Routledge
Taylor & Francis Group

International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

Development, validation, and factorial comparison of the McGill Self-Efficacy of Learners For Inquiry Engagement (McSELFIE) survey in natural science disciplines

Ahmed Ibrahim, Mark W. Aulls & Bruce M. Shore

To cite this article: Ahmed Ibrahim, Mark W. Aulls & Bruce M. Shore (2016): Development, validation, and factorial comparison of the McGill Self-Efficacy of Learners For Inquiry Engagement (McSELFIE) survey in natural science disciplines, International Journal of Science Education, DOI: 10.1080/09500693.2016.1249531

To link to this article: http://dx.doi.org/10.1080/09500693.2016.1249531

Published online: 04 Nov 2016.

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2016
http://dx.doi.org/10.1080/09500693.2016.1249531

Development, validation, and factorial comparison of the McGill Self-Efficacy of Learners For Inquiry Engagement (McSELFIE) survey in natural science disciplines

Ahmed Ibrahim[a,b†], Mark W. Aulls[b] and Bruce M. Shore[b]

[a]University of Chicago, Chicago, IL, USA; [b]Department of Educational and Counselling Psychology, McGill University, Montréal, QC, Canada

ABSTRACT

Sociocognitive theory [Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall; Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, *44*, 1175–1184. doi:10.1037/0003-066x.44.9.1175; Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, *50*, 248–287. doi:10.1016/0749-5978(91)90022-L] accords high importance to the mechanisms of human agency and how they are exercised through self-efficacy. In this paper, we developed and validated the McGill Self-Efficacy For Inquiry Engagement (McSELFIE) instrument with undergraduate students in natural science disciplines. We defined inquiry engagement as carrying out the practices of science (POS) that are supported by students' personality characteristics (SPCs) and that result in achieving inquiry-learning outcomes (ILOs). Based on these theoretical perspectives, the McSELFIE is a 60-item, learner-focused survey that addresses three components that are theoretically important for engaging in scientific inquiry: (a) SPCs, (b) ILOs, and (c) POS. Evidence for construct and content validity were obtained by using experts' judgments and confirmatory factor analysis with a sample of 110 undergraduate students enrolled in science disciplines. Internal consistency of the factors and instrument was also examined. The McSELFIE instrument is a reliable and valid instrument for measuring science undergraduate students' self-efficacy for inquiry engagement. Matched pairs analyses were conducted among the instruments' factors. Students reported the highest self-efficacy for openness, applying knowledge, and carrying out investigations. Students reported the lowest self-efficacy for extraversion, understanding metacognitive knowledge, and planning investigations. Theoretical and practical implications are discussed.

ARTICLE HISTORY
Received 15 January 2016
Accepted 13 October 2016

KEYWORDS
NGSS; practices of science; self-efficacy; scientific inquiry; science education

Social cognitive theory (Bandura, 1986, 1989, 1991) accords high importance to the mechanisms of human agency and how they are exercised through self-efficacy. Self-efficacy has been defined as the 'beliefs in one's capabilities to organize and execute the course of

CONTACT Ahmed Ibrahim aiibrahim@uchicago.edu
†Current address: University of Chicago, Chicago, IL, USA

action required to produce given attainments' (Bandura, 1997, p. 3). Self-efficacy was previously defined as being 'concerned with judgments of how well one can execute courses of action required to deal with prospective situations' (Bandura, 1982, p. 122), 'people's convictions in their own effectiveness [that] is likely to affect whether they will even try to cope with given situations' (Bandura, 1977, p. 193), and confidence about successfully performing in a specific domain (Bandura, 2006). According to social cognitive theory (Bandura, 1986, 1997), self-efficacy is affected by (a) actual performances, (b) vicarious performances, (c) forms of social persuasion, and (d) physiological indices. Several meta-analytic and systematic-review studies have shown that self-efficacy is strongly and positively related to academic outcomes and achievement (Bartimote-Aufflick, Bridgeman, Walker, Sharma, & Smith, 2015; Richardson, Abraham, & Bond, 2012). Self-efficacy is also a significant predictor of career options (Lent, Brown, & Larkin, 1984), learning strategies (Bong, 2006), academic performance (Honicke & Broadbent, 2016), and work-related performance (Stajkovic & Luthans, 1998). Self-efficacy affects students' choices of activities, effort, persistence, interest, and self-regulation (Bandura, 1997; Bandura, Barbaranelli, Caprara, & Pastorelli, 1996; Pajares, 1996; Pajares & Schunk, 2001; Schunk & Pajares, 2009). In undergraduate science education, the same strong effects of self-efficacy on learning outcomes and achievement persist. For example, Jansen, Scherer, and Schroeders (2015) found that academic self-efficacy had a stronger predictive impact on achievement in general science than academic self-concept, and Cavallo, Potter, and Rozman (2004) showed that self-efficacy significantly predicted physics understanding and achievement in college physics students. Self-efficacy in scientific disciplines is important because it could account for scientific-related career choices, which is a national priority in many countries including the US (Lamb, Vallett, & Annetta, 2014). The need for assessing self-efficacy in scientific disciplines leads to the necessity for creating instruments for that purpose.

Existing instruments of self-efficacy

Self-efficacy for science and self-efficacy for engineering

Several instruments have already been developed to measure students' self-efficacy for science, inquiry, engineering, or research. For example, the Sources Of Self-Efficacy in Science Courses (SOSEC) scale was developed to measure the roots of self-efficacy in typical introductory science-class experiences (Fencl & Scheel, 2004). The SOSEC instrument was primarily developed to measure the four sources of self-efficacy: (a) actual performance, (b) vicarious learning, (c) emotional arousal, and (d) social persuasion, and specifically targeted students in physics courses. Although SOSEC addressed self-efficacy, it only addressed the sources (predictors) of self-efficacy.

Ketelhut (2010) developed the Self-Efficacy in Technology and Science (SETS) scale to measure middle-school students' self-efficacy in technology and science. Lamb et al. (2014) developed a short version of SETS to help identify STEM students who need interventions based on the assessment of self-efficacy. The SETS instrument included only one holistic factor for self-efficacy in inquiry-science processes, and five factors related to self-efficacy in technology (videogaming, computer gaming, general computer use, problem-solving computer use, and synchronous chat use). The SETS instrument addressed mainly self-efficacy for technology, with the exception of one factor (consisting of 12 items) that addressed self-efficacy for inquiry. The 12 items in the SETS inquiry factor were based on another instrument that was developed for college biology students (Baldwin, Ebert-May, & Burns, 1999). Consequently, although SETS addressed self-efficacy and inquiry, it was limited in its scope and applicability. Instruments for self-efficacy in engineering design have also been developed (Seth, Ibrahim, & Tangora, 2015; Carberry, Lee, & Ohland, 2010), but as the name suggests, focused on engineering design.

Self-efficacy for research

Several instruments were also created to measure self-efficacy for research such as the Research Self-Efficacy Scale (RSES) (Bieschke, Bishop, & Garcia, 1996), and the Research Self-Efficacy (RSE) scale (Holden, Barker, Meenaghan, & Rosenberg, 1999). Healey, Jordan, Pell, and Short (2010) stated that it is generally accepted that students' experience of research could be one of the following: (a) research-led, (b) research-oriented, (c) research-based, and (d) research-tutored (Griffiths, 2004; Healey, 2005a, 2005b). According to this perspective, students experience inquiry-based activities only in formal research-based experiences in disciplines that commonly use the term research to describe inquiry. Consequently, students' self-efficacy for research cannot be equated with students' self-efficacy for inquiry because inquiry can describe settings and contexts that are not limited to research-based activities, such as problem-based learning or case-based learning.

Self-efficacy for inquiry

Self-efficacy has also been used to evaluate students' confidence in scientific inquiry (Ketelhut, 2007, 2010), although limited by using older definitions of inquiry (National Research Council, 1996, 2000) or equating scientific inquiry self-efficacy with science self-efficacy (Chen, Metcalf, & Tutwiler, 2014; Nelson & Ketelhut, 2008). These instruments did not address the idea of science as a set of practices, and thus may not be appropriate for use with current implementations of curricula that take such a view of science.

Qualities needed in a new measure of self-efficacy for science

The most recent standards of science, *Next Generation Science Standards* (NGSS; NGSS Lead States, 2013), advocated the perspective of science as a set of practices. A new instrument that incorporates this perspective is needed in the literature. Such an instrument, that takes the NGSS perspective into consideration, can provide educators and researchers with a tool to assess students' self-efficacy for the practices of science (POS). In addition to measuring students' self-efficacy for carrying out the POS, it would be useful to assess students' self-efficacy for achieving learning outcomes such as understanding content knowledge, applying content knowledge, and evaluating and creating knowledge. One other useful aspect to assess in the context of scientific inquiry in higher eduaction is students' self-efficacy for exhibiting their personality characteristics.

These three requirements of measurement, namely (a) self-efficacy for POS, (b) self-efficacy for achieving inquiry-learning outcomes (ILOs), and (c) self-effiacacy for demonstrating student personality characteristics (SPCs), led to the conceptualization of the new self-efficacy instrument that we present in this paper, the McGill Self-Efficacy of Learners For Inquiry Engagement (McSELFIE) instrument. Although some self-efficacy instruments are composite in nature (i.e. by addressing more than one self-efficacy dimension) such as the SETS (Ketelhut, 2010) instrument that included self-efficacy for technology and self-efficay for inquiry in a single instrument, none of the existing self-efficacy instruments addressed SPCs, POS, and ILOs in a single instrument. The development of a new instrument of self-efficacy that incorporates these three dimensions could provide new insights related to predicting students' confidence in achieving learning outcomes based on their confidence in carrying out the POS.

Practices of science

The current movement in standards, policies, and implementations of science teaching and learning has adopted the notion of science-as-practice (Erduran & Dagher, 2014). One of the main assumptions underlying this notion is that engaging in POS as well as understanding disciplinary core ideas of science promote a deeper understanding of scientific investigations and help students become more like experts and less like novices in disciplinary fields of knowledge. The NRC (2012) emphasized that this approach is considered a major improvement over previous views of describing the meaning of scientific inquiry. The idea of practices emerged from the ethnographic study of laboratories and research groups, and from the work in the history and philosophy of science, psychology, and sociology (National Research Council; NRC, 2012).

