999). These drops in science interest and achievement are exacerbated for students from ethnic minority groups who are also more likely to attend lower resourced schools (Morgan et al. 2016). It is possible that these trends can be partially explained by the stark differences in the types of science learning opportunities students encounter in middle school (e.g., Quinn et al. 2016; Quinn and Cooc 2015). Our findings provide preliminary evidence for this proposition. Not only do students from lower SES schools report less opportunities to learn science, we also found that these students reported lower engagement and self-efficacy in science compared to their peers who attended higher SES schools. Additionally, results from our external validation tests showed that the dimensions of OtP in science were significantly and moderately correlated to students' engagement and self-efficacy in science, indicating that these constructs are related in a meaningful way. Although examining causal outcomes of OtP in science is beyond the scope of this study, an implication of our findings is that lack of opportunity may have serious implications for how students connect to and persist in science, which ultimately impacts, their science achievement. Our findings clearly indicate opportunity gaps along socioeconomic lines, and point to the critical need to address discrepancies in the quality of science education in middle schools.

Limitations and Areas for Future Research

Some methodological limitations are worth noting. Although we used well-established cut-off criteria (Chen 2007; Cheung and Rensvold 2002) that provide rigorous model comparisons, and also applied these criteria to compare the models in the structural invariance tests, it has been suggested that the log likelihood tests may provide a more stringent test for nested model comparisons (Little 2013). Future research is needed to determine best practices for model comparisons, given that limited criterion for comparing models in tests of structural invariance exist in the literature (Little 2013). We would also like to note that the reliability of the first dimension, acquiring foundation knowledge was slightly below the .70 threshold. Possible reasons for this include the small number of items (three total) in this factor. Future research is needed to examine this factor, such as testing if adding items theoretically related to the acquiring foundational knowledge dimension improves the reliability of this factor.

Another methodological limitation includes the self-report nature of the data. Although there is evidence for the reliability and validity of self-reported information, including self-report data obtained from middle school-aged students (e.g., Morgan and Sonquist 1963; Skinner and Belmont 1993), future research is needed to examine opportunities to participate in science learning using alternative methods such as direct classroom observations.

Additionally, although we longitudinally examined OtP in science scores at two time points during the school year, the data across middle school grades is cross-sectional in nature. That is, students in 6th, 7th, and 8th grade represent unique samples. Future research is needed to longitudinally examine science learning opportunities over the trajectory of students' middle school career to make stronger claims about trends over time. Future research is also needed to validate this questionnaire in other grade levels (elementary, high school) and countries, in order to see if our findings generalize more broadly. In addition, the correlational analysis of the relationships among school SES, OtP in science, and students' engagement and self-efficacy limits our ability to establish causal effects. Future studies using experimental frameworks are needed to make conclusions about causality.

Conclusion

In this study, we provide two major findings to inform efforts to provide high-quality science learning opportunities for all students. First, we present a theoretically grounded and empirically validated OtP in science questionnaire to closely approximate the types of science learning opportunities that students have access to in their classrooms. Second, we present more detailed evidence to support trends reported in past studies regarding the opportunity gaps that exist in schools along socioeconomic lines, and discuss the implications of these gaps for diverse students' interest, persistence, and learning outcomes in science. Finally, we propose several lines of future study to further validate the psychometric properties of the OtP in science questionnaire, to better understand students' science learning opportunities longitudinally, and to examine causal relationships between OtP in science and important student learning outcomes, that together will contribute to efforts to increase equitable access to science learning and diversity in STEM.

Acknowledgments This research is based upon work supported by the National Science Foundation under Grant No. 096280. We would also like to thank the participating teachers and students, as well as the science education team at the Alameda County Office of Education.

References

Anderman, E. M., Anderman, L. H., & Griesinger, T. (1999). The relation of present and possible academic selves during early adolescence to grade point average and achievement goals. *The Elementary School Journal, 100*(1), 3–17.

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist, 48*(1), 35–44.

Anderson, T., & Shattuck, J. (2012). Design-based research: a decade of progress in education research. *Educational Researcher, 41*(1), 16–25.

Aschbacher, P. R., Li, E., & Roth, E. J. (2010). Is science me? High school students' identities, participation and aspirations in science, engineering, and medicine. *Journal of Research in Science Teaching, 47*(5), 564–582.

Baker, D. P., Goesling, B., & LeTendre, G. K. (2002). Socioeconomic status, school quality, and national economic development: a cross-national analysis of the "Heyneman-Loxley effect" on mathematics and science achievement. *Comparative Education Review, 46*(3), 291–312.

