JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 52, NO. 5, PP. 686–709 (2015)

Research Article

Measuring Choice to Participate in Optional Science Learning Experiences During Early Adolescence

Li Sha, Christian Schunn, and Meghan Bathgate

Learning Research and Development Center, University of Pittsburgh, 3939 O'Hara St, Pittsburgh, Pennsylvania 15213

Received 31 May 2013; Accepted 12 January 2015

Abstract: Cumulatively, participation in optional science learning experiences in school, after school, at home, and in the community may have a large impact on student interest in and knowledge of science. Therefore, interventions can have large long-term effects if they change student choice preferences for such optional science learning experiences. To be able to track K-12 students' intentions to participate in optional science learning experiences, we developed a new measure of science choice preferences in early adolescence. The present study with 284 5th and 894 6th graders from diverse school contexts (i.e., from the Bay Area and the Pittsburgh area) illustrates the value of applying Item Response Theory analyses to develop a measurement instrument. These analyses established the overall reliability of the instrument and each item in the scale, as well as the generalizability of the scale and individual items across subgroups by gender, by ethnicity, and by achievement levels in science. Further, preferences to participate in science were shown to be separate from preferences to participate in mathematics, engineering, or medicine. Finally, the science choice preferences measure is validated through replicated positive correlations with levels of science interest, self-efficacy, and learning achievement. © 2015 Wiley Periodicals, Inc. J Res Sci Teach 52: 686–709, 2015
Keywords: choice in science learning; early adolescence; item response theory IRT; psychological assessment

Strength in science, technology, engineering, and mathematics (STEM) fields is an indicator of a nation's ability to sustain itself and important for successful participation in the modern workforce. Therefore, STEM is an increasingly critical area of K-12 schooling. However, across ages and cohorts, students are becoming less motivated to choose and engage in science-related activities, courses, and careers (Glynn, Brickman, Armstrong, & Taasoobshirazi, 2011; Logan & Skamp, 2008; Vedder-Weiss & Fortus, 2011, 2012) leading to poor academic performance in science learning (Bryan, Glynn & Kittleson, 2011; Lee & Shute, 2010). For example, Bryan et al. (2011) found that students' achievement in science was significantly associated with their science self-efficacy and intrinsic motivation. From the perspective of social cognitive theory, motivated students are supposed to: i) proactively make academic choices and ii) actively engage in these choices in learning (Bandura, 2001).

Contract grant sponsor: Gordon and Betty Moore Foundation; Contract grant numbers: 2820; 3341; Contract grant sponsor: National Science Foundation; Contract grant number: DRL-1348468.
Correspondence to: L. Sha; E-mail: lishahn@yahoo.com
DOI 10.1002/tea.21210
Published online 20 March 2015 in Wiley Online Library (wileyonlinelibrary.com).

Numerous studies have been done in the fields of science education, educational psychology, and the learning sciences in measuring academic engagement and identifying the influences on engagement (Fredricks, Blumenfeld, & Paris, 2004; Jang, Reeve, & Deci, 2010; Linnenbrink & Pintrich, 2003; Pekrun & Linnenbrink-Garcia, 2012; Steele & Fullagar, 2009). However, choice has received much less attention as an outcome or indicator of motivation (e.g., Neuville, Frenay, & Bourgeois, 2007).

Therefore, more research is critically needed to explore what personal and contextual factors will facilitate or hinder K-12 students' choice of science activities, courses, and even science-related careers in the future. A necessary step is validly measuring their choice preferences in science learning, which will then support investigations of how science choice preference is associated with their personal and contextual factors. Accordingly, the present study is intended to develop a scale assessing middle school students' *choice preference* (CP) in science learning. Science choice preference in the present study is defined as the extent to which children prefer a science-related choice when given both science-related and non-science related alternative options (i.e., a psychological tendency toward a topical choice).

Theoretical Background

Children in the middle school years and beyond have a large amount of what the informal learning research community have called "free choice time," that is time in which children have a large amount of control over the topic and form of their experiences (Dierking & Falk, 2003). For example, most children in developed nations usually have large amount of time not occupied by schooling, eating, and sleeping when adding up after school, weekends, and vacation time. This total amount of this free choice time can exceed the number of hours in school, and thus presents an opportunity to significantly increase science learning time even if only a fraction of the free choice time is devoted to science (Bevan et al., 2010; Feder, Shouse, Lewenstein, & Bell, 2009). Some of this time can be spent on science learning in and around the home (e.g., reading science-related books, watching science-related TV shows or websites, exploring natural phenomena such as mixing chemicals or collecting insects) alone or with family and friends. Many children can also choose from a variety of optional organized activities related to science learning, such as participating in various after school clubs, weekend classes, or summer camps. While access to some of these opportunities is determined by family income and distance of home from urban centers, most communities have some relevant open-access opportunities (e.g., through public libraries or community organizations), and most homes have access to some relevant media (Madden, Lenhart, Duggan, Cortesi, & Gasser, 2013; Powers, Wilson, Keel, & Walton, 2013; Rectors & Sheffield, 2011).

If a significant amount of this free choice time is spent on science, science learning outcomes could be much higher. Also, given the diversity in content that could be accessed in free choice settings, science choice preference may create opportunities to deepen interest and identity in science, which would broaden participation in STEM-related careers and could also improve learning outcomes.

Prior Psychological and Educational Research on Choice

Choice in psychological research can be viewed either as an input for later motivation development or an outcome of motivation (Patall, 2012; Schunk, Pintrich, & Meece, 2008). Choice is most often treated as an input for or influence on motivation and learning, particularly in the framework of self-determination theory (SDT) (Katz & Assor, 2007; Patall, Cooper, & Robinson, 2008; Patall, Cooper, & Wynn, 2010; Patall, 2013). A basic tenant in SDT research is that people's intrinsic motivation will be improved when they consciously feel some degree of autonomy in controlling their own thoughts and actions (e.g., making a choice). Thus, providing students with choices is widely used to offer them a sense of autonomy for enhancing motivation and performance (Flowerday & Schraw, 2003; Patall et al., 2008). Researchers have also explored what personal factors (e.g., interest) (Patall, 2013) and context- and task-related factors (e.g., the type of choice, number of options) (Reber, Hetland, Weiqin, Norman, & Kobbeltvedt, 2009) influence the effects of choice on motivation and learning. In classroom settings, providing students with choices is usually manifested as a teaching strategy, such as allowing students generate their own solution to problem, or offering students choices about the materials to use in classroom (Katz & Assor, 2007; Patall et al., 2008). For example, Mortensen and Smart (2007) investigated how providing elementary school students with free-choice worksheets motivated them to participate in science learning activities in the science museum.

An alternative to treating choice as an input for improving motivation is to conceive of this construct as a behavioral outcome or indicator of motivation (Zimmerman, 2011). That is to say that students with higher motivation to learn science should choose more science-related activities in and out of school than those with low motivation. Prior research in this line have explored how individual motivational beliefs predict students' choices of tasks, activities, courses, and careers within the framework of expectancy-value theory (EVT, Simpkins, Davis-Kean, & Eccles, 2006; Wigfield & Eccles, 2000; Wigfield & Cambria, 2010). EVT involves two core motivational constructs. One is expectancy for success—individuals' self-beliefs about how well they will do on upcoming tasks/activities, which is conceptually similar to the notion of self-efficacy in social cognitive theory (Bandura, 1997). The other is a four-component construct—values that individuals attach to an immediate task/activity (e.g., learning activities in today's class), or future events (e.g., courses in the next term, career in the future). *Attainment value* refers to the importance of doing well on a given task; *interest value* refers to the enjoyment obtained from engagement of tasks/activities, which is similar to other motivational constructs such as intrinsic motivation (Ryan & Deci, 2000), and interest (Hidi & Renninger, 2006); *utility value* refers to the usefulness individuals attach to their future plan; and *cost* refers to any assessment of resources (e.g., time, efforts) one may need to accomplish a task. In EVT, students' choices of tasks/activities/courses are influenced by their expectations for success (personal efficacy) and the values they attach to the choices (Eccles, 2009).

Choice preferences in this motivation research have been defined and measured in two different ways: i) individuals' intended selection of courses/future college majors (which can be called as *hypothetical choices*), and ii) choices they had actually made (which can be called as *actual choices*). For example, Meece, Wigfield, and Eccles (1990) adopted a single item measure (a seven-point scale rating) of elementary and high school students' choices of mathematics course. That measure asked students whether they would take more math in the future if they no longer had to. It is an example of a measure of one's psychological tendency to make a choice from two options (taking or not taking math courses). Similarly, Hazari, Sonnert, Sadler, & Shanahan's study (2010) used a single item measure of high school students' intended choice of a physics career in which the students were asked to report the likelihood of choosing a career in the physical sciences (1—not at all likely to, 6—extremely likely). In these two studies, students only had two options (science-related versus not science-related), but the world actually consists of many complex alternatives such as art, sports, and business. For example, Parker et al., (2012) used a single-item measure of high school students' choices of university major from four options (physics/mathematics/engineering; life, biological, and medical sciences; humanities and social sciences; and law and business). The Science Aspiration and Career Choice scale (Archer et al., 2012; DeWitt, Archer, & Osborne, 2014) has five items, but is limited at least in two aspects: i) not providing any alternative non-science options, and ii) all the items are distant future-oriented (job/career) only. Stronger measures should consider a broader variety of choices and a broader variety of alternatives in each choice such that a choice measurement instrument is able to validly predict the tendency to participate in optional science learning opportunities (Kane, 2001; Morizot, Ainsworth, & Reise, 2007).