Practices are currently advocated as what should 'imbue articulations like "scientific method" and "inquiry" with authentic meaning' (Ford, 2015, p. 1047). The National Research Council (NRC, 2012) stated that, 'it is only through engagement in the practices that students can recognize how [scientific] knowledge comes about and why some parts of scientific theory are more firmly established than others' (p. 44). The older definitions of inquiry relied on conceptualizing inquiry as a set of skills or a succession of phases, which led to misconceptions about how scientific inquiry actually works (NRC, 2012). The idea of practices entails the coordination of knowledge of content, and skills of implementing tasks (NRC, 2012).

To elucidate the idea, in 2013, the NRC provided a set of specific practices of science and engineering (PSE) that offered a succinct and clear idea of what scientific inquiry should look like. Scientific inquiry as a set of practices should not be segregated from learning scientific content and disciplinary core ideas, and should involve students in understanding crosscutting concepts such as cause and effect and searching for patterns. Such a view would make students more expert-like, make them have a better appreciation for scientific knowledge, and allow them to experience better learning outcomes in the scientific or engineering disciplines. The practices proposed by the NRC (2012) were (a) asking questions (for science) and defining problems (for engineering), (b) developing and using models, (c) planning and carrying out investigations, (d) analyzing and interpreting data, (e) using mathematics and computational thinking, (f) constructing explanations (for science) and designing solutions (for engineering), (g) engaging in argument from evidence, and (h) obtaining, evaluating, and communicating information. In the current study, we focused on the POS by excluding the practice of designing solutions that is primarily addressed to engineers. We used the POS in our current study to investigate how confident undergraduate students in scientific disciplines would be about engaging in them within the context of scientific inquiry.

Students' personality characteristics

The Five-Factor Model (FFM; Digman, 1990; Goldberg, 1993; John, Naumann, & Soto, 2008; McCrae & John, 1992) presented five broad dimensions of personality characteristics that have attracted broad consensus that they can classify and account for the variation of several other traits (Morizot, 2014). These Big Five characteristics are (a) openness, (b) conscientiousness, (c) extraversion, (d) agreeableness, and (e) neuroticism. Openness includes attributes such as using imagination (McCrae, 1993), cleverness in solving problem (Buss, 1991, 1996), innovation, and unconventionality (Ashton & Lee, 2001, 2007). Conscientiousness includes attributes such as being organized and orderly (Ashton & Lee, 2001, 2007), and adhering to plans and timelines (McCrae & Costa, 1996, 1997). Extraversion includes attributes such as expressiveness and sociability (Ashton & Lee, 2001). Agreeableness refers to being gentle and sympathetic versus being harsh and cold-hearted; and neuroticism refers to emotional instability (Ashton & Lee, 2001). These characteristics can be conceptualized and assessed as motivational constructs, and this provides the advantage of recognizing that they can interact with environmental features to result in behavior and action (Denissen & Penke, 2008).

From this perspective, SPCs should play an important role in engaging in inquiry because engaging in scientific inquiry entails being open to ideas and experiences, being collaborative with colleagues and teachers, and being organized and disciplined (Aulls & Shore, 2008). In the current study, we included only openness, conscientiousness, and extraversion as SPCs. Neuroticism was excluded because it has been negatively related to learning styles (Komarraju, Karau, Schmeck, & Avdic, 2011). There were no theoretical or conceptual reasons for including agreeableness in the current instrument, and thus it was not included in the instrument development.

Inquiry-learning outcomes

Engaging in PSE without understanding the importance of attaining specific learning outcomes represents an incomplete representation of the nature of engagement. Learning outcomes represent the change that happens to a person as a result of a learning experience (Watson, 2002), or what a student is expected to be able to do as a result of a learning experience (Gogus, 2012). In the current paper, the learning experience was defined through engagement in POS in the context of inquiry. Outcomes of engagement in scientific inquiry are broad and diverse (Saunders-Stewart, Gyles, Shore, & Bracewell, 2015). For example, from a literature review Saunders-Stewart, Gyles, and Shore (2012) derived a 23-item, criterion-referenced inventory for student outcomes of engaging in inquiry. Examples of these ILOs included knowledge acquisition, motivation, understanding, metacognition, application of knowledge, and development of expertise. These learning outcomes present similarities to learning outcomes in the cognitive domain identified in the revised Bloom's taxonomy (Krathwohl, 2002). Several studies used similar learning outcomes in the context of inquiry (e.g. Redden, Simon, & Aulls, 2007; Bunterm et al., 2014; Spronken-Smith, Walker, Batchelor, O'Steen, & Angelo, 2012) to assess students' levels of learning outcomes. In our current study, we included items that measured students' self-efficacy for five cognitive ILOs that are consistent with the revised Bloom's taxonomy (Krathwohl, 2002). We were interested in assessing how students feel confident in achieving learning outcomes that range from understanding to applying, evaluating, and creating. Accordingly, we included five ILOs: (a) understanding conceptual knowledge, (b) understanding metacognitive knowledge, (c) application, (d) evaluation, and (e) creation.

Understanding conceptual knowledge was represented by items that referred to understanding important concepts and forming a coherent mental structure of knowledge by linking previous and current knowledge and representing this knowledge in the form of a knowledge cluster. Understanding metacognitive knowledge was represented by items that referred to understanding the effects of personal views on learning, understanding the effects of engaging in inquiry learning, and externalizing metacognitive awareness by addressing or talking about any internal doubts about learning. Applying knowledge was represented by items that referred to using previous knowledge in new situations or using new knowledge in new contexts or outside the boundaries of a classroom. Evaluating was represented by items that referred to making judgments about the learning experience, which is based on reflection and taking into consideration other people's opinions. Creating was represented by items that referred to creating new questions, creating new knowledge, and working in a creative environment.

Objectives

In this paper, we defined and operationalized inquiry engagement in the context of science education as the engagement in the POS, which is supported by SPCs and leads to ILOs. POS were based on the NGSS standards and articulated PSE, by focusing on science and excluding engineering design. Our specific goals were to:

1. Develop a learner-focused survey instrument that measures the self-efficacy for inquiry engagement (as defined by carrying out the POS, in addition to SPCs and ILOs) for undergraduate students who are enrolled in scientific disciplines.
2. Validate this instrument by examining its construct and content validity.
3. Compare the students' ratings of self-efficacy for the different components of inquiry engagement. The comparison of the students' ratings sought to answer the following sub-questions: (a) Which SPCs do undergraduate students enrolled in natural science disciplines feel more and less self-efficacious to exhibit in inquiry engagement? (b) Which ILOs do undergraduate students enrolled in natural science disciplines feel more and less self-efficacious to achieve in inquiry engagement? and (c) Which POS do undergraduate students enrolled in natural science disciplines feel more and less self-efficacious to carry out in inquiry engagement?

Method

Developing the items of the McSELFIE instrument

In order to develop the current instrument, we searched the literature for items that could specifically address the PSE in addition to SPCs and ILOs. We relied on the McGill

Strategic Demands of Inquiry Questionnaire (MSDIQ) (Shore, Chichekian, Syer, Aulls, & Frederiksen, 2012) to extract items that were classified into the PSE, SPCs, and ILOs. Analysis of the MSDIQ resulted in a 67-item, criterion-referenced, learner-focused questionnaire that addressed the specific 'strategic demands' of engaging in inquiry. These 67 items were used to develop the MAVIES instrument (Ibrahim, Aulls, & Shore, 2016). The MAVIES instrument consisted of four components that dealt with the attainment value (or importance) assigned by students to teachers' roles (7 items), SPCs (13 items), PSE (32 items), and ILOs (15 items).

The items included in the MAVIES represented a broad repertoire of different inquiry activities, SPCs, and ILOs that represented the dimensions we sought to include in our instrument. Thus, we found it possible to convert the items in the MAVIES instrument to self-efficacy items to be used in McSELFIE by (a) excluding the seven items that asked about the teachers' roles, resulting in the 60-item core of the McSELFIE instrument), (b) preceding each of the items by a stem that asked the students about their confidence to do a certain POS-related task, and (c) slightly rephrasing some of the items when necessary to be more specific about what the students could express confidence in doing. The revised items' stem was 'I believe that I am able to.' The scale ranged from 0, corresponding to 'definitely cannot,' to 10, corresponding to 'definitely can.' The new instrument sought to ask the learners to express judgments of their confidence levels about inquiry-related tasks and activities. The McSELFIE items were reviewed by an expert panel that confirmed the adequacy of using the items as a basis for a survey to assess the self-efficacy for inquiry engagement for undergraduate students enrolled in scientific disciplines. After confirming the relevance of the remaining 60 items, the survey was administered to a representative sample of undergraduate students enrolled in scientific disciplines.

Sampling strategy and survey administration

We used a 'typical case' sampling approach (Kuzel, 1999; Patton, 1990) to ensure a normal or average population and increase the 'confidence in conclusions' (Miles & Huberman, 1994, p. 28). Our purpose was to represent the range of scientific disciplines by including a large sample of undergraduates who studied in a scientific disciplinary program. Traditional scientific disciplines refer to traditional natural sciences including physics, chemistry, biology, and recently earth, space, and environmental sciences (NRC, 2012). These disciplines differ in the phenomena studied, the basic concepts and vocabulary, and the theories that explain and predict phenomena within each discipline (Toulmin, 1972). However, they share a common empiricist epistemology, assumptions about observation and experimentation, experimental methods for data collection, the credibility of scientific explanations (Repko & Szostak, 2017; Rosenberg, 2000), and an organized intellectual social community of practice (Becher, 1994; Whitley, 1984).

Although there are several educational policies that call for an integrated STEM education that would prepare students in an integrative perspective in science, technology, engineering, and mathematics (Committee on Integrated STEM education, Honey, Pearson, Schweingruber, 2014; NRC, 2012), this paper focused on students who have indicated a disciplinary career in the natural sciences. We recognized that different disciplines within STEM would normally deal with different content knowledge, and that students in disciplines such as mathematics and the sciences normally have different ways of knowing (Lederman & Lederman, 2013). Accordingly, we excluded from the original sample students who had indicated that they were enrolled in engineering or mathematics disciplines despite the fact that they may have background in the natural sciences, or had indicated that they elected a minor in a scientific discipline. Details of the sample composition are presented in the Results section. After constructing the instrument and specifying the sampling strategy, the instrument was administered to the students. Data were collected using surveys that were administered in the university classrooms.