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). Report of the 2012 National Survey of science and mathematics education. *Horizon Research, Inc.* (NJ1).

Baroody, A. J. (2003). *The development of adaptive expertise and flexibility: the integration of conceptual and procedural knowledge*. The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise, pp. 1–33.

Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching, 43*(5), 485–499.

Brown, G. T., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing, 9*(2), 105–121.

Bybee, R. W., & Kennedy, D. (2005). Math and science achievement. *Science, 307*(5709), 481.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456–466.

Byrnes, J. P., & Miller, D. C. (2007). The relative importance of predictors of math and science achievement: an opportunity–propensity analysis. *Contemporary Educational Psychology, 32*(4), 599–629.

Caldas, S. J., & Bankston, C. (1997). Effect of school population socioeconomic status on individual academic achievement. *The Journal of Educational Research, 90*(5), 269–277.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: a reconceptualization and proposed new method. *Journal of Management, 25*(1), 1–27.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255.

Chi, M. T., Slotta, J. D., & De Leeuw, N. (1994). From things to processes: a theory of conceptual change for learning science concepts. *Learning and Instruction, 4*(1), 27–43.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: a theoretical framework for evaluating inquiry tasks. *Science Education, 86*(2), 175–218.

Cromley, J. G., Perez, T., & Kaplan, A. (2016). Undergraduate STEM achievement and retention: cognitive, motivational, and institutional factors and solutions. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 4–11.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*(3), 287–312.

Duncan, O. D., Featherman, D. L., & Duncan, B. (1972). *Socio-economic background and achievement*. NY: Seminar Press.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399–412.

Duschl, R. A. (2007). Quality argumentation and epistemic criteria. In *Argumentation in Science Education* (pp. 159–175). Dordrecht: Springer.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88*(6), 915–933.

Forbes, C. T., Biggers, M., & Zangori, L. (2013). Investigating essential characteristics of scientific practices in elementary science learning environments: the practices of science observation protocol (P-SOP). *School Science and Mathematics, 113*(4), 180–190.

Fredricks, J. A., Wang, M. T., Linn, J. S., Hofkens, T. L., Sung, H., Parr, A., & Allerton, J. (2016). Using qualitative methods to develop a survey measure of math and science engagement. *Learning and Instruction, 43*, 5–15.

Gao, S. (2014). Relationship between science teaching practices and students' achievement in Singapore, Chinese Taipei, and the US: An analysis using TIMSS 2011 data. *Frontiers of Education in China, 9*(4), 519–551.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care, 44*(11 Suppl 3), S78–S94.

Greiff, S., Holt, D., & Funke, J. (2013). Perspectives on problem solving in cognitive research and educational assessment: analytical, interactive, and collaborative problem solving. *Journal of Problem Solving (The), 5*, 71–91.

Grolnick, W. S., Price, C. E., Beiswenger, K. L., & Sauck, C. C. (2007). Evaluative pressure in mothers: effects of situation, maternal, and child characteristics on autonomy supportive versus controlling behavior. *Developmental Psychology, 43*(4), 991–1002.

Guiton, G., & Oakes, J. (1995). Opportunity to learn and conceptions of educational equality. *Educational Evaluation and Policy Analysis, 17*(3), 323–336.

Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. *Handbook of the Economics of Education, 2*, 1051–1078.

Hartry, A., Dorph, R., Shields, P., Tiffany-Morales, J., & Romero, V. (2012). *The status of middle school science education in California*. Sacrament: The Center for the Future of Teaching and Learning at WestEd.

Harwell, M., & LeBeau, B. (2010). Student eligibility for a free lunch as an SES measure in education research. *Educational Researcher, 39*(2), 120–131.

Hayes, K. N., Lee, C. S., DiStefano, R., O'Connor, D., & Seitz, J. (2016). Measuring scienceinstructional practices: A survey tool for the age of NGSS. *Journal of Science Teacher Education, 27*(2), 137–164.

Hayes, K. N., & Trexler, C. J. (2016). Testing predictors of instructional practice in elementary science education: The significant role of accountability. *Science Education, 100*(2), 266–289.

Hmelo-Silver, C. E. (2004). Problem-based learning: what and how do students learn? *Educational Psychology Review, 16*(3), 235–266.

Hogan, K., Nastasi, B. K., & Pressley, M. (1999). Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. *Cognition and Instruction, 17*(4), 379–432.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424–453.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118.