In science education research studying actual choices, a common outcome measure is students' selection of STEM courses, majors, and careers (Bøe, 2012; Cerinsek, Hribar, Glodez, & Dolinsek, 2012; Sjaastad, 2011). But such measures are of little practical use for studies of early adolescents because they are not given such choices, and many secondary schools throughout the world give students few to no choices in STEM course taking. Most importantly, beyond students' life-defining choices, few researchers have examined choices students need to make in their daily science learning in and outside schools (e.g., choices of learning tasks/activities/courses in school; choices of TV programs related to science at home, visiting science museums outside school).

Simpkins et al., (2006) measured student choices at a more micro level: the extent to which they actually had chosen to participate in math or science activities over the past year on a seven-point scale (0 = never, 6 = almost every day for a lot of time). They used children's actual choices of activities in Grade 5 to predict their later motivation (expectancy for success, value, and interest), and in turn how those motivational variables at Grade 6 and 10 predicted the number of math or science courses they actually chose throughout high school. However, Simpkins et al.'s (2006) choice measure was problematic as it did not consider whether it was the students who actually made the choice and what alternatives students had. In other words, it was a measure of the context as well as a measure of the child. This suggests that measuring the aspect of choice controlled by the student (i.e., the target of student experiences) will need to focus on hypothetical choices and consider a range of both science related and non-science related options (e.g., art, music, sports) that compete for free choice time and resources. Of course, contextual factors may also influence student motivation which in turn influences choices that were made. But it is important to differentiate choices made at least partially by students from choices made exclusively by others.

It should also be noted that the total of all actual choices over short time periods might be highly subject to particular competing time periods that may be very localized (e.g., a particular sports season or time of a school play). However, a sum of all actual choices over long time periods is not a convenient outcome measure for most research studies: it requires waiting a long time to collect, and either high vigilance if information is directly collected or regularly collected, or high noise from poor memory if collected infrequently.

In sum, actual choice measurement instruments have many practical and theoretical problems. First, any given actual choice will vary across children in the degree to which children are allowed to make choices entirely on their own (versus with family or friends also influencing the final choices). Second, the available actual choices are different across locations, and may involve too long a list to comprehensively study on a regular basis, and one that would need to be regularly updated based on changing programming. Hypothetical choice measures hold greater promise, but more diverse types of choices are required, particularly ones including both macro (e.g., choosing to be a scientist) and micro science choices (e.g., choosing to visit science museums) that also consider explicitly the alternative choices that compete for time and resources.

Conceptualizing Science Choice Preference

In daily life situations, choice may occur when one selects one alternative from among similarly attractive but indeterminate options (Williams, 1998). Individuals are motivated to choose the best or most rewarding option among the alternatives given their anticipations and the information they have about the circumstances under which they are restricted in making that choice (Patall, 2012). Making a choice essentially involves psychological processes and behaviors of judgment and decision making. From a cognitive view of decision making, Fuzzy-Trace Theory postulates that people's judgment and decision making are based on simple, gist-like schema of options (Reyna & Brainerd, 1995; Reyna, Adam, Poirier, LeCroy, & Brainerd, 2005). Gist refers to one's semantic representation of information reflecting her or his knowledge, value, culture, and developmental level through direct and indirect experiences (Reyna et al., 2005). In other words, the mental representation driving decision making is multi-dimensional in terms of content and development.

Students in school or out-of-school often need to make a choice from among multi-attributes alternatives differing in importance or preference. In schools in the US, students typically have the option of stopping after two high school courses in science (Sheppard & Robbins, 2005), and many do opt out. In addition, out-of-school learning represents a large proportion of waking hours; as children become older, the amount of time spent in child-selected, optional, out-of-school learning contexts (reading, club, and summer learning) can produce significant science learning opportunities (Bevan et al., 2010). Finally, career interest by the end of eighth grade has been found to be good predictor of who obtains a STEM degree (Tai, Liu, Maltese, & Fan, 2006).

The large learning opportunity represented by the choices and the predictiveness of the choice tendencies suggest that researchers and teachers in science education need to know the extent to which children prefer a science-related choice when given both science-related and non-science related alternative options, i.e., science choice preference. Specifically, for researchers to be able to fully examine the ways in which motivation influences learning, researchers need to study children's science choice preferences rather than studying just engagement and learning behaviors. For classroom teachers, assessing science choice preference provides a tangible intermediate outcome of their instruction that can have larger later effects on student learning. It could also be used to assess other arenas of educator intervention such as students' openness to suggestions for out-of-school learning opportunities, or gaps between student preferences and current regional opportunities for students (e.g., to organize an after-school club).

Therefore, framed within Fuzzy-Trace Theory (Reyna & Brainerd, 1995; Reyna et al., 2005), we conceptualize students' choice of science experiences as reflecting their holistic mental representation of experiences with science. We define science choice preference as the tendency to choose a variety of activities (e.g., attending science class, participating in a scientific experiment), in various settings (e.g., in school, outside of school), at different points in time (e.g., present, proximal, or distant future). Science choice preference is influenced through changing motivational levels and changing conceptions of what science is, and it also drives later motivation levels and conceptions of science through expanded learning opportunities. In other words, it is rooted in the contemporary view that cognition is dynamic and situational (Tschacher & Scheier, 1999).

Design Principles for an Effective Science Choice Preference Instrument

A core part of the mental representation that determines individuals' science choice preferences is the issue of *what science is* (i.e., their perception of science) and thus which choices reflect this conception. Students' conception of science is shaped and developed over time by the context in which students experience science (Zimmerman, 2012), and especially the authenticity of the science that is experienced (Chinn & Malholtra, 2002). Accordingly, several measurement considerations should guide the design of an effective science choice preference instrument.

First, science education research has revealed variation in students' perception of what kinds of situations in real life are scientific (Mantzicopoulos, Samarapungavan, & Patrick, 2009). While there may be learning domains with indirect relationships to science (e.g., medicine and engineering), they can only be included as part of science choice preference to the extent to which children see the choices as science-related, as preferences are built upon perceptions of the world. For instance, when individuals are required to make a choice from among possible careers, student A may chose doctor or engineer because in her mind it is a science-related career. Alternatively student B, who is generally interested in science-related careers, may not choose doctor or engineer simply because, in her mental model of science, neither doctor nor engineer has anything to do with science. Mathematics, sometimes described as the language of science in science standards documents, is similarly ambiguous. For example, a child might choose to participate in a mathematics club because it is part of their larger science schema. But math-related items should not be included in a science choice preference scale if only a few students select those options because of their connections to science but most other students make choices about the options because of some other reason (e.g., presence or absence of math anxiety).

Second, in accordance with sociocultural perspective of learning, context may play a vital role in individuals' perception of science, regardless of expertise or developmental stages (Buldu, 2006; Mantzicopoulos et al., 2009). Thus science choice preference is conceptualized as an individual's psychological tendency (a state) to make choices toward science in the contexts in which choices are made. A scale aimed to measure individuals' choice of science should embrace the options that are concrete and representative of the broader set that are commonly available, i.e., a mix of in school and out-of-school activities. At the same time, this consideration may require the creation of a scale that is specific to particular age groups because the commonly available choices or the choices perceived as relevant to science may be specific to particular age groups. We focus on the choices broadly relevant to early adolescents because of prior research suggesting that middle school is a key transitional point across K-12 schooling at which a decline in science motivation begins (Wigfield, Byrnes, & Eccles, 2006).

Third, contextual influences such as family background and individual factors such as gender may frame perceived choices. For example, Buldu's study (2006) found young children's perception of scientists varied as a function of their socio-economic status. Both Mantzicopoulos et al. (2009) and Buldu (2006) suggest a relation between the children's age and their stereotyped perception of science and scientists. Many existing studies in science education research revealed a pervasive gender effect on K-12 students' attitude toward science and science learning (Britner, 2008; Patrick, Mantzicopoulos, & Samarapungavan, 2009; Simpkins et al., 2006). Gender differences in academic interest usually begin to emerge in the middle school years (Meece & Painter, 2008). We examine whether gender or age shapes which choices reflect one's science choice preferences. For example, certain choices might be avoided because they are seen as not appropriate for girls or for kids of this age rather than representing participation in science.

The Current Study

The present study was aimed to develop and validate a new scale measuring early adolescents' *science choice preferences* that better reflects the concerns identified from the review of the past literature about conceptualization and measurement via actual and hypothetical choice. A valid measure with good precision should have acceptable internal consistency and unidimensionality (homogeneity). Internal consistency indicates the overall degree of interrelation among a set of items (Simms & Watson, 2007). Unidimensionality refers to the extent to which all of the items in a scale converge on a single latent trait, i.e., that participants differ only on a single latent trait (e.g., science choice) regardless of gender, ethnicity, self-efficacy, interest, or achievement.

Factor analysis methods can be used to establish unidimensionality, but Item Response Theory (IRT) methods provide additional insights into critical measurement properties of a scale. IRT methods have been widely applied to educational assessments such as aptitude tests, knowledge, language tests, and are now beginning to be used in the analysis of other kinds of scales (Embretson & Reise, 2000; Fraley, Waller, & Brennan, 2000; John & Soto, 2007). The present study applied IRT methods to test explore the breadth of student's conceptions of science in science choice preferences as well as to test measurement invariance of the scale across time, demographics, and motivation levels. Scales are not useful for studying effects of demographics of motivation levels on choice or cummulating results across studies done in different contexts if location, demographics, or motivation levels change the measurement properties themselves. The IRT method and its use in this study are detailed in the Methods section.