Establishing construct validity using confirmatory factor analysis

Construct validity refers to the 'degree of agreement with theoretical expectations' (Knapp & Mueller, 2010, p. 340). Our reliance on theoretical foundations to develop the instrument from its earliest conception offered support for the validity of the overall construct of inquiry engagement and its constituent constructs of (a) SPCs, (b) ILOs, and (c) POS, which are based on their respective theories or conceptual frameworks.

Model specification using substantive justification

In the current instrument, the factorial constructs (Openness, Conscientiousness, and Extraversion) were well supported by the FFM of personality. As discussed in the section on the qualities needed in a new measure of self-efficacy for science, the FFM additionally includes agreeableness and neuroticism as part of the model; however, we did not include these two constructs in our instrument because the link between these constructs of being gentle and sympathetic or emotionally unstable and engagement in PSE is not established in the literature. Our purpose was not to test the FFM in scientific inquiry, but to test and confirm a subset of the FFM as part of our instrument.

Similarly, the ILOs included conceptual understanding, metacognitive understanding, applying knowledge, evaluation, and creation. These ILOs are not all the learning outcomes identified in the modified Bloom's taxonomy (Krathwohl, 2002). However, the ILOs identified in the current instrument are well supported theoretically, conceptually, and empirically. The current instrument does not address all the PSE identified by the NRC, and this is discussed in the Limitations of the current study. The included POS are recommended by the NRC and form the basis of the NGSS dimension of practices.

We used confirmatory factor analysis (CFA) to evaluate the psychometric properties of our hypothesized models depicting the item–factor relations in McSELFIE. CFA can be used to evaluate an instrument's construct validity and requires theoretical or conceptual frameworks to specify the item-factor relations (Brown, 2015). We used the FFM of personality, Bloom's revised taxonomy of educational objectives in the cognitive domain, and the NRC (2012) PSE as theoretical reference frameworks.

Selecting the CFA fitting function

Robust Maximum Likelihood (MLR) (Yuan & Bentler, 2000) provides parameter estimates with standard errors (SEs) and a mean adjusted χ^2 test statistic that are robust to non-normality (Brown, 2015; Lei & Wu, 2012), and was used in the CFA model estimation. Maximum Likelihood (ML) cannot be used in the case of non-normal data, because ML requires multivariate normality.

Evaluating model-fitting results

We used the data-model-fit indices recommended by West, Taylor, and Wu (2012) and their associated recommended cut-off criteria (Brown, 2006; Geiser, 2012; Hu & Bentler, 1998; Schermelleh-Engel, Moosbrugger, & Müller, 2003; West et al., 2012) to assess data-model-fit. These indices are (a) Standardized Root Mean Square Residual (SRMR), (b) Root Mean Square Error of Approximation (RMSEA), (c) Tucker Lewis Index (TLI), and (d) Comparative Fit Index (CFI). Goodness-of-data-model-fit should be further investigated by checking the modification indices (MIs) and the Expected Parameter Changes (EPCs), and the absence of out of range parameter estimates (Heywood cases).

Evaluating model results

A cut-off value of 0.3 was chosen to determine the salient loadings of items on factors. A cut-off value should ideally preserve the items with high loadings (the loadings that have a value higher than the cut-off value), and allow for interpreting these items together as part of a cluster of items (that have a loading value that is greater than that of the cut-off value) and load on the same factor. In factor analytic studies, one of the purposes is to determine the relations among measured items (indicators), and latent constructs (factors). In applied research, a cut-off value of .30 or .40 is often used to determine a 'salient; factor loading, which means that the item or indicator is 'meaningfully' related to the primary factor upon which it loads (Brown, 2015). However, the use of a specific cut-off may lead to eliminating an item (or more than one item) altogether that has a loading value below the cut-off value. Such elimination would lead, in turn, to retaining a smaller number of items per factor compared to the number of items that could be retained if a less conservative cut-off value were chosen. The advantages of this strategy are that it should normally lead to obtaining items that are more coherent per factor, and also lead to a better interpretation of the factor. However, the disadvantage of using a higher and more conservative cut-off value is the risk of eliminating items that are vital to the construction of the instruments, and that are 'presumably chosen for some reason and that eliminating some of them changes the definition of one's construct(s) to some extent' (Bandalos & Finney, 2010, p. 100). In the current analysis, the items with the highest item-factor loading (IFL) values above the cut-off value were retained, and the structure of the components of the instrument was interpreted based on the retained item-factor relations that resulted after applying the cut-off value.

Establishing content validity using expert judgment

Content validity is currently considered to be the core and essential form of validity for tests and instruments (Gorin, 2007; Kane, 2008). Content validity refers to 'expert judgments of the representativeness of items with respects to skills, knowledge, etc. domain to be covered' (Knapp & Mueller, 2010, p. 340).

We sought to establish that the items in McSELFIE represented their respective latent factors. We relied on three experts in higher education, science education, and scientific inquiry to evaluate the adequacy of the items in their representation of the factorial structure of the three subsets of the McSELFIE instrument. The experts agreed 100% that the items represented corresponding factors.

Establishing reliability

The internal consistency of an instrument is defined as 'how well the items that make up an instrument or one of its subscales fit together' (Pett, Lackey, & Sullivan, 2003, p. 175). Bandalos and Finney (2010) stated that 'internal consistency estimates for multidimensional instruments should be obtained for the dimensional level at which the scale will be interpreted and used' (p. 105). Accordingly, we computed Cronbach's alpha for all the McSELFIE factors as a measure of reliability of the instrument.

Matched pairs analyses

We used Matched Pairs Analysis (Ferguson & Takane, 1989) to compare the difference between the means of the different factor scores (for each component of McSELFIE), and compute the significance of that difference. We also calculated the respective effect sizes.

Results

Sample summary and adequacy

Students in our sample indicated that they had on average 3.9 years (SD = 2.8) experience in inquiry-based education, ranging from 1 to 11 years; the median and mode were 5 years. The sample consisted of three student groups, which were (a) biology ($n = 14$), (b) chemistry ($n = 29$), and (c) biochemistry ($n = 67$). The sample included students in five undergraduate years, namely (a) first ($n = 12$), (b) second ($n = 12$), (c) third ($n = 36$), (d) fourth ($n = 36$), and (e) fifth year ($n = 14$). The total sample size was $N = 110$. Table 1 shows the number of students in the different disciplines and years in the sample.

We used the Monte Carlo method to determine the adequacy of the sample size for statistical power (Brown, 2015). Because of our theoretical stance, we treated all item clusters as belonging to single factors representing self-efficacy for SPCs, ILOs, and POS.

In the Monte Carlo population model, we used equal loadings for the items on the single factor of each model tested. We used 10,000 replications and a random seed. The Monte Carlo power analysis with $N = 110$ supported that this sample size would give us 100% power in each tested CFA model.

Item summary statistics

Item summary statistics are provided in Tables 2–4. Tables 2–4 provide means (M), standard deviation (SD), skewness (SK), and kurtosis (KR) values for all items of the instrument, as well as factor means (M) and standard errors of the mean (SEM). In CFA, both univariate

Table 1. Distribution of the sample by university year and science discipline.

Year	Biochemistry	Biology	Chemistry	Total
First	2	8	2	12
Second	6	4	2	12
Third	28	1	7	36
Fourth	21	0	15	36
Fifth	10	1	3	14
Total	67	14	29	110

and multivariate normality should be assessed because the violation of the normality assumption leads to an underestimation of SEs and inflation of chi-square values leading to biased fit indices based on chi-square values (Bandalos & Finney, 2010). Acceptable univariate skewness and kurtosis values are less than |2.0| (or |7.0| for kurtosis according to a more liberal standard), and acceptable value of Mardia's normalized multivariate kurtosis coefficient is less than 3.0 (Bandalos & Finney, 2010). All items except one shown in Tables 2–4 had skewness and kurtosis values less than |2.0| (item 49 had kurtosis = 2.5, which is acceptable according to the liberal standard of kurtosis less than |7.0|), and thus were acceptable, and the condition of univariate normality was satisfied.

Regarding multivariate normality, we examined the value of Mardia's normalized multivariate kurtosis coefficient. The value of the Mardia's normalized multivariate kurtosis coefficient was above the cut-off value of 3.0. Accordingly, the ML estimation method was not used in the CFA model parameter estimation.

The alternative method of estimation that can be used in CFA is the MLR (Yuan & Bentler, 2000) method. The MLR estimation method provides parameter estimates with SEs and a mean adjusted χ^2 test statistic which are robust to non-normality (Brown, 2015; Lei & Wu, 2012; Muthén & Muthén, 2012). Because our data devidated from normality and MLR was suitable for this deviation, we used MLR in the CFA model estimation.

Construct validity using confirmatory factor analysis

The McSELFIE instrument measures the self-efficacy of learners for inquiry engagement in scientific disciplines in higher education. We defined and operationalized inquiry engagement as carrying out the POS, which lead to ILOs, and which are supported by SPCs. The McSELFIE was built by using a literature-based inventory of items from two existing, peer-reviewed, published instruments (MSDIQ and MAVIES).

The origin of the items of the McSELFIE instrument ensured that these items were clearly stated and understood by the students because the items were previously tested in the original MSDIQ and subsequently in MAVIES. The MSDIQ was based on an extensive review of literature on inquiry processes and science practices. The resulting items of the original MSDIQ were criterion-referenced. We established construct validity based on

Table 2. Descriptive statistics for SPCs items.