Jacob, B. A. (2007). The challenges of staffing urban schools with effective teachers. *The Future of Children, 17*, 129–153.

Jöreskog, K. G. (1993). Testing structural equation models. *Sage Focus Editions, 154*, 294–294.

Kahn, P. H., & Kellert, S. R. (2002). *Children and nature: psychological, sociocultural, and evolutionary investigations*. MIT press.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., Puntambekar, S., & Ryan, M. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: putting learning by design (tm) into practice. *The Journal of the Learning Sciences, 12*(4), 495–547.

Krajcik, J. S., Blumenfeld, P. C., Marx, R. W., & Soloway, E. (1994). A collaborative model for helping middle grade science teachers learn project-based instruction. *The Elementary School Journal, 94*(5), 483–497.

Kuhn, D. (2015). Thinking together and alone. *Educational Researcher, 44*(1), 46–53.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: a descriptive analysis. *Educational Evaluation and Policy Analysis, 24*(1), 37–62.

Lee, C. S., Hayes, K. N., Seitz, J. C., DiStefano, R., & O'Connor, D. (2016). Examiningmotivational structures that differentially predict engagement and achievement in middle school science. *International Journal of Science Education, 38*(2), 192–215.

Lee, O., & Buxton, C. A. (2010). *Diversity and equity in science education: research, policy, and practice*. Multicultural education series. Teachers College Press.

Lee, O., & Luykx, A. (2007). Science education and student diversity: Race/ethnicity, language, culture, and socioeconomic status. *Handbook of Research on Science Education, 1,* 171–197.

Lee, O., & Luykx, A. (2005). Dilemmas in scaling up innovations in elementary science instruction with nonmainstream students. *American Educational Research Journal, 42*(3), 411–438.

Lemke, J. L. (2001). Articulating communities: sociocultural perspectives on science education. *Journal of Research in Science Teaching, 38*(3), 296–316.

Lewis, R. W., & Farkas, G. (2017). Using an opportunity-propensity framework to estimate individual-, classroom-, and school-level predictors of middle school science achievement. *Contemporary Educational Psychology, 51*, 185–197.

Little, R. J. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association, 88*(421), 125–134.

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford press.

Llewellyn, D. (2005). *Teaching high school science through inquiry: A case study approach*. Corwin Press.

Manz, E., & Suárez, E. (2018). Supporting teachers to negotiate uncertainty for science, students, and teaching. *Science Education, 102*(4), 771–795.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004a). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal, 11*(3), 320–341.

Marsh, H. W., Wen, Z., & Hau, K. T. (2004b). Structural equation models of latent interactions: evaluation of alternative estimation strategies and indicator construction. *Psychological Methods, 9*(3), 275–300.

McCoach, D. B., O'connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: a hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology, 98*(1), 14–28.

McGinn, M. K., & Roth, W. M. (1999). Preparing students for competent scientific practice: implications of recent research in science and technology studies. *Educational Researcher, 28*(3), 14–24.

McKenna, M. C., Conradi, K., Lawrence, C., Jang, B. G., & Meyer, J. P. (2012). Reading attitudes of middle school students: results of a US survey. *Reading Research Quarterly, 47*(3), 283–306.

McNeill, K. L., & Krajcik, J. (2008). Scientific explanations: characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 45*(1), 53–78.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543.

Messick, S. (1989). Meaning and values in test validation: the science and ethics of assessment. *Educational Researcher, 18*(2), 5–11.

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching, 47*(4), 474–496.

Mo, Y., Singh, K., & Chang, M. (2013). Opportunity to learn and student engagement: a HLM study on eighth grade science achievement. *Educational Research for Policy and Practice, 12*(1), 3–19.

Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association, 58*(302), 415–434.

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher, 45*(1), 18–35.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*(4), 557–585.

Muthén, L. K., & Muthén, B. O. (2011). *Mplus statistical modeling software (version 6.12)*. Los Angeles: Muthén & Muthén.

Muthén, L.K. and Muthén, B.O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles: Muthén & Muthén.

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

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.

Oakes, J. (1990). *Multiplying inequalities: the effects of race, social class, and tracking on opportunities to learn mathematics and science*. RAND Coporation: Washington D.C.

OECD. (2016). *PISA 2015 Results (Volume) I003A Excellence and Equity in Education*. Paris: PISA, OECD Publishing.

Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S. Y. (2016). The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching, 53*(6), 821–846.