We test scale validity from multiple perspectives (Kane, 2001). First, from a content validity perspective, we begin with a broader set of choices to overcome the identified limitations of the prior narrow science choice scales (Archer et al., 2012; DeWitt et al., 2014; Simpkins et al., 2006). Second, because motivational research conceives of choice as an important outcome of motivation, from the perspective of concurrent validity, we test whether our science choice preference scale is significantly associated with motivation to learn science. Specifically, research has found that both self-efficacy and interest are powerful predictors for students' learning behaviors and achievement (Britner & Pajares, 2006; Hidi & Renninger, 2006), so we examine the associations of choice preferences with science self-efficacy and interest.

Thus, to strongly establish the reliability and validity of the new science choice preferences scale, we apply factor analysis, IRT analysis, and cross-scale correlation analyses. No prior study of choice has gone beyond a simple reliability analysis of the measures they used. Our in depth analyses are organized around three specific research questions:

1. How broadly conceptualized are student conceptions of science choice preferences? Specifically, are choice preferences toward mathematics, engineering, or medicine part of science preferences in middle school students?
2. Are the psychometric properties (e.g., discrimination) of the science choice preferences scale consistent over location, time, gender, ethnicity groups, self-efficacy levels, interest levels, and achievement levels?
3. To what extent is the measure of science choice preferences associated with self-efficacy beliefs and interest in science learning?

Methods

The research questions required testing the reliability, validity, and generalizability of the science choice preference scale across very different contexts. Therefore, two cohorts of students were selected to be different on many dimensions. First, Cohort 1 was from the Pittsburgh area, a historically blue-collar, industrial-focused urban region in the Eastern US; Cohort 2 was from the Bay Area, a highly diverse, technology-focused urban region in the Western US. Those two regions are different from one another in many aspects such as ethnic composition (see Table 1 below). Second, since the participants from the Pittsburgh area were 6th graders, and those from the Bay Area were 5th graders, this makes it possible to examine if whether the scale functioned similarly across ages. Third, within the Pittsburgh cohort, the data were collected at two points in time (i.e., Test 1 and Test 2), enabling us to assess whether the psychometric properties of the scale is consistent over time (i.e., appropriate for longitudinal research). In short, the study design was not created to provide a clean test of age or region, but instead to provide a generalization test of the scale's properties across substantially different contexts, using the larger and more complete Pittsburgh dataset as the in-depth investigation and the smaller Bay Area dataset as a

Table 1
Demographic information in the two cohorts of participants

	Pittsburgh Cohort Mean Age = 12.0 (SD = 0.5)	Bay Area Cohort Mean Age = 10.5 (SD = 0.5)
Gender		
Boy	350 (50.7%)	135 (51.9%)
Girl	340 (49.3%)	125 (48.1%)
Ethnic information		
White	261 (47.9%)	73 (32.6%)
Asian	15 (2.8%)	8 (3.6%)
African American	246 (45.1%)	25 (11.2%)
Hispanic	19(3.5%)	111 (49.6%)
Native American	4 (0.7%)	7 (3.1%)

generalization test. Related to instrument validity, data were also collected on two motivational variables and an achievement test.

Participants and Procedure

For the Pittsburgh cohort, 894 Grade 6 students from 10 public schools (38 classrooms) in the Pittsburgh area, participated in Test 1 and Test 2. For the Bay Area cohort, 284 Grade 5 students from four schools participated. Table 1 shows that the Pittsburgh and Bay Area cohorts simultaneously differ in age and ethnicity distributions. The majority in the Pittsburgh cohort was Caucasian and African Americans, while the majority in the Bay Area cohort was Hispanic and Caucasian.

For the Pittsburgh cohort, Test 1 was conducted in September and early October, and Test 2 was conducted approximately 5 months later, in early February. The choice preference scale was administrated in both Test 1 and Test 2. Meanwhile, the students also completed a survey on their levels of self-efficacy and interest in science learning at Test 1. All 10 schools in Pittsburgh were currently teaching the same 5-month-long unit on weather and climate from Full Option Science System (FOSS, http://www.lhsfoss.org/), a hands-on curriculum developed by the Lawrence Hall of Science. A knowledge test on weather and climate was conducted as a part of Test 2. The Bay Area cohort participated only in Test 1 in late October of that same year, completing the survey questionnaire on the science choice preferences, and self-efficacy and interest.

Science Choice Preference Scale

We developed a five-item CP survey involving a total of 29 different options as our CP (i.e., five items, each with 4–7 alternative response options). Total scale information (e.g., reliability coefficient) is a function of the number of items in a scale, and scale lengths could affect responses (Hinkin, 1995). Scales with too many items can create problems with respondent fatigue or response biases, but scales with too few items may lack content and construct validity (Hinkin, 1995). Adequate internal consistency reliabilities can be obtained with as few as three items (Cook et al., 1981, as cited in Hinkin, 1995). According to this logic, we believe that five items could be appropriate for the CP scale.

We explore through IRT analyses in which the 29 options are consistently associated with an overall science choice preferences construct. This multi-item, multidimensional scale is intended to expand the existing choice instruments in science learning (Archer et al., 2012; DeWitt et al., 2014; Hazari et al., 2010; Simpkins et al., 2006) in four aspects: *hypothetical-focus*, *content*, *context*, and *time*. Specifically, it uses only hypothetical choices rather than reporting of actual choices. As to content and context, one item is about class choice (i.e., item 1); three items are about different science-related activity choices (i.e., item 3, 4, and 5); and one item is about career choice (i.e., item 2). In addition, the choice preference scale involves diverse science-related choices (e.g., engineering, medicine, and mathematics), as well as non-scientific options (e.g., history, art; item 1 and 2). As to time, three items are about present/immediate choices (today; item 3 and 4); one is about proximal future (next year; item 1), and one item is about distant future (choosing to be scientist; item 2). The distant future item is similar to the distant-future focus of the Science Aspiration and Career scale (Archer et al., 2012; DeWitt et al., 2014).

All but one item required students to choose one option among the choices because at a given moment in time, only one from the set is typically possible, and items are simply scored as selecting the science choice or not (e.g., selecting the science museum versus the other museums). As future choice situations sometimes allow for selecting multiple options, for the item asking children to pick classes for the next year, the participants were allowed as many classes as they wanted. As long as science class was selected, 1 is given to this item regardless the number of choices a student selected (e.g., someone chose more than one class) otherwise, 0 is given to it.

For three of the items, there were choices that were not precisely about science, but were sufficiently related to science that children may have encoded them as science-related choices. Therefore, we explored whether these choices should also be counted as evidence of a science choice. For example, the career choice item included options of engineer and doctor. Similarly, the class choice item included a math class option and the activity choice item included a mathematics puzzle. We used IRT analyses (described below) to determine whether mathematics, medicine, and engineering were part of a science-related field of choices in the minds of children at this age or whether they were in fact conceptually distinct choices that are not part of pro-science choice preferences.

Measures on Self-Efficacy, Interest, and Learning Achievement

In order to provide concurrent validity information about the choice scales and to establish measurement invariance across learner populations, data were also collected on children's *self-efficacy* and *interest in science* (see supplemental materials for full scales). Each scale was developed as part of a larger research effort aimed at understanding 6th grade science learning and motivation development in and out of school; items were adapted from the literature to apply to both in and out of school (rather than typical items that focus on science in just school or just a out-of-school) and be relevant to 6th graders (rather than those more relevant to college-aged or high school-aged populations). Cognitive interviews were conducted with children to validate new items.

Both self-efficacy and interest data were collected at Test 1. Detailed introduction to the validation work of the two self-developed motivational instruments is beyond the present study, but reliability and confirmatory factor analysis fit statistics are included here. Self-efficacy was measured as a mean score across 9 four-point (YES!, yes, no, NO!) Likert scale items (Cronbach's $\alpha = 0.85$; e.g., I think I am pretty good at: Coming up with questions about science). Its measurement is not only subject specific (i.e., science), but also specific to critical aspects of science (e.g., capability of coming up with questions about science) rather than simply asking the degree of general confidence in learning science. By doing so, the students at this age were brought into a more specific context in which a fine-grained measurement of their self-efficacy can be obtained. The CFA fit indices ($\text{CFI} = 0.96$, $\text{RMSEA} = 0.06$) indicate acceptable homogeneity of this nine-item self-efficacy scale, according to the typical thresholds for an acceptable CFA model of $\text{CFI} >= 0.90$ and $\text{RMSEA} < 0.08$ (Brown, 2006).

The interest variable was computed as a mean score across 10 four-point Likert scale items (Cronbach's $\alpha = 0.91$; e.g., "When I work on science at school, I: like it-dislike it"). The 10 items were intended to not only cover a wide range of possible interest-related psychological states while learning science in and out of school, but also includes both affective and cognitive elements of interest. In contemporary research, interest is composed of both affective and cognitive aspects (Hidi & Renninger, 2006). Thus, rather than simply asking if they are "interested in" or "liked" science, the items included related psychological states like happy, amazed, curious, excited as elements of the interest construct. The CFA fit indices (CFI = 0.96, RMSEA = 0.07) indicate acceptable homogeneity of this 10-item interest scale.

The achievement test, collected at Test 2, was developed to assess the big ideas found in the curriculum unit on weather and climate, and consisted of 21 multiple choice questions (Cronbach's $\alpha = 0.78$; e.g., What is the primary energy source that drives all weather events, including precipitation, hurricanes, and tornadoes? (a) the Sun, (b) the Moon, (c) Earth's gravity, or (d) Earth's rotation. Only the Grade 6 students in Pittsburgh took part in the knowledge test, taking the test after 5 months of instruction on this topic.