Factor	#	I believe that I am able to	*M*	SD	SK	KR
Extraversion (*M* = 6.62, SEM = .15)	1	Share the direction of an inquiry with the teacher	6.7	2.0	−0.1	0.2
	44	Help others to observe what happens in the task	7.5	2.0	−0.4	−0.5
	2	Contribute to creating parts of the curriculum	5.7	2.4	−0.5	0
	28	Share my emotions, feelings, ideas, and opinions	7.2	2.3	−0.3	−1.1
	3	Contribute to the decision-making process with my teacher	6.0	2.3	−0.4	0.2
Openness (*M* = 7.96, SEM = .13)	22	Use my imagination freely	7.7	2.1	−0.6	−0.8
	60	Realize that there may be more than one answer to the same problem				
	42	Be open to change	7.8	2.0	−0.7	−0.1
	50	Look for different opinions	7.8	1.7	−0.3	−0.9
Conscientiousness (*M* = 6.97, SEM = .18)	6	Organize my time needed to do inquiry	7.2	2.0	−0.5	0.4
	7	Organize my work space needed to do inquiry	7.4	2.1	−0.7	0.5
	21	Create a back-up plan	6.7	2.3	−0.5	0.1
	20	Create more than one plan	6.6	2.5	−0.4	−0.3

M = mean, SD = standard deviation, SK = skewness, KR = kurtosis, SEM = standard error of the mean.

Table 3. Descriptive statistics for ILOs items.

Factor	#	I believe that I am able to	*M*	SD	SK	KR
Understanding conceptual knowledge (*M* = 7.55, SEM = .14)	16	Link what I know to what I learned	8.0	1.7	−0.4	−0.8
	12	Understand the important concepts	8.0	1.8	−0.6	−0.7
	10	Make a concept map or web or cluster	6.7	2.5	−0.6	0.1
Understanding metacognitive knowledge (*M* = 7.26, SEM = .16)	40	Understand how my personal views affect what I learn	7.3	1.9	−0.2	−0.8
	41	Be aware how the inquiry affects me personally	7.0	2.0	−0.4	0.3
	43	Talk openly about my doubts	7.4	2.1	−0.7	0.1
Application (*M* = 7.93, SEM = .14)	39	Relate what I know to the new ideas I learn about	7.8	1.8	−0.4	−1.1
	61	Use what I learned in the future	8.2	1.9	−0.9	0.5
	5	Use inquiry outside of school	6.0	2.3	−0.4	0.2
Evaluation (*M* = 7.52, SEM = .14)	65	Reflect upon the inquiry experience	7.5	1.9	−0.3	−0.9
	67	Evaluate the inquiry experience	7.1	1.9	−0.4	0.5
	46	Treat everyone's opinions as important	8.0	1.9	−0.6	−0.6
Creating (*M* = 7.42, SEM = .15)	53	Create new knowledge	6.7	2.2	−0.1	−1.0
	68	Follow-up the project with a new set of questions	7.3	2.1	−0.2	−1.0
	4	Work in an encouraging and creative environment	8.3	2.0	−1.3	1.8

M = mean, SD = standard deviation, SK = skewness, KR = kurtosis, SEM = standard error of the mean.

theoretical groundings and CFA, and content validity by using experts' judgments on the adequacy of item-factor relations.

All the specified models that related the different measured items (observed variables) to their respective constructs (latent variables) within each component of the McSELFIE instrument had good data-model-fit, and all the MIs were better than the recommended values (Brown, 2006; Geiser, 2012; Hu & Bentler, 1998; Schermelleh-Engel et al., 2003; West et al., 2012). All factor loadings were positive and less than 1.00, indicating the absence of Heywood cases.

Tables 5–7 list the factors, items per factor, item-factor loadings (IFLs), corresponding standard errors, and residual variances (RVs) for the three components of the McSELFIE instrument Tables 8 and 9 show the goodness-of-fit indices for the tested models.

Content validity using expert judgment

Interrater reliability analysis was conducted for the experts' matching of the survey's self-efficacy items to each of the three components of the McSELFIE instrument. We calculated percentages of agreement among the experts. We wanted to ensure that the self-efficacy components of the instrument (SPCs, ILOs, and POS) were truly represented by the meanings of the self-efficacy items (the individual observed and measured items) with as little overlap as possible between items. We did not hypothesize overlapping items, which means that each item was assigned to a single factor. The experts were asked to judge whether each item represented the corresponding self-efficacy factor of the McSELFIE instrument. They agreed 100% that the items represented corresponding factors.

Reliability

Cronbach's alpha was computed for each latent factor of McSELFIE. Cronbach's alpha was .96 for the entire set of factors. Cronbach's alpha was above .95 for all factors in the three

Table 4. Descriptive statistics for POS items.

Factor	#	I believe that I am able to	*M*	SD	SK	KR
Define the problem (*M* = 7.78, SEM = .14)	13	Understand how to follow instructions	8.8	1.5	−1.2	0.5
	26	Say in my own words what the problem is	7.7	1.9	−0.6	−0.3
	8	Understand the goal of an inquiry task	7.4	2.0	−0.7	0.5
	9	Divide an inquiry task into smaller steps	7.2	2.1	−0.5	0.1
Obtain and evaluate information (*M* = 8.24, SEM = .14)	37	Search the Internet and World Wide Web	9.0	1.6	−1.6	1.7
	36	Look for information beyond textbooks	7.9	2.0	−0.5	−1.1
	38	Separate important from unimportant information	7.8	1.9	−0.4	−1.0
Plan investigation (*M* = 7.37, SEM = .15)	29	Start thinking about what will happen next during inquiry	7.4	1.8	−0.1	−1.1
	27	Make suggestions	7.4	2.1	−0.5	−0.6
	11	Imagine different outcomes during inquiry	7.3	2.2	−0.6	−0.2
	14	Describe how to solve the problem	8.5	1.6	−1.0	0.8
	32	Figure out where to obtain data	7.4	1.8	−0.1	−1.1
	19	Make a plan for the inquiry	6.8	2.2	−0.3	0
Carry out investigation (*M* = 8.12, SEM = .15)	35	Organize data	8.3	1.8	−1.0	0.3
	34	Write down data during inquiry	8.2	1.8	−0.5	−1.1
	31	Make careful observations	7.9	1.7	−0.4	−0.8
Analyze and interpret data (*M* = 7.68, SEM = .14)	45	Find patterns in data	7.5	2.0	−0.4	−0.5
	33	Make sense of how I will use the data to help me solve the problem	7.6	1.8	−0.1	−1.2
	51	Test ideas and hypotheses	7.5	1.6	−0.3	−0.5
	47	Double-check (verify) the data and information	8.1	1.7	−0.4	−0.8
Construct explanations and engage in argument from evidence (*M* = 7.79, SEM = .14)	63	Explain the results	8.1	1.9	−0.7	−0.4
	64	Question the results	7.8	1.9	−0.6	−0.4
	66	Compare and talk about what I learned to what I knew before	7.7	1.8	−0.3	−0.6
	62	Write down how I did an inquiry, the results, and conclusions	8.1	1.9	−0.7	−0.4
	30	Say what I think are the reasons for what happened during an inquiry	6.8	2.0	−0.2	0.1
	48	Compare and contrast my data with someone else's	8.3	1.7	−0.5	−0.8
	49	Explain how the same data can have different meanings for different people	7.6	1.7	−1.0	2.5
Communicate knowledge (*M* = 7.84, SEM = .14)	58	Use words that are good fit to audience and topic	7.8	1.9	−0.4	−1.0
	57	Organize how I will show what I learned	7.6	1.9	−0.5	−0.4
	59	Show data in tables and graphs	7.8	1.9	−0.4	−1.0
	56	Think about various ways to communicate what I learned	7.6	1.8	−0.2	−0.8
	55	Communicate what I learned with others	7.7	1.8	−0.4	−0.8

M = mean, SD = standard deviation, SK = skewness, KR = kurtosis, SEM = standard error of the mean.

components of McSELFIE. These internal consistency reliability results indicated that the internal factorial structure of McSELFIE was homogenous, which was expected based on the theoretical grounding of the instrument.

Results of factor comparisons

Table 10 shows the results of the comparisons of the means of the factors within each McSELFIE component. Figure 1 shows the factorial comparisons for the factors within the four components.

Table 5. Item loadings for SPCs.

Factor	Item	I believe that I am able to	IFL	SE	RV
Extraversion	1	Share the direction of an inquiry with the teacher	.73	.10	.47
	44	Help others to observe what happens in the task	.60	.10	.65
	2	Contribute to creating parts of the curriculum	.56	.10	.69
	28	Share my emotions, feelings, ideas, and opinions	.49	.10	.76
	3	Contribute to the decision-making process with my teacher	.39	.13	.85
Openness	60	Realize that there may be more than one answer to the same problem	.66	.11	.56
	42	Be open to change	.60	.09	.64
	22	Use my imagination freely	.56	.11	.69
	50	Look for different opinions	.55	.11	.70
Conscientiousness	7	Organize my work space needed to do inquiry	.91	.04	.17
	6	Organize my time needed to do inquiry	.87	.05	.24
	21	Create a back-up plan	.61	.08	.63
	20	Create more than one plan	.56	.09	.68

IFL = item-factor loading, SE = standard error, RV = residual variance

Table 6. Item loadings for ILOs.

Factor	Item	I believe that I am able to	IFL	SE	RV
Understanding conceptual knowledge	16	Link what I know to what I learned	.81	.12	.35
	12	Understand the important concepts	.71	.12	.49
	10	Make a concept map or web or cluster	.36	.12	.87
Understanding metacognitive knowledge	40	Understand how my personal views affect what I learn	.92	.10	.16
	41	Be aware of how the inquiry affects me personally	.61	.14	.63
	43	Talk openly about my doubts	.56	.07	.68
Application	39	Relate what I know to the new ideas I learn about	.69	.13	.53
	61	Use what I learned in the future	.60	.14	.64
	5	Use inquiry outside of school	.47	.13	.78
Evaluation	65	Reflect upon the inquiry experience	.90	.11	.20
	67	Evaluate the inquiry experience	.71	.16	.50
	46	Treat everyone's opinions as important	.37	.10	.86
	68	Follow-up the project with a new set of questions	.67	.11	.55
	4	Work in an encouraging and creative environment	.41	.10	.84

IFL = item-factor loading, SE = standard error, RV = residual variance.