Owens, D. C., Sadler, T. D., Barlow, A. T., & Smith-Walters, C. (2017). Student motivation from and resistance to active learning rooted in essential science practices. *Research in Science Education*, 1–25.

Pajares, F., Britner, S. L., & Valiante, G. (2000). Relation between achievement goals and self beliefs of middle school students in writing and science. *Contemporary Educational Psychology, 25*(4), 406–422.

Passmore, C., & Stewart, J. (2002). A modeling approach to teaching evolutionary biology in high schools. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 39*(3), 185–204.

Podschuweit, S., & Bernholt, S. (2017). Composition-effects of context-based learning opportunities on students' understanding of energy. *Research in Science Education*, 1–36.

President's Council of Advisors on Science and Technology (PCAST). (2010). *Report to the president: prepare and inspire: K-12 education in science, technology, engineering, and mathematics (STEM) for America's future*. Washington, DC: Executive Office of the President.

Quinn, D. M., & Cooc, N. (2015). Science achievement gaps by gender and race/ethnicity in elementary and middle school: Trends and predictors. *Educational Researcher, 44*(6), 336–346.

Quinn, D. M., Cooc, N., McIntyre, J., & Gomez, C. J. (2016). Seasonal dynamics of academic achievement inequality by socioeconomic status and race/ethnicity: updating and extending past research with new national data. *Educational Researcher, 45*(8), 443–453.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373.

Rinke, C. R., Gimbel, S. J., & Haskell, S. (2013). Opportunities for inquiry science in Montessori classrooms: learning from a culture of interest, communication, and explanation. *Research in Science Education, 43*(4), 1517–1533.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: an iterative process. *Journal of Educational Psychology, 93*(2), 346–362.

Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: searching for instructional sensitivity. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 39*(5), 369–393.

Saçkes, M., Trundle, K. C., Bell, R. L., & O'Connell, A. A. (2011). The influence of early science experience in kindergarten on children's immediate and later science achievement: evidence from the early childhood longitudinal study. *Journal of Research in Science Teaching, 48*(2), 217–235.

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Shwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching, 46*(6), 632–654.

Shim, S. S., Ryan, A. M., & Anderson, C. J. (2008). Achievement goals and achievement during early adolescence: examining time-varying predictor and outcome variables in growth-curve analysis. *Journal of Educational Psychology, 100*(3), 655–671.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: a meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85*(4), 571–581.

Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: differences between educational groups in human values measurement. *Quality & Quantity, 43*(4), 599–616.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation, 9*(4), 1–19.

Tekkumru-Kisa, M., Stein, M. K., & Coker, R. (2018). Teachers' learning to facilitate high-level student thinking: impact of a video-based professional development. *Journal of Research in Science Teaching, 55*(4), 479–502.

TIMSS 2015. *Assessment Frameworks. Copyright © 2013 International Association for the Evaluation of Educational Achievement (IEA)*. Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Tobin, K., Seiler, G., & Walls, E. (1999). Reproduction of social class in the teaching and learning of science in urban high schools. *Research in Science Education, 29*(2), 171–187.

U.S. Department of Education. (2000). *National Center for Education Statistics, The Condition of Education 2000, NCES 2000–062*. Washington, DC: U.S. Government Printing Office.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70.

Wang, M. T., Fredricks, J. A., Ye, F., Hofkens, T. L., & Linn, J. S. (2016). The math and science engagement scales: Scale development, validation, and psychometric properties. *Learning and Instruction, 43*, 16–26.

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: model-based inquiry as a new paradigm of preference for school science investigations. *Science Education, 92*(5), 941–967.

Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education, 96*(5), 878–903.

Zhang, Z., & Yuan, K. H. (2016). Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data: Methods and software. *Educational and Psychological Measurement, 76*(3), 387–411.

Affiliations

Christine L. Bae[1] • **Morgan DeBusk-Lane**[1] • **Kathryn N. Hayes**[2] • **Fa Zhang**[1]

Morgan DeBusk-Lane
debusklaneml@vcu.edu

Kathryn N. Hayes
kathryn.hayes@csueastbay.edu

Fa Zhang
zhangf22@mymail.vcu.edu

[1] Department of Foundations of Education, Virginia Commonwealth University, 1015 W. Main Street, P.O. Box 842020, Richmond, VA 23284, USA

[2] Department of Educational Leadership, California State University East Bay, 25800 Carlos Bee Blvd., Hayward, CA 94542, USA