The ranges of percentage of missing data (item non-response rate) for each of the nine self-efficacy items, the 10 interest item, and the 21 achievement test items are respectively, 1.7–5.7%, 2.2–11.7%, and 8.4%. Missing items were dropped from the computation of mean scores. While there is currently no agreement upon threshold defining problematic percentage in psychological studies, published papers commonly range from 5% to 20% missing items (Schlomer, Bauman, & Card, 2010). Thus, the percentages of missing data in the present study should not be problematic.

A Brief Introduction to IRT and its Use in the Present Study

Since the science choice preference scale will be systematically validated here through more advanced methods from IRT (Embretson & Reise, 2000; Fraley et al., 2000; John & Soto, 2007; Morizot et al., 2007), a brief introduction to key aspects of this psychometric method and details of our use of IRT are given below.

IRT is a psychometric approach which assumes that an individual's response to a particular item is influenced by a combination of an overall characteristic of the individual (i.e., a trait or ability) and properties of the item. Different IRT models vary in how many properties (called parameters) are associated with each item. In the present choice measurement, whether or not individuals endorse a choice item (e.g., choosing to do science experiments in school) is assumed to be affected by two item features (and therefore is called a two parameter model): i) the discriminability of the item and ii) the difficulty level of the item. An item's discrimination indicates the relevance of the item to the trait being measured by the test, similar to factor loading in confirmatory factor analysis (CFA). A large positive discrimination value indicates relatively high consistency between the item and the underlying trait being measured. A discrimination value of 0 means that the item is unrelated to the underlying trait. Discrimination values will be used to determine whether the items about engineering, medicine, and mathematics choices are associated with science choice preferences. Discrimination values can also be tested statistically across subgroups of students test whether the item is stable in the extent to which it assesses the construct across subgroups (e.g., whether both boys and girls consider science club participation as indicative of general choice preferences toward science).

Mathematically, item difficulty is quantified in terms of trait level denoted as θ and represented by an IRT scale score rather than a raw score (Embretson & Reise, 2000). IRT scores are usually computed on a standardized metric with a mean of 0 and standard deviation of 1. Specifically, an item's difficulty is defined as the trait level required for participants to have a 0.5 probability of endorsing the item. For example, if an item's difficulty is 0, an individual with an average trait level (i.e., whose trait level is 0) will have a 50/50 chance of endorsing the item. If an item's difficulty is 1.5, then an individual having a trait level of 1.5 will have a 50/50 chance of endorsing the item. Participants whose trait levels are lower than 1.5 will have lower than 50/50 likelihood of endorsing the item. Good scales have a distribution of items that vary in difficulty values in order to optimally differentiate students at both high and low ends of the scale. Motivational scales can sometimes have poor differentiation of students at the high end of the scales because they have no high difficulty items (Stake & Mares, 2001).

The IRT models in the present study were run in IRTPro (SSI, http://www.ssicentral.com/irt/IRTPRO_by_SSI.pdf). It has Graphical User Interface and runs in Microsoft Windows. The data to be processed can be imported into IRTPRO directly from SPSS data files, and outcomes for the selected IRT model (e.g., Two-Parameter Logistic) are generated automatically in table form (e.g., discrimination, difficulty, Goodness-of-Fit parameters).

Results

The results are organized by three research questions. First, to characterize the breadth/narrowness of student choice preferences, we begin with IRT analyses focused on identifying which items are associated with science choice preferences as a scale. Second, we examine the consistency of the scale over time and across student subgroups by gender, ethnicity, self-efficacy levels, interest levels, and achievement levels. Third, we validate the choice preferences scale by testing whether it is associated with two motivational variables (science self-efficacy and interest).

Description of the Items and the Scale

Table 2 presents the frequencies with which each possible science item was selected in each of the three data collections. Note that Ns vary across items because of students skipping items or failing to complete a measure from absences. There is a small amount of variation in item frequencies across data collections. While there is considerable variation in popularity of items, all items are selected by some children. What is unclear from simple frequencies, however, is whether item popularity is driven by relevance to science interests or just general popularity with children at this age, regardless of interest in science (e.g., visiting science museums could simple reflect an interest in seeing what dinosaur looks like).

Table 2
Selection frequencies of each choice preference item across age and locations

	Bay Area	Pittsburgh	
	Early 5th Grade ($241 < n < 284$)	Early 6th Grade ($620 < n < 894$)	Mid 6th Grade ($703 < n < 894$)
Science class	57%	39%	37%
School activity	25%	17%	13%
Home activity	30%	25%	23%
Science museum	50%	28%	39%
Scientist	5%	5%	5%
Math class	43%	50%	49%
Math activity	7%	5%	5%
Math puzzle	14%	8%	24%
Doctor	12%	9%	9%
Engineer	6%	3%	5%

Are Choices of Mathematics, Engineer, and Doctor Items Indicators of Preferences Toward Science?

In order to answer the above theoretical question and basic scale construction question, Figure 1 shows the discrimination of each item in the 10-item choice preference IRT model. An item's discrimination indicates the relevance of the item to the trait being measured by the test, which corresponds to factor loading in confirmatory factor analysis (CFA; Embretson & Reise, 2000). A large positive discrimination value indicates a good consistency between the item and the underlying trait being measured. A discrimination value of 0 means that the item is unrelated to the underlying trait. An item with a negative discrimination value is inversely related to the underlying trait, i.e., individuals with high trait level are less likely to endorse the item.

The three math items and the doctor and engineer items always had discrimination values below +0.5, a common threshold for acceptable discrimination (Fraley et al., 2000). This consistently low discrimination indicates that they are not indicators of preferences toward science and should be removed from the choice preference scale. By contrast, the discrimination values of the other five science items are larger than 1 in the three cases, and clearly should be retained in the scale. After deleting the problematic items, the remaining analyses focus on the psychometric properties of the five-item CP scale. Figure 1 shows the discrimination values of each item in the choice preference scale over time and in the Bay Area early 5th grade, Pittsburgh early 6th grade, and middle 6th grade in the 10-item IRT model.

Overall Scale and Item-Level Psychometric Properties

Past psychometric research has found that χ^2 is driven too heavily by sample size (Healey, 2012), and depends on full information contained in the contingency table (Cai, Maydeu-Olivares, Coffman, & Thissen, 2006; Cai & Hansen, 2013). When there are empty cells in the table (i.e., limited information), Cai and colleagues' studies (2006, 2013) suggest that researchers use limited information Goodness-of-Fit M_2 instead of χ^2. The Goodness-of-Fit parameters in Table 3

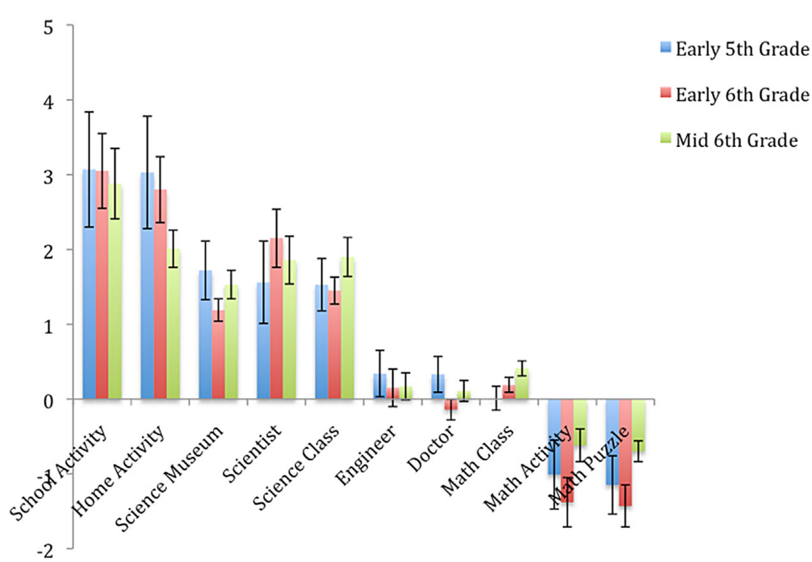

Figure 1. Discrimination values of each item in the choice preference scale over time and across settings in the 10-item IRT model.

Table 3
Overall Goodness-of-Fit parameters for the three IRT models

	Bay Area	Pittsburgh Cohort	
	Early 5th Grade	Early 6th Grade	Middle 6th Grade
M_2 value	3.28 (df = 5, p = 0.66)	8.04 (df = 5, p = 0.15)	2.05 (df = 5, p = 0.84)
RMSEA	0.00	0.03	0.00

indicate that overall the five-item 2PL IRT model fits the three datasets well according to the commonly used thresholds of $p > 0.05$ and RMSEA < 0.05.

The key IRT parameters (difficulty and discrimination) for each item level are shown in Table 4. The discrimination values of all five items were generally good across all contexts (i.e., well above 1). Difficulty also varied considerably across items from high positive to negative items, indicating the scale comprised items useful for differentiating students across a wide spectrum of overall science choice preferences. The scientist item had the highest difficulty, suggesting that choosing it require a high level of science choice preference compared to choices of other activities. By contrast, taking an extra science class was relatively easiest to endorse overall.

Note that the difficulty of some items (the museum visit item) varied substantially across contexts, even when the discrimination did not. That is, the choices always reflected choices in science, even when the relative intensity of preferences required to endorse a particular item varied.

Are the Items Equally Informative Across Gender, Ethnic Groups, Levels of Achievement, Level of Self-Efficacy, and Level of Interest?