Discussion and conclusions

Development and validation of the McSELFIE instrument

The CFA confirmed our hypothesized structures for the McSELFIE instrument. We showed that the McSELFIE is a reliable and valid instrument that can be used with undergraduate students in scientific disciplines to assess their self-efficacy for inquiry engagement. The McSELFIE is composed of three self-efficacy components, namely (a) POS, (b) ILOs, and (c) SPCs. Each component consists of its constituent factors that are represented through their respective items. In our CFA model specifications, we did not specify any double factor loadings. In other words, every item loaded on a single factor. Factor loadings in a CFA completely standardized solution are interpreted as standardized regression coefficients and, additionally, when the measurement model contains no double-loading indicators (as is the case in our specified CFA models), a completely standardized factor loading can also be interpreted as the correlation of the indicator with the factor, because the factor is the only predictor of the indicator, and thus squaring the completely standardized factor loading provides the proportion of variance in the indicator which is explained by the factor (Brown, 2015). IFLs are also important for factor

Table 7. Item loadings for POS.

Factor	#	I believe that I am able to	IFL	SE	RV
Define the problem	13	Understand how to follow instructions	.76	.12	.42
	26	Say in my own words what the problem is	.67	.12	.55
	8	Understand the goal of an inquiry task	.59	.12	.66
	9	Divide an inquiry task into smaller steps	.43	.12	.73
Obtain and evaluate information	37	Search the Internet and World Wide Web	.85	.11	.27
	36	Look for information beyond textbooks	.67	.09	.55
	38	Separate important from unimportant information	.57	.10	.68
Plan investigation	29	Start thinking about what will happen next during inquiry	.84	.05	.30
	27	Make suggestions	.79	.04	.37
	11	Imagine different outcomes during inquiry	.72	.07	.49
	14	Describe how to solve the problem	.66	.06	.57
	19	Make a plan for the inquiry	.66	.06	.57
	32	Figure out where to obtain data	.64	.07	.59
Carry out investigation	35	Organize data	.91	.04	.17
	34	Write down data during inquiry	.82	.06	.32
	31	Make careful observations	.80	.05	.35
Analyze and interpret data	45	Find patterns in data	.78	.07	.39
	33	Make sense of how I will use the data to help me solve the problem	.75	.06	.44
	47	Double-check (verify) the data and information	.67	.08	.55
	51	Test ideas and hypotheses	.67	.07	.55
Construct explanations and engage in argument from evidence	63	Explain the results	.89	.03	.21
	64	Question the results	.83	.06	.32
	66	Compare and talk about what I learned to what I knew	.77	.05	.41
	62	Write down how I did an inquiry, the results, and conclusions	.74	.08	.46
	30	Say what I think are the reasons for what happened during an inquiry	.69	.05	.53
	48	Compare and contrast my data with someone else's	.65	.07	.58
	49	Explain how the same data can have different meanings for different people	.52	.07	.73
Communicate knowledge	55	Communicate what I learned with others	.78	.10	.39
	56	Think about various ways to communicate what I learned	.72	.11	.48
	57	Organize how I will show what I learned	.67	.08	.56
	59	Show data in tables and graphs	.62	.13	.62
	58	Use words that are good fit to audience and topic	.50	.10	.75

IFL = item-factor loading, SE = standard error, RV = residual variance.

interpretability, because items that load the highest on a given factor are considered to be more representative of that factor, while items that load the lowest on a given factor are considered to be less representative of that factor (Pett et al., 2003).

Accordingly, extraversion is highly represented by sharing and helping others during inquiry (.73 and .60, respectively). Openness is highly represented by flexibility in realizing that multiple solutions exist to a problem and by being open to changes (.66 and .60, respectively). Conscientiousness is highly represented by being organized in workspace and in time, which were two items that exhibited high factor loadings (.91 and .87, respectively). Similarly, the item of linking and relating knowledge to what one learns showed a high IFL (.81) for understanding conceptual knowledge, and the items of understanding how personal views affect what one learns showed a high IFL (.92) for understanding metacognitive knowledge. Reflecting upon the inquiry experience loaded highly (.90) on the ILO of evaluation. Creating new knowledge was a high representative item (.78) for the ILO of creation.

Table 8. Goodness-of-fit indices for SPCs and ILOs.

		SPCs			ILOs				
Fit index	Recommended value	Extraversion	Openness	Conscientiousness	Understand conceptual	Understand metacognitive	Apply	Evaluate	Create
SRMR	<.08	0.026	0.025	0.009	0.000	0.000	0.000	0.000	0.000
RMSEA	<.06	0.000	0.000	0.097	0.000	0.000	0.000	0.000	0.000
90% CI		0.000–0.118	0.000–0.170	0.000–0.293	0.000–0.000	0.000–0.000	0.000–0.000	0.000–0.000	0.000–0.000
p (RMSEA) $\leq$ .05		0.737	0.587	0.209	0.000	0.000	0.000	0.000	0.000
CFI	>.95	1.000	1.000	0.989	1.000	1.000	1.000	1.000	1.000
TLI	>.95	1.043	1.042	0.935	1.000	1.000	1.000	1.000	1.000

Table 9. Goodness-of-fit indices for the POS.

		POS						
Fit index	Recommended value	Define the problem	Obtain and evaluate	Plan investigation	Carry out investigation	Analyze and interpret	Construct explanations	Communicate knowledge
SRMR	<.08	0.004	0.000	0.045	0.000	0.017	0.040	0.010
RMSEA	<.06	0.000	0.000	0.100	0.000	0.000	0.059	0.155
90% CI		0.000 to 0.186	0.000–0.000	0.033–0.163	0.000–0.000	0.000–0.160	0.000–0.117	0.000–0.212
p (RMSEA) $\leq$ .05		0.734	0.000	0.092	0.000	0.649	0.367	0.602
CFI	>.95	1.000	1.000	0.951	1.000	1.000	0.980	1.000
TLI	>.95	1.060	1.000	0.918	1.000	1.030	0.970	1.050

Items with IFL that are above .80 were kept because they were representative of the factors and non-redundant to them. Additionally, because several factors were limited in the number of items that represented them, high loading items were preserved. The case of small number of items per factor (in some cases, three items per factor) is a limitation of the current study, and we intend to expand the number of items per factor in a future revision of the instrument. However, the current high-loading and low-loading items were kept because they were chosen to be consistent with the definitions of their respective factor constructs presumably chosen for some reason and that eliminating some of them changes the definition of one's construct(s) to some extent (Bandalos & Finney, 2010, p. 100), and thus we kept items with high loadings (above .80). The highest IFL was .91 (less than .95), so such an item is still considered salient and independent, but with a high correlation with its respective factor.

IFLs for the self-efficacy for POS showed that the highest loading items represented their corresponding factors meaningfully. The items of self-efficacy for 'communicating what I learned with others,' and self-efficacy for 'thinking about various ways to communicate what I learned' were salient items for the self-efficacy for communicating knowledge. The items of self-efficacy for 'explaining the results,' and self-efficacy for 'questioning the results' were salient items for the self-efficacy for constructing explanations and engaging in argument from evidence. The items of self-efficacy for 'finding patterns in data,' and self-efficacy for 'making sense of how to use the data to solve the problem' were salient items for the self-efficacy for analyzing and interpreting data. The items of self-efficacy for 'organizing data,' and self-efficacy for 'writing down data during inquiry' were salient items for the self-efficacy for carrying out investigation. The items of self-efficacy for 'starting thinking about what will happen next during inquiry,' and self-efficacy for 'making suggestions' were salient items for the self-efficacy for planning investigations. The items of self-efficacy for 'searching the Internet,' and self-efficacy for 'looking for information beyond textbooks' were salient items for the self-efficacy for obtaining and evaluating information. Finally, the items of self-efficacy for 'understanding how to follow instructions,' and self-efficacy for 'saying in one's own words what the problem is' were salient items for the self-efficacy for defining a problem.

Comparison of the McSELFIE factors

The differences among the McSELFIE factors making up the components of the instrument allowed us to compare how students in natural science disciplines differentially evaluated their confidence in the different POS, ILOs, and SPCs. Such comparisons highlight significant differences among the self-efficacy ratings assigned to POS, ILOs, and students' SPCs. Comparing the different factors within each component of the McSELFIE instrument provided a portrayal of students' self-efficacy and for asking questions about the practical and theoretical implications of the difference results.

Openness has the highest self-efficacy and extraversion has the lowest self-efficacy in SPCs

Undergraduate students in the natural sciences assigned a statistically significantly higher confidence (with a strong effect size; see Table 10) to openness compared to either conscientiousness or extraversion. Having a high self-efficacy for openness is in line with

Table 10. Matched pair analyses for the McSELFIE instrument.

Factor 1	*M*1	SD1	Factor 2	*M*2	SD2	*r*	*t*-value	*p*-value	Cohen's *d*
Openness	7.96	1.34	Extraversion	6.62	1.51	.51	$t(102) = 9.17$	<.0001*	.95
Openness	7.96	1.34	Conscientious	6.97	1.87	.39	$t(105) = 5.48$	<.0001*	.52
Conscientious	6.97	1.87	Extraversion	6.6	1.50	.57	$t(104) = 2.59$	=.0055*	.25
Application	7.93	1.42	Conceptual	7.55	1.50	.63	$t(107) = 2.92$	=.0021*	.32
Application	7.93	1.42	Metacognitive	7.26	1.64	.63	$t(105) = 5.18$	<.0001*	.46
Application	7.93	1.42	Evaluation	7.52	1.49	.58	$t(106) = 3.13$	=.0011*	.30
Application	7.93	1.42	Creation	7.42	1.59	.73	$t(106) = 4.56$	<.0001*	.45
Evaluation	7.52	1.49	Conceptual	7.50	1.50	.50	$t(104) = 0.07$	=.53	–
Evaluation	7.52	1.49	Metacognitive	7.26	1.64	.60	$t(104) = 1.85$	=.034*	.14
Evaluation	7.52	1.49	Creation	7.42	1.59	.56	$t(104) = 0.60$	=.27	–
Creation	7.42	1.59	Conceptual	7.50	1.50	.56	$t(104) = 0.67$	=.25	–
Creation	7.42	1.59	Metacognitive	7.26	1.64	.52	$t(104) = 2.50$	=.86	–
Conceptual	7.55	1.50	Metacognitive	7.26	1.64	.50	$t(104) = 1.70$	=.05*	.13
Obtain	8.24	1.49	Define	7.78	1.48	.65	$t(102) = 3.59$	=.0003*	.37
Obtain	8.24	1.49	Plan	7.37	1.52	.69	$t(102) = 7.51$	<.0001*	.73
Obtain	8.24	1.49	Carry out	8.12	1.60	.70	$t(102) = 0.85$	=.20	–
Obtain	8.24	1.49	Analyze	7.68	1.42	.77	$t(102) = 5.82$	<.0001*	.57
Obtain	8.24	1.49	Explain	7.79	1.41	.70	$t(102) = 4.36$	<.0001*	.40
Obtain	8.24	1.49	Communicate	7.84	1.44	.67	$t(102) = 3.29$	=.0007*	.34
Carry out	8.12	1.60	Define	7.78	1.48	.70	$t(102) = 2.86$	=.0026*	.28
Carry out	8.12	1.60	Plan	7.37	1.52	.75	$t(102) = 7.15$	<.0001*	.68
Carry out	8.12	1.60	Analyze	7.68	1.42	.77	$t(102) = 4.58$	<.0001*	.32
Carry out	8.12	1.60	Explain	7.79	1.41	.84	$t(102) = 4.38$	<.0001*	.38
Carry out	8.12	1.60	Communicate	7.84	1.44	.69	$t(102) = 2.42$	=.0087*	.23
Plan	7.37	1.52	Define	7.78	1.48	.85	$t(102) = 5.51$	<.0001*	.50
Plan	7.37	1.52	Analyze	7.68	1.42	.76	$t(102) = 3.03$	=.0031*	.30
Plan	7.37	1.52	Explain	7.79	1.41	.81	$t(102) = 4.38$	<.0001*	.46
Plan	7.37	1.52	Communicate	7.84	1.44	.72	$t(102) = 4.48$	<.0001*	.42