Differential item function (DIF) analyses were conducted on the item discrimination values to determine whether the measures of choice preference can be generalized across gender, ethnic groups, levels of self-efficacy, levels of interest, and levels of learning achievement. We created median-split high/low groups on the two motivational variables (self-efficacy, interest) among both the Pittsburgh and Bay Area cohorts, and on achievement in the Pittsburgh cohort. Students are excluded from a given analysis when the relevant motivation, performance, or demographic information was missing. We combine White and Asian as the traditionally over-represented

Table 4
Mean (and standard error) discrimination and difficulty of science choice preference across ages and locations in the 5-item models

	Discrimination			Difficulty		
	Bay Area	Pittsburgh Cohort		Bay Area	Pittsburgh Cohort	
Science choice preference	Early 5th Grade	Early 6th Grade	Mid 6th Grade	Early 5th Grade	Early 6th Grade	Mid 6th Grade
School activity	2.7 (0.6)	3.2 (0.6)	2.8 (0.5)	0.7 (0.1)	0.7 (0.1)	0.7 (0.1)
Home activity	3.2 (0.8)	2.7 (0.4)	1.7 (0.2)	0.5 (0.1)	0.4 (0.1)	0.4 (0.1)
Scientist	1.8 (0.6)	2.2 (0.4)	1.8 (0.3)	2.3 (0.4)	1.9 (0.2)	1.8 (0.2)
Extra class	1.9 (0.4)	1.4 (0.2)	1.7 (0.2)	−0.2 (0.1)	−0.1 (0.1)	−0.2 (0.1)
Museum visit	1.7 (0.3)	1.4 (0.2)	1.5(0.2)	−0.3 (0.1)	0.4 (0.1)	−0.4 (0.1)

groups in science and the remaining ethnicities as the traditionally under-represented groups in science. Table 5 shows that of the 70 possible differences on the five binary variables, only a small number was statistically significant (e.g., the discrimination value of the scientist item for the early 5th grade girls in Bay area is higher than boys at $p < 0.001$).

We note several important points in these DIF analyses. First, all of the items are positive indicators of science preferences regardless of gender, ethnicity, achievement, self-efficacy, and interest. Second, the discrimination of all items is relatively consistent across subgroups, but the

Table 5
Discrimination values (with SEs) of each science choice preference item as a function of gender, ethnicity, classroom achievement, interest, self-efficacy in science across ages and locations

	Bay Area		Pittsburgh Cohort			
	Early 5th Grade		Early 6th Grade		Mid 6th Grade	
	Girls $n=125$	Boys $n=151$	Girls $n=346$	Boys $n=350$	Girls $n=346$	Boys $n=350$
School activity	3.7 (1.6)	2.3 (0.6)	2.5 (0.5)	3.6 (0.5)	3.6 (1.3)	3.3 (1.0)
Home activity	3.1 (1.1)	3.0 (1.0)	3.3 (0.2)	1.9 (0.3)	1.7 (0.3)	1.6 (0.3)
Scientist	23.9 (7.9)	1.5* (0.6)	1.9 (1.1)	1.9 (0.4)	2.2 (0.8)	1.6 (0.4)
Extra class	1.4 (0.4)	1.8 (0.5)	1.1 (0.6)	1.6 (1.3)	2.6 (0.8)	1.9 (0.4)
Museum visit	2.5 (0.9)	1.8 (0.5)	1.3 (0.3)	1.2 (0.2)	1.5 (0.3)	1.7 (0.3)
	Others $n=130$	W/A $n=67$	Others $n=275$	W/A $n=276$	Others $n=275$	W/A $n=276$
School activity	2.7 (0.8)	3.3 (1.5)	25.2 (1.9)*	3.0 (0.8)	3.8 (2.3)	3.7 (2.4)
Home activity	2.3 (0.7)	5.6 (4.8)	2.0 (0.4)	2.7 (0.7)	1.5 (0.4)	1.6 (0.4)
Scientist	2.1(1.1)	0.7 (0.9)	1.9 (0.6)	2.2 (0.7)	1.9 (0.7)	1.9 (0.6)
Extra class	2.1 (0.6)	1.9 (0.8)	1.4 (0.3)	1.4 (0.3)	1.5 (0.4)	2.9 (1.0)
Museum visit	3.0 (1.1)	1.8 (0.7)	1.1 (0.3)	1.2 (0.3)	1.5 (0.4)	1.3 (0.3)
	Low Ach (N/A)	High Ach (N/A)	Low Ach $n=415$	High Ach $n=407$	Low Ach $n=415$	High Ach $n=407$
School activity			4.2 (1.5)	3.2 (0.6)	3.1 (1.0)	4.0 (1.7)
Home activity			3.1 (0.8)	3.1 (0.6)	1.5 (0.3)	2.0 (0.4)
Scientist			2.1 (0.6)	3.2 (0.8)	2.6 (0.7)	1.6 (0.4)
Extra class			1.4 (0.3)	1.6 (0.8)	1.7 (0.3)	2.1 (0.4)
Museum visit			1.3 (0.3)	1.8 (0.3)	1.6 (0.3)	1.4 (0.3)
	Low SE $n=102$	High SE $n=106$	Low SE $n=308$	High SE $n=293$	Low SE $n=437$	High SE $n=378$
School activity	1.9 (0.6)	3.3 (1.3)	2.0 (0.4)	3.1 (0.9)	4.6 (4.2)	2.4 (0.4)
Home activity	39.9* (2.5)	1.6 (0.5)	6.7* (2.0)	2.0 (0.4)	1.5 (0.3)	1.9 (0.3)
Scientist	1.2 (1.3)	1.8 (0.8)	2.6 (0.8)	1.4 (0.9)	2.4 (0.7)	1.7 (0.4)
Extra class	2.0 (0.6)	1.4 (0.5)	1.0 (0.2)	1.3 (0.3)	2.2 (0.5)	1.6 (0.3)
Museum visit	2.0 (0.6)	2.7 (1.3)	0.9 (0.2)	1.5 (0.3)	1.2 (0.2)	2.5* (0.5)
	Low Int $n=124$	High Int $n=92$	Low Int $n=300$	High Int $n=306$	Low Int $n=417$	High Int $n=403$
School activity	15.4* (2.2)	1.3 (0.4)	2.8 (0.8)	3.2 (1.1)	2.5 (0.9)	2.4 (0.6)
Home activity	1.8 (0.5)	37.4* (4.0)	3.7 (1.4)	2.4 (0.6)	1.3 (0.3)	1.5 (0.3)
Scientist	0.6 (0.7)	1.4 (0.7)	2.0 (0.7)	2.2 (0.6)	2.7 (1.0)	1.4 (0.3)
Extra class	1.6 (0.5)	1.2 (0.4)	1.4 (0.3)	1.1 (0.3)	1.6 (0.4)	1.4 (0.3)
Museum visit	1.6 (0.5)	4.8 (4.0)	1.2 (0.3)	1.7 (0.4)	1.1 (0.3)	1.6 (0.3)

*$p < 0.05$.

occasional variance of each item in isolated cases highlights the importance of having a scale with more than just one item to be a reliable and valid indicator of science choice preferences across subgroups (Hazari et al., 2010; Simpkins et al., 2006). For example, the most discriminating choices, school activity and home activity, show high variability in discriminability across subgroups, and the commonly used career preference is sometimes the weakest indicator. Every item shows at least 2:1 variations in discriminability across subgroups even in the larger Pittsburgh cohort.

Table 5 shows that there are four exceptionally high discrimination values of the four items in the Bay Area cohort, and one exceptionally high discrimination value in the Pittsburgh cohort (roughly 3.5% of the total 140 discrimination values in the two cohorts). For example, the discrimination of the home activity (39) is extremely high among the low self-efficacy group. Mathematically, the discrimination parameter represents the slope of the middle section of the item characteristic curve (an index of steepness of the curve), and theoretically its range is between $-\infty$ and $+\infty$, with typical value less than 2.5 (Baker & Kim, 2004). The x-axis of the item characteristic curve is the standardized latent-trait continuum (i.e., choice preference levels in this case); the y-axis represents the probability of endorsing the item with a level of CP (standardized). In the case of positive discrimination values, an extremely high discrimination indicates a very steep curve, which, in turn, suggests the probability of endorsing the item is almost 1 for the students whose total CP scores are greater than the mean score (upper group), and the probability of endorsing the item is almost 0 for the students whose total CP scores are less than the mean score (in the lower group). This suggests that an item with extremely high discrimination makes a clear-cut distinction between the upper group and the lower group, but poorly discriminates the students within either of the two groups, compared to the items with typical discrimination values. It is possible that the activities with very high discrimination values were rarely available to these students and thus required strong demonstration of preferences among these students. Alternatively, in the low motivation cases, it may be that too few learners had high science choice preferences and thus the estimate of the more difficult items was unstable (i.e., can be treated as outliers). It is unclear why the home activity was so discriminating among the high interest subgroup in the Bay Area 5th graders. Compared to the data obtained from the Bay Area, the discrimination values in the much larger Pittsburgh cohort are generally consistent across a variety of subgroups.

Are Science Choice Preferences Associated With Self-Efficacy Beliefs, Interest, and Achievement in Science Learning?

Prior to addressing this research question, we first examined the scale reliability issue from the perspective of Classical Test Theory by computing the internal consistency between all items that the scale contains, i.e., the scale's reliability coefficient. Because the five CP items are binary, reliability was computed using Armor's θ (1974) instead of Cronbach's α. Mathematically, Armor's θ is a function of the number of items in the scale and the largest eigenvalue from the principal component analysis of the correlation matrix of the items of the scale (Armor, 1974). Statistically, unlike α, θ is not affected by the skewness of the item response distribution, and thus provides a better reliability estimate than coefficient α for all scales with skewed distributions (Zumbo, Gadermann, & Zerisser, 2007). The reliability coefficients from the three measurements of science choice preference were all good (see Table 4). These high coefficients further justify the conclusion drawn from the above IRT analyses that the scale developed in the present study is psychometrically reliable in general, in other words, the five CP items statistically converge on a unique psychological construct—*science choice preference*.

We assessed concurrent validity by examining the correlations of individual choice preference scores with two motivational variables: self-efficacy and interest, and the achievement score (only from the Mid 6th graders). Within each context, children were divided on self-efficacy, interest, or achievement measure using a median split. ANOVA analyses (shown on Table 6) revealed shows that interest effects on choice preference were significant in both locations for CP, and the significant self-efficacy effect was observed among both the early 6th graders and mid 6th graders in Pittsburgh. The mid 6th graders with high CP got high achievement scores in science learning.

Did the CP measure show sensitivity to either self-efficacy or interest differences across a wide range of each variable? The four panels shown in Figure 2 show that there was indeed a monotonic (and nearly linear) relation of self-efficacy and interest with choice preference. The *x*-axis on the two figures represents equal proportion bins by self-efficacy (on the left), and interest (on the right); the *y*-axis represents the mean choice preference score for participants falling in each bin. This visual pattern was found to be statistically significant with a standard multiple regression ($F(3,451) = 19.75$, adjusted $R^2 = 0.11$, $p < .001$) in which the interest ($\beta = 0.22$, $p < .001$) and self-efficacy ($\beta = 0.10, p < 0.05$) were the predictors, and the choice preference was the outcome.

Discussion

We focus on two significant contributions from the current student of early adolescents' science choice preference, which we defined as the extent to which children tend to prefer a science-related choice when given both science-related and non-science related alternative options. Overall, the developed science choice preference scale was found to be reliable and valid across a diverse group of students and contexts. Second, the empirical patterns across the items in the instrument have implications for science teaching and learning. We discuss each point below.

An Innovative Expansion to the Measurement of Choices in Science

It is widely realized that K-12 students become less motivated to choose and engage in science-related activities, courses, and careers (Glynn et al., 2011; Logan & Skamp, 2008; Vedder-Weiss & Fortus, 2011, 2012), thus, exploring motivational factors influencing children's science choices is an important focus for researchers in the fields of science education and

Table 6

Reliability coefficients (Armor's θ) and ANOVA on group difference of mean CP scores (with standard errors) between boys and girls, and levels of interest and self-efficacy across ages and locations, and levels of learning achievement (Middle 6th grade only)

	Bay Area	Pittsburgh Cohort	
	Early 5th Grade	Early 6th Grade	Middle 6th Grade
Armor's θ	0.81	0.82	0.79
Boys	1.79 (0.13)	1.70* (0.08)	1.33 (0.07)
Girls	1.58 (0.14)	1.29 (0.07)	1.29 (0.07)
Low interest	1.42 (0.14)	1.26 (0.08)	0.79 (0.05)
High interest	1.96* (0.14)	1.93* (0.08)	1.84* (0.07)
Low self-efficacy	1.53 (0.14)	1.33 (0.08)	0.94 (0.05)
High self-efficacy	1.88 (0.14)	1.92* (0.09)	1.73* (0.07)
Low achievement			1.32 (0.08)
High achievement			1.43* (0.08)

*$p < 0.01$.

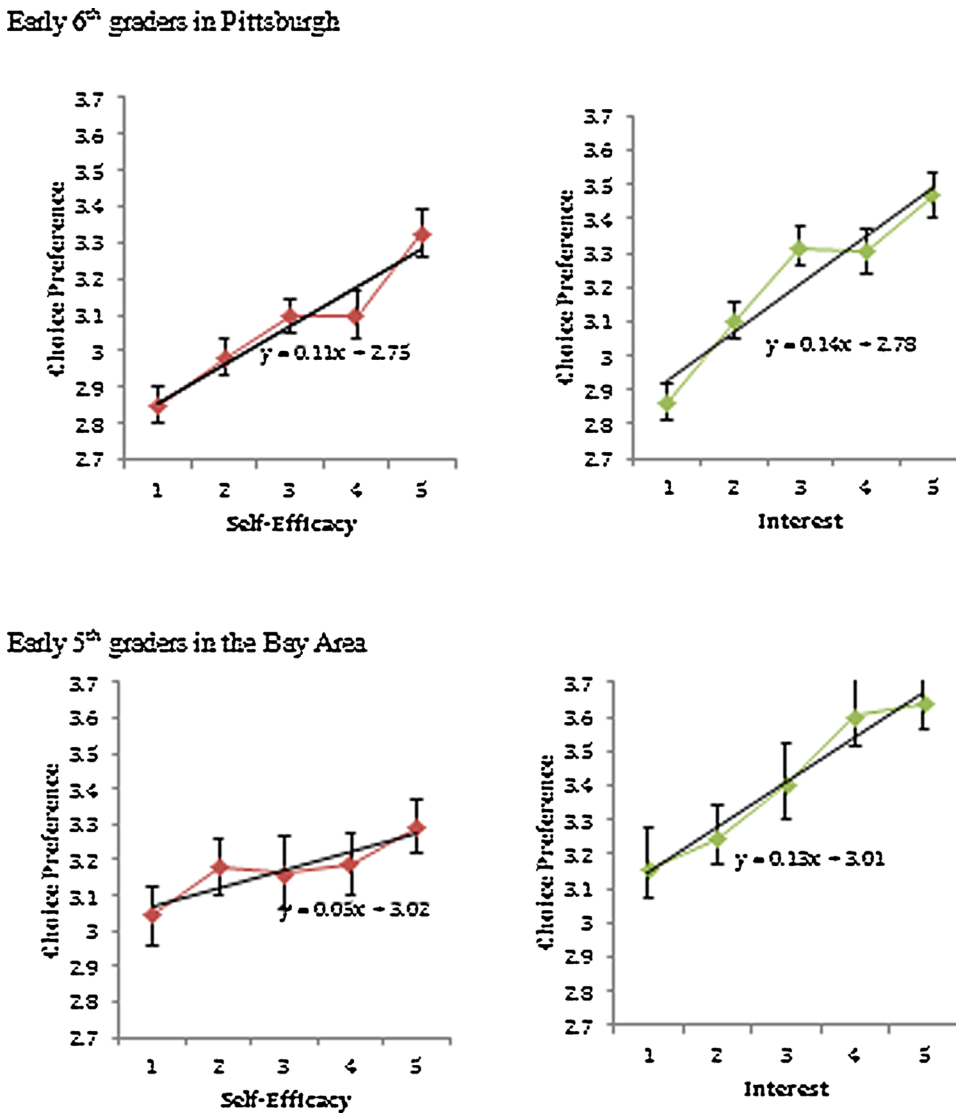

Figure 2. Mean choice preference (SE bars) as a function of binned science self-efficacy and interest across ages and locations.

motivation. Although research on motivations fostering choice is prolific, the prior conceptualization and measurement of science choices had a number of important challenges, with respect to pragmatics of efficient data collection, measuring the child rather than the context, taking into account competition among free choice time and resources, and broadly measuring diverse aspects of choice.

The new science choice preference scale is more inclusive than those in the existing research on choices in terms of construct spectrum. Almost all existing studies we have reviewed measured students' science-related choices simply by a single item, and restricted choices within two options (e.g., choosing science courses or not). Our scale expands the existing choice measurement from the following three aspects. The first aspect is quantitative. The current scale contains five survey items rather than single item. In terms of psychometric quality of survey instruments, a construct is better measured by a number of different items that converge on the theoretical meaning of the construct, as according to psychometric theory, no single item is a pure measure of the construct of interest (Braithwaite & Scott, 1991). Indeed our IRT analyses reveal problems in relying on any one of the particular items across important subgroups of students.

The second aspect is that this quantitative expansion enables us to investigate the extent to which children tend to make a science-related choice from a wide range of alternatives rather than merely two options adopted in past research, particularly when the alternatives not only include apparently non science-related options like art, music, history, etc., but subjects that look like science in some people's conceptions such as math, engineering, or medicine. The real world in which children live is diverse and complex, so the conditions under which they need to make a choice are also diverse and complex. The strategy of measuring science choice preference adopted here reflects more of that diversity and complexity, and thus provides a scale with greater external validity.

Third, the new scale captures diversity of science choices available to students along three important dimensions to further improve scale external validity: choice content (activity, course, or career), time (immediate, proximal, or distant future choice), and settings (in school versus out of school). The multi-faceted nature of science learning opportunities requires a multi-faceted measurement tool to enable researchers and teachers to more accurately grasp the nature of children's choice pattern as a critical expression of their motivation to learn science.

Implications for Science Education

In validating this choice preference scale, we have learned that students' choices of mathematics classes and activities, and engineer or doctor careers are not a part children's preferences toward science at this age, even though they belong to an integrated notion of STEM. This empirical finding not only indicates the scope of valid science choice preference measurement, but also a feasible approach to looking into the boundary of early adolescents' conception of science, which in turn can help us study the development of their science identity. Science identity plays a considerable role in predicting students' actual choices of science-related college majors or careers (Hazari et al., 2010). Repeating such IRT analyses with data collected across science experiences (e.g., ones showing connections of engineering, mathematics, or medicine to science) could be used to track changing choice-relevant conceptions of science.