*Significance at .05 level.

inquiry because inquiry encourages curiosity, imagination, and flexibility. Thus, the high rating of self-efficacy for openness is consistent with the ideas of investigating and exploring that are advocated in inquiry. Inquiry and engaging in scientific investigations in education and research also includes collaboration between the student and teacher or mentor and with peers. Latour and Woolgar (1986) mentioned, 'scientific activity is just one social arena in which knowledge is constructed' (p. 31). In a context in which scientific activity is highly social, it is not encouraging to see that students in the natural sciences rate their confidence in sharing and collaborating significantly less than openness (with a strong effect size). This result points to the attention that needs to be directed towards highlighting and encouraging the collaborative and social nature of engaging in science and inquiry. Similarly, undergraduate students in the natural sciences did not rate their self-efficacy for conscientiousness as highly as their self-efficacy for openness; conscientiousness was also rated as significantly less important than openness (with a strong effect size; see Table 10).

Applying knowledge has the highest self-efficacy and understanding metacognitive knowledge has the lowest self-efficacy in ILOs

Undergraduate students in the natural sciences rated their self-efficacy for achieving the 'application' learning outcome statistically higher (with a strong effect size; see Table 10) than their self-efficacy for achieving any other learning outcome, including conceptual understanding, metacognition, evaluation, and creation. The self-efficacy for applying knowledge included applying prior knowledge to new concepts and applying new

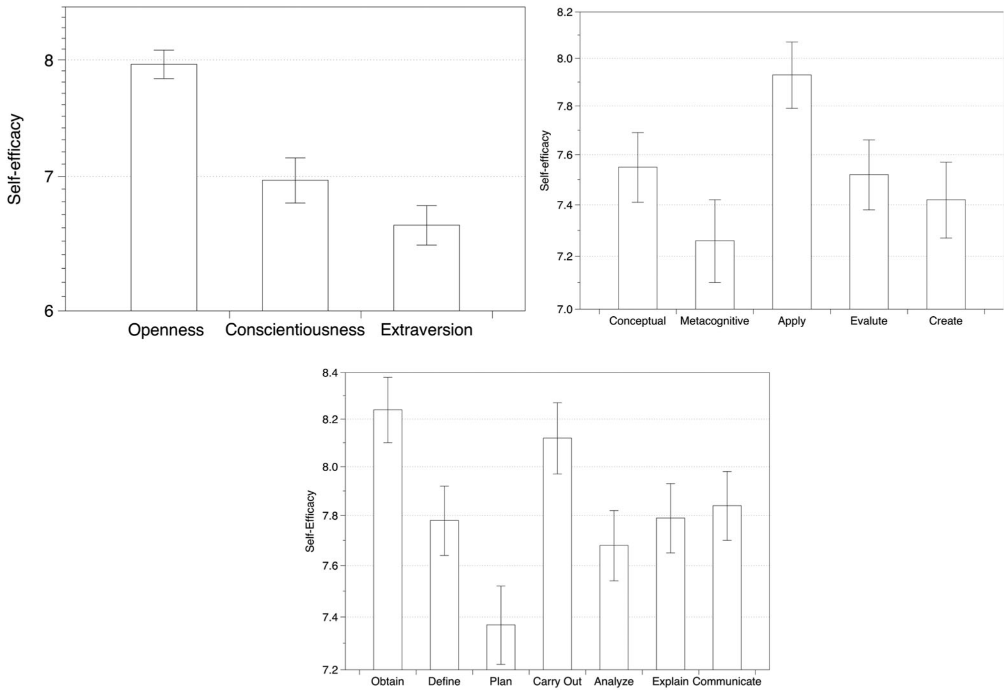

Figure 1. Comparison of McSELFIE factors.

knowledge to future experiences. Although applying knowledge is considered a high-level cognitive educational objective (Gogus, 2012), developing interpretive knowledge (Broudy, 1977; Schwartz, Bransford, & Sears, 2005) and conceptual understanding (Bransford, Brown, & Cocking, 1999) are essential for applying knowledge in more complex situations. Undergraduate students in the natural sciences in our study did not rate their self-efficacy for conceptual understanding or metacognition as highly as their self-efficacy for applying knowledge. The NRC's (2012) emphasis on understanding disciplinary core ideas called to attention the importance of conceptual understanding of core ideas in the disciplines. The emphasis on application may have been carried over from previous standards that emphasized 'doing' as opposed to 'knowing.' In the current study, we faced this situation with undergraduate students in the natural sciences who felt more confident in applying knowledge than understanding. This situation calls for more research on instructional practices and educational objectives.

Obtaining and evaluating information and carrying out investigations have the highest self-efficacy and planning investigation has the lowest self-efficacy in POS

Undergraduate students in the natural sciences rated their self-efficacy for obtaining and evaluating information and their self-efficacy for carrying out investigations statistically significantly higher (with strong effect sizes; see Table 10) than their self-efficacy for all other POS. Undergraduate students in the natural sciences rated their self-efficacy for obtaining and evaluating information (searching the Internet, searching for resources, and separating relevant and irrelevant information) statistically significantly higher (with strong effect sizes; see Table 10) than their self-efficacy for other POS including planning investigations, analyzing and interpreting data, and communicating knowledge. Although obtaining and evaluating information is an essential POS, it is striking to see other POS having a significantly lower self-efficacy compared to obtaining and evaluating information. Technological tools have made information search seemingly fast and accessible, and this may have influenced the students' confidence in obtaining and evaluating information. Further research needs to be conducted to investigate students' understanding of the complexity of information search including designing search strategies and implementing high-quality searches in credible sources. It might also be useful to separate the obtaining and evaluating parts of this POS; easily finding evidence does not equate with finding good evidence or knowing how to evaluate its quality.

Undergraduate students in the natural sciences rated their self-efficacy for carrying out investigations (organizing data, collecting data, making observations) statistically significantly higher (with strong effect sizes; see Table 10) than their self-efficacy for other POS including defining the problem, planning investigations, analyzing and interpreting data, and explaining results. The rating of high self-efficacy for carrying out investigations is consistent with the high self-efficacy of applying knowledge as evidenced by a strong linear relation between the two variables.

Undergraduate students in the natural sciences rated their self-efficacy for planning investigations (hypothesizing outcomes, making suggestions, making plans) statistically significantly less (with strong effect sizes; see Table 10) than their self-efficacy for all other POS including defining the problem, analyzing and interpreting data, and explaining results. The rating of low self-efficacy for planning investigations is consistent with the low self-efficacy of conscientiousness as evidenced by a strong linear relation between the two variables. Perhaps undergraduate students in the natural sciences did not have opportunities to plan investigations and thus feel less confident in carrying out this POS. It is conceivable that assignments may require students to obtain and evaluate information from articles and books and this may be the reason why, in our sample, students expressed a high confidence for obtaining and evaluating information. It is also conceivable that undergraduate students in the natural sciences have opportunities to work in laboratory setting to carry out experiments, and this may be the reason they expressed a higher confidence in carrying out investigations.

Our results point to the importance of highlighting the practice of planning investigations and trying to implement methods that could enhance students' self-efficacy for carrying out this practice. This practice would ensure preparing students to work on experiments or investigations that they designed and planned and not only work on experiments or investigations that have been set for them by their instructors or teachers.

Theoretical implications

Inquiry engagement

Inquiry engagement is a new construct (Ibrahim et al., 2016) that involves SPCs, teachers' roles, POS, and ILOs. In the context of science education, we defined inquiry engagement as students' carrying out of POS to achieve ILOs. Carrying out POS entails students' involvement through their personality characteristics (SPCs) and other cognitive, affective, motivational, metacognitive, and epistemic variables. By identifying inquiry engagement as an important construct, we add several contributions to the literature on inquiry and science education.

First, in the context of science education, this contribution helps ease tension between inquiry as a contested construct with multiple definitions and the POS that are more recently advocated as an educational imperative in science curricula and instruction. In this way, inquiry is not abandoned and practices are integrative and essential in inquiry, especially in the context of science in which POS become integrative to scientific inquiry. Second, adding engagement to inquiry emphasizes the dynamic nature of participating in science. Third, identifying and operationalizing inquiry engagement allows for creating instruments that can be used to measure students' motivations, valuing, self-efficacy, or attitudes towards this construct.

Components of inquiry engagement

By operationalizing inquiry engagement, we were able to construct a self-efficacy instrument that measured undergraduate students' confidence for exhibiting SPCs, carrying out POS, and achieving ILOs. These components could help further investigate relations among the subcomponents of the instrument. Figure 2 illustrates examples of possible relations (shown in dotted lines) that can be tested among the subcomponents of the instrument.