The strong difference in choice preference patterns between various forms of science and mathematics, engineering, and medicine in the minds of late elementary/early middle school student may seem puzzling. For example, STEM is tightly integrated at the adult level where science uses mathematics quite heavily, and engineering and medicine uses science quite heavily. Further, there is considerable emphasis on STEM as a whole in the policy sphere and STEM integration as new pedagogical/curricular approach. But the data reported here clearly indicated that choice preferences in children reveal they treat engineering, mathematics, and medicine as different. Theories and research of decision making (Williams, 1998; Reyna et al., 2005) suggest that choice making is rooted in individuals' existing knowledge and perception of the choices themselves. Moreover, children's perceptions of science are formulated in a great variety of science-related experiences in school, at home, during extracurricular activities, and so on (Mantzicopoulos et al., 2009). It may be that children with greater exposure to science in applications will see clearer STEM connections and make choices that reflect those connections.

The IRT analyses also showed that two items, science activity in school and science activity at home, have the highest relevance to one's science choice preference (i.e., high discrimination), the relevance of the scientist item is middling, and the items of science class and science museum visits have the lowest relevance to science choices. This pattern was consistent across ages and locations. This difference in discrimination values may reflect the reality that children of this age (10–12 years old) face in which they may not have much freedom in choosing intended courses in school, and choosing to visit science museums. In other words, for them, choosing a course or visiting museums may be construed as heavily controlled by parents or other adults. By contrast, choosing to be scientist as future career is as a distant outcome that may reflect a child's true attitude toward science or true level of science identity. These patterns highlight the importance of culturally/situationally relevant choices in measurement. They also highlight the challenges for educators, parents, and communities in providing equal access and some child autonomy in diverse optional science learning opportunities.

The scientist item has the highest difficulty across ages and locations (i.e., required the highest levels of science choice preferences in order to endorse this item; see Table 4). Overall, commitment to a science career is only common among individuals with very high levels of science choice preference. Pragmatically, this points to the importance of measuring more than career choice because the choices of science career discriminate only students at the high end of choice preferences; thus it is unfortunate that much of the literature has focused on this particular choice alone. Multidimensional choice instruments including "less difficult" items are psychometrically needed like the one developed in the present study.

In a broader pragmatic sense, brief optional learning activities in science are only powerful learning experiences when the choices are at least semi-regularly made, and thus measuring choice preferences at the medium to high end are the most practically important. That is, there may be no difference in self-efficacy, interest, or learning outcomes between one child who would *never* consider an optional science learning experience and another child who would *almost never* consider an optional science learning experience. But, for research purposes, it may be useful to study effects along the whole range, and thus future measures should explore adding items with even lower difficulties levels, perhaps through more non-forced alternative items. This is a limitation of the present study.

In general, research within various theories of motivation revealed that both competence beliefs (expectances for success, self-efficacy) and interest are two good predictors of students' choices of activities/tasks, courses, or careers (Schunk et al., 2008). But, it was still open whether the two are equally important. In examining the concurrent validity of the choice preference scale in the current study (see Figure 2), science interest seemed to play a more significant role than science self-efficacy in predicting their science choice preferences across ages. Moreover, it seems that the younger children are the weaker association between self-efficacy and choice preferences. It may be that these relatively low stakes free choice situations are less psychologically risky for failure, and thus more influenced by interest. In addition, the concurrent validity of the CP instrument is also manifested in its positive relation with the students' science learning achievement (see Table 6). The finding that the boys as a whole reported higher CP score than did girls (see Table 6) is in accordance with the pattern of gender effect on science learning that has been discovered by numerous studies in the literature (Britner, 2008; Patrick et al., 2009; Simpkins et al., 2006), suggesting a strong concurrent validity of the CP scale as well.

The IRT analyses also revealed the limitations of the past single-choice or unidimensional choice instruments (Archer et al., 2012; DeWitt et al., 2014; Simpkins et al., 2006). The present study illustrated the complexity and diversity of children's psychological processes in choice making for science learning; any one particular type of choice is not broadly indicative of student choice preferences across the subgroups that science educators and science education researchers frequently examine. The complexity and diversity stem from a variety of sources such as individual adolescents' inclusive conception of science, settings whereby choices may occur, or time-oriented choices (immediate, proximal, distant future choices). Either single-item or unidimensional choice scales cannot capture the complexity and diversity of children's science choice preferences. For researchers seeking to understand the ways in which motivation and situation factors explain differences in choice behaviors, it is critical to have robust measures that are valid across situations and motivation levels. Similarly, teachers need instruments that work in the wide variety of contexts in which they teach.

With this new developed CP scale, researchers can now examine more comprehensively the ways in which motivation influences long term learning outcomes (e.g., through enabling more optional learning experiences). Researchers might also use the CP scale for studying the effects of such choice preference tendencies on later motivation levels (e.g., are students who are high on science CP more likely to grow in self-efficacy or develop deeper interests in science than those who are low on science CP?). While actual choices are likely important factors, they are much harder to study in quantitative ways given the large number of possible optional science learning experiences to document, the ways in which the list of possibilities will vary great by demographic and geography, and the likely poor memory that students with have for many of them in retrospect.

For science teachers, the current science CP scale can function as reliable and convenient formative assessment of motivational interventions. For example, did a visit to the science museum or an extended science project result in effective motivational changes? For teachers and informal science learning providers, the CP scale could provide information about current student appetite for additional science learning experiences. For example, the scale provides information on which groups of students are most interested in additional opportunities.

Taken together, the present study goes beyond developing and validating a science choice preference scale, in that it provides a conceptualization of the multidimensional nature of student science choices, as well as providing additional insights into the relations between motivational factors and choices in science learning.

Work on this project was supported by grants #2820 and #3341 from the Gordon and Betty Moore Foundation and by grant DRL-1348468 from the National Science Foundation.

References

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2012). Science aspirations, capital, and family habitus: How families shape children's engagement and identification with science. American Educational Research Journal, 49(5), 881–908.

Armor, D. J. (1974). Theta reliability and factor scaling. In H. Costner (Ed.), Sociological methodology (pp. 17–50). San Francisco: Jossey-Bass.

Baker, F. B. & Kim, S.-H. (2004). Item response theory: Parameter estimation techniques (2nd edition). New York: Marcel Dekker.

Bandura, A. (1986). Social foundations of thought and action: A social cognition theory. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1997). Self-efficacy: The exercise of control. New York: Freeman.

Bandura, A. (2001). Social-cognitive theory: An agentic perspective. Annual Review of Psychology, 52, 1–26.

Bereby-Meyer, Y., Assor, A., & Katz, I. (2004). Children's choice strategies: The effects of age and task demands. Cognitive Development, 19(1), 127–146.

Beghetto, R. (2007). Factors associated with middle and secondary students' perceived science competence. Journal of Research in Science Teaching, 44(6), 800–814.

Bevan, B., Michalchik, V., Bhanot, R., Rauch, N., Remold, J., Semper, R., & Shields, P. (2010). Out-of-school time STEM: Building experience, building bridges. San Francisco: Exploratorium, Retrieved April 29, 2013 from http://informalscience.org/research_documents/0000/0629/STEM_OST_Conf_Report.pdf .

Bøe, M. V. (2012). Science choices in Norwegian upper secondary school: What matters? Science Education, 96(1), 1–20.

Bong, M. (2001). Role of self-efficacy and task-value in predicting college students' course performance and future enrollment intentions. Contemporary Educational Psychology, 26(4), 553–570.

Braithwaite, V. A. & Scott, W. W. (1991). Value. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), Measures of personality and social psychological attitudes (pp. 662–690). San Diego, CA: Academic Press.

Bryan, R. R., Glynn, S. M., & Kittleson, J. M. (2011). Motivation, achievement, and advanced placement intent of high school students learning science. Science Education, 95(6), 1049–1065.

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.

Britner, S. L. (2008). Motivation in high school science students: a comparison of gender differences in life, physical, and earth science classes. Journal of Research in Science Teaching, 45(8), 955–970.

Brown, T. A. (2006). Confirmatory factor analysis for applied research. New York, NY: The Guilford Press.

Buldu, M. (2006). Young children's perceptions of scientists: A preliminary study. Educational Research, 48(1), 121–132.

Cai, L., Maydeu-Olivares, A., Coffman, D. L., & Thissen, D. (2006). Limited-information goodness-of-fit testing of item response theory models for sparse 2p tables. British Journal of Mathematical and Statistical Psychology, 59, 173–194.

Cai, L. & Hansen, M. (2013). Limited-information goodness-of-fit testing of hierarchical item factor models. British Journal of Mathematical and Statistical Psychology, 66, 245–276.

Cavallo, A. M. L. & Laubach, T. A. (2001). Students' science perceptions and enrollment decisions in differing learning cycle classrooms. Journal of Research in Science Teaching, 38(9), 1029–1062.

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.

Cerinsek, G., Hribar, T., Glodez, N., & Dolinsek, S. (2012). Which are my future career priorities and what influenced my choice of studying science, technology, engineering or mathematics? Some insights on educational choice, Ä̀Case of Slovenia. International Journal of Science Education, 35(17), 2999–3025.

DeWitt, J., Archer, L., & Osborne, J. (2014). Science-related aspirations across the primary-secondary divide: Evidence from two surveys in England. International Journal of Science Education, 36(10), 1609–1629.

Dierking, L. D. & Falk, J. H. (2003). Optimizing out-of-school time: The role of free choice learning. New Directions for Youth Development, 2003(97), 75–88.

Durik, A. M., Vida, M., & Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. Journal of Educational Psychology, 98(2), 382–393.

Eccles, J. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. Educational Psychologist, 44(2), 78–89.

Embretson, S. E. & Reise, S. P. (2000). Item response theory for psychologists. Mahwah New Jersey: Lawrence Erlbaum Associates.

Feder, M. A., Shouse, A. W., Lewenstein, B., & Bell, P. (Eds.) (2009). Learning science in informal environments: People, places, and pursuits. Washington, DC: National Academies Press.

Feest, U. (2005). Operationism in psychology: What the debate is about, what the debate should be about. Journal of the History of the Behavioral Sciences, 41(2), 131–149.

Flowerday, T. & Schraw, G. (2003). Effect of choice on cognitive and affective engagement. The Journal of Educational Research, 96(4), 207–215.

Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. Journal of Personality and Social Psychology, 78(2), 350–365.

Fredricks, J., Blumenfeld, P., & Alison, P. (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, 74(1), 59–109.

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. Journal of Research in Science Teaching, 48(10), 1159–1176.

Hazari, Z., Sonnert, G., Sadler, P. M., & Shanahan, M.-C. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study. Journal of Research in Science Teaching, 47(8), 978–1003.

Healey, J. (2012). The essentials of statistics: A tool for social research (3rd edition). Wadsworth: Cengage Learning.

Hinkin, T. R. (1995). A review of scale development practices in the study of organization. Journal of Management, 21(5), 967–988.

Hidi, S. & Renninger, K. A. (2006). The four-phase model of interest development. Educational Psychologist, 41(2), 111–127.

Jang, H., Reeve, J., & Deci, E. (2010). Engaging students in learning activities: It is not autonomy support or structure but autonomy support and structure. Journal of Educational Psychology, 102(2), 588–600.

John, O. P., & Soto, C. J. (2007). The importance of being valid: Reliability and the process of construct validation. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), Handbook of research methods in personality psychology (pp. 461–494). New York: The Guilford Press.

Katz, I. & Assor, A. (2007). When choice motivates and when it does not. Educational Psychology Review, 19(4), 429–442.

Kane, M. T. (2001). Current concerns in validity theory. Journal of Educational Measurement, 38(4), 319–342.

Lee, J. & Shute, V. J. (2010). Personal and social-contextual factors in K-12 academic performance: An integrative perspective on student learning. Educational Psychologist, 45(3), 185–202.

Logan, M. & Skamp, K. (2008). Engaging students in science across the primary secondary interface: Listening to the students' voice. Research in Science Education, 38(4), 501–527.

Madden, M., Lenhart, A., Duggan, M., Cortesi, S., & Gasser, U. (2013). Teens and technology 2013. Pew Internet & American Life Project..

Mantzicopoulos, P., Samarapungavan, A., & Patrick, H. (2009). ÄúWe learn how to predict and be a scientist, Äú: Early science experiences and kindergarten children's social meanings about science. Cognition and Instruction, 27(4), 312–369.

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. Journal of Educational Psychology, 82(1), 60–70.

Meece, J. L., Glienke, B. B., & Burg, S. (2006). Gender and motivation. Journal of School Psychology, 44(5), 351–373.

Meece, J., & Painter, J. (2008). Gender, self-regulation, and motivation. In D. H. Schunk & B. J. Zimmerman (Eds.), Motivation and self-regulated learning: Theory, research, and applications (pp. 339–367). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Morizot, J., Ainsworth, A. T., & Reise, S. P. (2007). The construct validation approach to personality scale construction. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), Handbook of research methods in personality psychology (pp. 407–423). New York: The Guilford Press.

Mortensen, M. F. & Smart, K. (2007). Free-choice worksheets increase students' exposure to curriculum during museum visits. Journal of Research in Science Teaching, 44(9), 1389–1414.

Neuville, S., Frenay, M., & Bourgeois, E. (2007). Task value, self-efficacy and goal orientations: Impact on self-regulated learning, choice and performance among university students. Psychologica Belgica, 47(1–2), 95–117.

Parker, P. D., Schoon, I., Tsai, Y-M., Nagy, G., Trautwein, U., & Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: A multicontext study. Developmental Psychology, 48(6), 1629–1642.

Patall, E., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. Psychological Bulletin, 134(2), 270–300.

Patall, E., Cooper, H., & Wynn, S. (2010). The effectiveness and relative importance of choice in the classroom. Journal of Educational Psychology, 102(4), 896–915.

Patall, E. (2012). The motivational complexity of choosing: A review of theory and research. In R. M. Ryan (Ed.), The Oxford handbook of human motivation (pp. 248–279). New York: Oxford University Press.

Patall, E. A. (2013). Constructing motivation through choice, interest, and interestingness. Journal of Educational Psychology, 105(2), 522–534.

Patrick, H., Mantzicopoulos, P., & Samarapungavan, A. (2009). Motivation for learning science in kindergarten: Is there a gender gap and does integrated inquiry and literacy instruction make a difference. Journal of Research in Science Teaching, 46(2), 166–191.

Pomerantz, E. M., Cheung, C. S.-S., & Qin, L. (2012). Relatedness between children and parents: Implications for motivation. In R. M. Ryan(Ed.), The Oxford handbook of human motivation (pp. 579–349). New York, NY: Oxford University Press.

Powers, R. S., Wilson, K., Keels, M. M., & Walton, M. (2013). Who's Connected? Trends from 1999 to 2011 in Home Internet Access in North Carolina. Sociation Today. .

Reber, R., Hetland, H., Weiqin, C., Norman, E., & Kobbeltvedt, T. (2009). Effects of example choice on interest, control, and learning. Journal of the Learning Sciences, 18(4), 509–548.

Rector, R. & Sheffield, R. (2011). Understanding poverty in the United States: Surprising facts about America's poor. The Heritage Foundation Leadership for America. .

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. Psychological Bulletin, 114(3), 552–566.

Reise, S. P. & Waller, N. G. (2002). Item response theory for dichotomous assessment data. In S. P. Reise & N. G. Waller (Eds.), Measuring and analyzing behavior in organizations: Advances in measurement and data analysis (pp. 88–122). San Francisco, CA: Jossey-Bass.

Reyna, V. F. & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. Learning and Individual Differences, 7(1), 1–75.

Reyna, V. F., Adam, M. B., Poirier, K. M., LeCroy, G. W., & Brainerd, C. J. (2005). Risky decision making in childhood and adolescence: A fuzzy-trace theory approach. In J. E. Jacobs & P. A. Klaczynski (Eds.), The development of judgment and decision making in children and adolescents (pp. 77–106). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Robbins, P. & Aydede, M. (2009). The Cambridge handbook of situated cognition. New York, NY: Cambridge University Press.

Ryan, R. M. & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. Contemporary Educational Psychology, 25(1), 54–67.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. Journal of Counseling Psychology, 57(1), 1–10.

Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). Motivation in education: Theory, research, and applications (3rd edition). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

Sheppard, K. & Robbins, D. M. (2005). Chemistry, the central science? The history of the high school science sequence. Journal of Chemistry Education, 82(4), 561–566.

Sjaastad, J. r. (2011). Sources of inspiration: The role of significant persons in young people's choice of science in higher education. International Journal of Science Education, 34(10), 1615–1636.

Simms, L. J. & Watson, D. (2007). The construct validation approach to personality scale construction. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), Handbook of research methods in personality psychology (pp. 240–258). New York: The Guilford Press.

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. Developmental Psychology, 42(1), 70–83.

Smith, L. B. (2005). Cognition as a dynamic system: Principles from embodiment. Developmental Review, 25(3–4), 278–298.

Stake, J. E. & Mares, K. R. (2001). Science enrichment programs for gifted high school girls and boys: Predictors of program impact on science confidence and motivation. Journal of Research in Science Teaching, 38(10), 1065–1088.

Steele, J. P. & Fullagar, C. J. (2009). Facilitators and outcomes of student engagement in a college setting. Journal of Psychology, 143(1), 5–27.

Tai, R. T., Liu, C. Q., Maltese, A. V., & Fan, X. T. (2006). Planning early for careers in science. Science, 312(5777), 1143–1144.

Tschacher, W. & Scheier, C. (1999). The perspective of situated and self-organizing cognition in cognitive psychology. In P. Van Loocke (Ed.), The nature of concepts: Evolution, structure and representation (pp. 68–90). Florence, KY: Taylor & Frances/Routledge.

Vedder-Weiss, D. & Fortus, D. (2011). Adolescents' declining motivation to learn science: Inevitable or not? Journal of Research in Science Teaching, 48(2), 199–216.

Vedder-Weiss, D. & Fortus, D. (2012). Adolescents' declining motivation to learn science: A follow-up study. Journal of Research in Science Teaching, 49(9), 1057–1095.

Vigneau, F. O. & Bors, D. A. (2005). Items in context: Assessing the dimensionality of Raven's advanced progressive matrices. Educational and Psychological Measurement, 65(1), 109–123.

Walker, C. M. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation? Journal of Psychoeducational Assessment, 29(4), 364–376.

Wigfield, A. & Eccles, J. (2000). Expectancy-value theory of achievement motivation. Contemporary Educational Psychology, 25(1), 68–81.

Wigfield, A., Byrnes, J. P., & Eccles, J. S. (2006). Development during early and middle adolescence. In P. A. W. Alexander & H. Philip (Eds.), Handbook of educational psychology (pp. 87–113). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Wigfield, A. & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. Developmental Review, 30(1), 1–35.

Williams, S. (1998). An organizational model of choice: A theoretical analysis differentiating choice, personal control, and self-determination. Genetic, Social, and General Psychology Monographs, 124(4), 465–491.

Zimmerman, H. T. (2012). Participating in science at home: Recognition work and learning in biology. Journal of Research in Science Teaching, 49(5), 597–630.

Zumbo, B. D., Gadermann, A. M., & Zerisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. Journal of Modern Applied Statistical Methods, 6(1), 21–29.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.