Consequently, examining the relations in the above figure should enhance our understanding of the involvement in carrying out of POS and how this could contribute to achieving ILOs. This work contributes to understanding learning through engaging in POS and inquiry in science education and has implications for understanding self-efficacy and sociocognitive theory in science education.

Practical implications

The McSELFIE instrument

The McSELFIE instrument fills a gap in the science education literature by providing a measure of self-efficacy that incorporates the NGSS and PSE (NGSS Lead States, 2013). The NGSS publication is influencing many changes in teaching, curriculum, and assessment. It is important to align our assessment of students' psychological constructs, such as self-efficacy, with the actual implementations of curriculum and actual performance in science classrooms. It is important to use instruments that are aligned with reform in science education. The McSELFIE is a suitable example of an instrument that can be used to assess students' self-efficacy for the POS, and to try to predict ILOs. We propose that the McSELFIE can inform instructors about several important aspects about students and their confidence in scientific activities.

The factors of the McSELFIE instrument

The comparison of the McSELFIE factors within the instruments' subcomponents (SPCs, ILOs, and POS) showed results that could have important implications for instruction and curriculum. Instruction in science education should focus on enhancing students' self-efficacy for sharing and collaborating and for planning and managing projects in addition

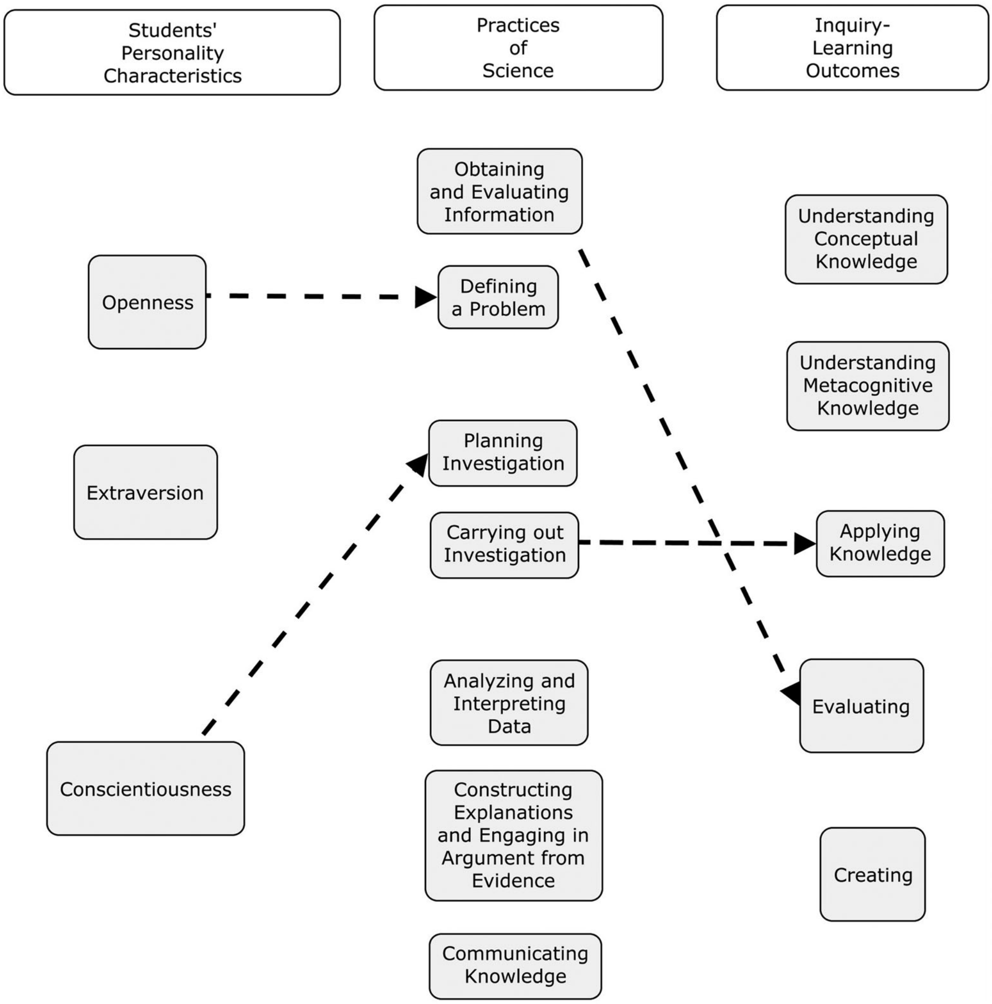

Figure 2 .Testing relations among the McSELFIE factors.

to openness and flexibility. Instruction and interventions should also focus more on students' conceptual understanding and metacognitive knowledge in addition to applying knowledge in order to reach more flexibility in transfer of learning to more complex and novel situations and problems. Finally, instructional activities should incorporate more opportunities for students to plan investigations in addition to other POS such as carrying out investigations, defining problems, and obtaining and evaluating information or evidence in general.

Limitations and future research

The items of the McSELFIE

McSELFIE could be improved by expanding the item pool of the instrument. Several factors of the instrument are limited to three items. In order to capture the full meaning of the factors, more items could be added to increase the precision of the instrument.

The factors of the McSELFIE instrument

McSELFIE did not include 'asking questions,' 'using mathematics and computational thinking,' and 'using and developing models.' In order to fully capture the essence of scientific inquiry as a set of practices that cover a wide spectrum of activities that students can be engaged in, we need to include all the POS. The practices that were not included in the current instrument are generally missing in other instruments because the statement of these practices and the emphasis on their inclusion in the framework of science education and the NGSS is new.

Future research

Future research can examine the validity and reliability of the McSELFIE, especially with added items, and the possible use of the instrument in other natural science disciplines. Future research can also test the different relations shown in Figure 2.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Ashton, M. C., & Lee, K. (2001). A theoretical basis for the major dimensions of personality. *European Journal of Personality*, *15*, 327–353. doi:10.1002/per.417

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO Model of personality structure. *Personality and Social Psychology Review*, *11*, 150–166. doi:10.1177/1088868306294907

Aulls, M. W., & Shore, B. M. (2008). *Inquiry in education (Vol. 1): The conceptual foundations for research as a curricular imperative*. New York, NY: Erlbaum.

Baldwin, J. A., Ebert-May, D., & Burns, D. J. (1999). The development of college biology self-efficacy instrument for nonmajors. *Science Education*, *83*, 397–408. doi:10.1002/(SICI)1098-237X(199907)83:4<397::AID-SCE1>3.0.CO;2-#

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93–114). New York, NY: Routledge.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191–215. doi:10.1037/0033-295x.84.2.191

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*, 122–147. doi:10.1037/0003-066x.37.2.122

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

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, *44*, 1175–1184. doi:10.1037/0003-066x.44.9.1175

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, *50*, 248–287. doi:10.1016/0749-5978(91)90022-L

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

Bandura, A. (2006). Guide to constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich, CT: Information Age.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Development*, *67*, 1206–1222. doi:10.2307/1131888

Bartimote-Aufflick, K., Bridgeman, A., Walker, R., Sharma, M., & Smith, L. (2015). The study, evaluation, and improvement of university student self-efficacy. *Studies in Higher Education*, 1–25. doi:10.1080/03075079.2014.999319

Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher Education*, *19*, 151–161. doi:10.1080/03075079412331382007

Bieschke, K. J., Bishop, R. M., & Garcia, V. L. (1996). The utility of the Research Self-Efficacy Scale. *Journal of Career Assessment*, *4*, 59–75. doi:10.1177/106907279600400104

Bong, M. (2006). Asking the right question: How confident are you that you could successfully perform these tasks? In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 287–303). Greenwich, CT: Information Age.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academies Press.

Broudy, H. S. (1977). Types of knowledge and purposes of education. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 1–17). Hillsdale, NJ: Erlbaum.

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

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Bunterm, T., Lee, K., Ng Lan Kong, J., Srikoon, S., Vangpoomyai, P., Rattanavongsa, J., & Rachahoon, G. (2014). Do different levels of inquiry lead to different learning outcomes? A comparison between guided and structured inquiry. *International Journal of Science Education*, *36*, 1937–1959. doi:10.1080/09500693.2014.886347

Buss, D. M. (1991). Evolutionary personality psychology. *Annual Review of Psychology*, *42*, 459–491.

Buss, D. M. (1996). Social adaptation and five major factors of personality. In J. S. Wiggins (Ed.), *The five factor model of personality: Theoretical perspectives* (pp. 180–207). New York, NY: Guilford Press.

Carberry, A. R., Lee, H. S., & Ohland, M. W. (2010). Measuring engineering design self-efficacy. *Journal of Engineering Education*, *99*, 71–79. doi:10.1002/j.2168-9830.2010.tb01043.x

Cavallo, A. M. L., Potter, W. H., & Rozman, M. (2004). Gender differences in learning constructs, shifts in learning constructs, and their relationship to course achievement in a structured inquiry, yearlong college physics course for life science majors. *School Science and Mathematics*, *104*, 288–300. doi:10.1111/j.1949-8594.2004.tb18000.x

Chen, J. A., Metcalf, S. J., & Tutwiler, M. S. (2014). Motivation and beliefs about the nature of scientific knowledge within an immersive virtual ecosystems environment. *Contemporary Educational Psychology*, *39*, 112–123. doi:10.1016/j.cedpsych.2014.02.004

Committee on Integrated STEM Education, Honey, M., Pearson, G., & Schweingruber, H. A. (2014). *STEM integration on K-12 education*. Washington, DC: National Academies Press.

Denissen, J. J. A., & Penke, L. (2008). Motivational individual reaction norms underlying the five-factor model of personality: First steps towards a theory-based conceptual framework. *Journal of Research in Personality*, *42*, 1285–1302. doi:10.1016/j.jrp.2008.04.002

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41*, 417–440. doi:10.1146/annurev.ps.41.020190.002221

Erduran, S., & Dagher, Z. (2014). *Reconceptualizing the nature of science education: Scientific knowledge, practices, and other family categories*. Dordrecht: Springer.

Fencl, H. S., & Scheel, K. R. (2004). Pedagogical approaches, contextual variables, and the development of student self-efficacy in undergraduate physics courses. In J. Marx, S. Franklin, & K.

Cummings (Eds.), *Proceedings of the 2003 physics education research conference* (Vol. 720, pp. 173–176). College Park, MD: American Institute of Physics.

Ferguson, G. A., & Takane, Y. (1989). *Statistical analysis in psychology and education* (6th ed.). New York, NY: McGraw-Hill.

Ford, M. J. (2015). Educational implications of choosing 'practice' to describe science in the next generation science standards. *Science Education*, *99*, 1041–1048. doi:10.1002/sce.21188

Geiser, C. (2012). *Data analysis with Mplus*. New York, NY: Guilford Press.

Gogus, A. (2012). Outcomes of learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 2534–2539). Boston, MA: Springer.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*, 26–34. doi:10.1037/0003-066X.48.1.26

Gorin, J. S. (2007). Reconsidering issues in validity theory. *Educational Researcher*, *36*, 456–462. doi:10.3102/0013189X07311607

Griffiths, R. (2004). Knowledge production and the research-teaching nexus: The case of the built environment disciplines. *Studies in Higher Education*, *29*, 709–726. doi:10.1080/0307507042000287212

Healey, M. (2005a). Linking research and teaching: Disciplinary spaces. In R. Barnett (Ed.), *Reshaping the university: New relationships between research, scholarship and teaching* (pp. 30–42). Maidenhead: Open University Press.

Healey, M. (2005b). Linking research and teaching to benefit student learning. *Journal of Geography in Higher Education*, *29*, 183–201. doi:10.1080/03098260500130387

Healey, M., Jordan, F., Pell, B., & Short, C. (2010). The research-teaching nexus: A case study of students' awareness, experiences and perceptions of research. *Innovations in Education and Teaching International*, *47*, 235–246. doi:10.1080/14703291003718968

Holden, G., Barker, K., Meenaghan, T., & Rosenberg, G. (1999). Research self-efficacy: A new possibility for educational outcomes assessment. *Journal of Social Work Education*, *35*, 463–476. Retrieved from http://www.jstor.org/stable/23043572.

Honicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review*, *17*, 63–84. doi:10.1016/j.edurev.2015.11.002

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424–453. doi:10.1037/1082-989x.3.4.424

Ibrahim, A., Aulls, M. W., & Shore, B. M. (2016). Teachers' roles, students' personalities, inquiry learning outcomes, and practices of science and engineering: The development and validation of the McGill Attainment Value for Inquiry Engagement Survey in STEM disciplines. *International Journal of Science and Mathematics Education*. Advance online publication. doi:10.1007/s10763-016-9733-y

Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology*, *41*, 13–24. doi:10.1016/j.cedpsych.2014.11.002

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). New York, NY: Guilford Press.

Kane, M. T. (2008). Terminology, emphasis, and utility in validation. *Educational Researcher*, *37*, 76–82. doi:10.3102/0013189X08315390

Ketelhut, D. J. (2007). The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in 'River City,' a multi-user virtual environment. *Journal of Science Education and Technology*, *16*, 99–111. doi:10.1007/sl0956-006-9038-y

Ketelhut, D. J. (2010). Assessing gaming, computer and scientific inquiry self-efficacy in a virtual environment. In L. Annetta & S. Bronsak (Eds.), *Serious educational game assessment: Practical methods and models for educational games, simulations and virtual worlds* (pp. 1–18). New York, NY: Sense.

Knapp, T. R., & Mueller, R. O. (2010). Reliability and validity of instruments. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 337–342). New York, NY: Routledge.

Komarraju, M., Karau, S. J., Schmeck, R. R., & Avdic, A. (2011). The big five personality traits, learning styles, and academic achievement. *Personality and Individual Differences*, *51*, 472–477. doi:10.1016/j.paid.2011.04.019

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, *41*, 212–218. doi:10.1207/s15430421tip4104_2

Kuzel, A. J. (1999). Sampling in qualitative inquiry. In B. F. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (2nd ed., pp. 33–45). Thousand Oaks, CA: Sage.

Lamb, R. L., Vallett, D., & Annetta, L. (2014). Development of a short-form measure of science and technology self-efficacy using Rasch analysis. *Journal of Science Education and Technology*, *23*, 641–657. doi:10.1007/S10956-014-9491-Y

Latour, B., & Woolgar, S. (1986). *Laboratory life: The social construction of scientific facts*. Princeton, NJ: Princeton University Press.

Lederman, N. G., & Lederman, J. S. (2013). Is it STEM or 'S & M' that we truly love? *Journal of Science Teacher Education*, *24*, 1237–1240. doi:10.1007/s10972-013-9370-z

Lei, P. W., & Wu, Q. (2012). Estimation in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 164–180). New York, NY: Guilford Press.

Lent, R. W., Brown, S. D., & Larkin, K. C. (1984). Relation of self-efficacy expectations to academic achievement and persistence. *Journal of Counseling Psychology*, *31*, 356–362. doi:10.1037/0022-0167.31.3.356

McCrae, R. R. (1993). Openness to experience as a basic dimension of personality. *Imagination, Cognition and Personality*, *13*, 39–55. doi:10.2190/h8h6-qykr-keu8-gaq0

McCrae, R. R., & Costa, P. T. (1996). Toward a new generation of personality theories: Theoretical contexts for the Five-Factor Model. In J. S. Wiggins (Ed.), *The Five Factor Model of personality: Theoretical perspectives* (pp. 51–87). New York, NY: Guilford Press.

McCrae, R. R., & Costa, P. T. (1997). Conceptions and correlates of openness to experience. In R. Hogan, J. A. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 825–847). San Diego, CA: Academic Press.

McCrae, R. R., & John, O. P. (1992). An introduction to the Five-Factor Model and its applications. *Journal of Personality*, *60*, 175–215. doi:10.1111/j.1467-6494.1992.tb00970.x

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.

Morizot, J. (2014). Construct validity of adolescents' self-reported big five personality traits: Importance of conceptual breadth and initial validation of a short measure. *Assessment*, *21*, 580–606. doi:10.1177/1073191114524015

Muthén, L. K., & Muthén, B. O. (2012). Special modeling issues. In L. K. Muthén & B. O. Muthén (Eds.), *Mplus user's guide* (pp. 459–502). Los Angeles, CA: Authors.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academies Press.

National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, DC: National Academies Press.

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

Nelson, B., & Ketelhut, D. (2008). Exploring embedded guidance and self-efficacy in educational multi-user virtual environments. *International Journal of Computer-Supported Collaborative Learning*, *3*, 413–427. doi:10.1007/s11412-008-9049-1

NGSS Lead States. (2013). *Next generation science standards: For states, by states* (Vol. 1). Washington, DC: National Academies Press.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*, 543–578. doi:10.3102/00346543066004543

Pajares, F., & Schunk, D. H. (2001). Self-beliefs and school success: Self-efficacy, self-concept, and school achievement. In R. Riding & S. Rayner (Eds.), *Perception* (pp. 239–266). London: Ablex.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.
Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oaks, CA: Sage.
Redden, K. C., Simon, R. A., & Aulls, M. W. (2007). Alignment in constructivist-oriented teacher education: Identifying pre-service teacher characteristics and associated learning outcomes. *Teacher Education Quarterly*, *34*, 149–164.
Repko, A. F., & Szostak, R. (2017). *Interdisciplinary research: Process and theory* (3rd ed.). Thousand Oaks, CA: Sage.
Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*, 353–387. doi:10.1037/a0026838
Rosenberg, A. (2000). *Philosophy of science* (2nd ed.). New York, NY: Routledge.
Saunders-Stewart, K. S., Gyles, P. D. T., & Shore, B. M. (2012). Student outcomes in inquiry instruction. *Journal of Advanced Academics*, *23*, 5–31. doi:10.1177/1932202x11429860
Saunders-Stewart, K. S., Gyles, P. D. T., Shore, B. M., Bracewell, R. J. (2015). Student outcomes in inquiry: Students' perspectives. *Learning Environments Research*, *18*, 289–311. doi:10.1007/s10984-015-9185-2
Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*, 23–74. Retrieved from http://www.dgps.de/fachgruppen/methoden/mpr-online/issue20/art2/mpr130_13.pdf.
Schunk, D. H., & Pajares, F. (2009). Self-efficacy theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 35–53). New York, NY: Routledge.
Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. P. Mestre (Ed.), *Transfer of learning: From a modern multidisciplinary perspective* (pp. 1–52). Greenwich, CT: Information Age.
Seth, D., Ibrahim, A., & Tangorra, J. (2015). *Measuring undergraduate students' self-efficacy in engineering design in a project-based design course*. Proceedings of the IEEE Frontiers in Education 2015 conference, 1375–1382. Retrieved from http://fie2015.org/sites/file2015.fie-conference.org/files/FIE-2015_Proceedings_v11.pd
Shore, B. M., Chichekian, T., Syer, C. A., Aulls, M. W., & Frederiksen, C. H. (2012). Planning, enactment, and reflection in inquir-based learning: Validating the McGill Strategic Demands of Inquiry Questionnaire. *International Journal of Science and Mathematics Education*, *10*, 315–337. doi:10.1007/s10763-011-9301-4
Spronken-Smith, R., Walker, R., Batchelor, J., O'Steen, B., & Angelo, T. (2012). Evaluating student perceptions of learning processes and intended learning outcomes under inquiry approaches. *Assessment and Evaluation in Higher Education*, *37*, 57–72. doi:10.1080/02602938.2010.496531
Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin*, *124*, 240–261. doi:10.1037/0033-2909.124.2.240
Toulmin, S. (1972). *Human understanding*. Princeton, NJ: Princeton University Press.
Watson, P. (2002). The role and integration of learning outcomes into the educational process. *Active Learning in Higher Education*, *3*, 205–219. doi:10.1177/1469787402003003002
West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). New York, NY: Guilford Press.
Whitley, R. (1984). *The intellectual and social organization of the sciences*. New York, NY: Oxford University Press.
Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. In M. E. Sobel & M. P. Becker (Eds.), *Sociological methodology* (pp. 165–200). Washington, DC: American Sociological Association.