Western SGraduate & Postdoctoral Studies

Western University Scholarship@Western

Electronic Thesis and Dissertation Repository

5-27-2021 10:00 AM

Systematic Review of STEM Attitude Measures

Meagan Koufis, The University of Western Ontario

Supervisor: Neil, Nicole, *The University of Western Ontario* A thesis submitted in partial fulfillment of the requirements for the Master of Arts degree in Education © Meagan Koufis 2021

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Science and Mathematics Education Commons

Recommended Citation

Koufis, Meagan, "Systematic Review of STEM Attitude Measures" (2021). *Electronic Thesis and Dissertation Repository*. 7929. https://ir.lib.uwo.ca/etd/7929

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Abstract

The aim of this review was to identify measurement tools used to assess students' attitudes towards science, technology, engineering, and/or mathematics (STEM) in publications, and evaluate their usefulness and appropriateness based on reported psychometric properties. Researchers look to also evaluate the content/subject area coverage as well as the attitude construct coverage of identified measures. Electronic databases were searched for peer-reviewed articles that utilized a STEM attitude measurement tool with elementary or high school students. Publications were examined for reported psychometric properties of tools and the quality of each measure was evaluated using the psychometric grading framework (PGF). A total of 104 STEM attitude measures were identified within the literature. Many (n = 83) identified measures lacked pertinent reliability and/or validity data and could not be evaluated using the PGF within the current review. Fifteen measures were evaluated, with eleven receiving the highest overall grade of "Good" using the PGF.

Keywords

Science, Technology, Engineering, Mathematics (STEM), Self-Report Measure, Attitudes, Interest, Psychometric Grading Framework (PGF), Social Cognitive Career Theory

Summary for Lay Audience

As the world becomes more technologically advanced, more career positions are available in the field of science, technology, engineering, and mathematics (STEM). However, student enrollment in said academic courses is decreasing and reported attitudes towards STEM are low in adolescence. It is pertinent to increase STEM engagement and interest in students due to the relationship between various internal/motivational factors such as self-efficacy and interest and its subsequent effect on future outcomes. Researchers are attempting to accomplish this via various educational programs. Unfortunately, research that evaluates STEM attitude measurement tools is lacking, thereby making it difficult to accurately ensure educational efforts are achieving their desired goals of targeting student attitudes. This review aims to identify measurement tools used to assess students' attitudes towards science, technology, engineering, and/or math in publications, as well as evaluate said tools usefulness and appropriateness based on reported psychometric properties. Electronic databases were searched for articles that used a STEM attitude measurement tool with elementary or high school students. Articles were examined for reported reliability and validity information of tools and the quality of each measure was evaluated using this extracted data. A total of eleven STEM attitude measures were found within the literature, that were identified within multiple research articles, publishing reliability and validity information. These eleven measures were evaluated for their definition of attitudes, the STEM subject areas included, and the strength of psychometric data. These measures were ultimately awarded a high grade and recommended for use by the present researchers.

Acknowledgments

I wish to express my deepest gratitude to my supervisor, Dr. Nicole Neil, who provided me with an abundance of guidance, support, and nurturing throughout my graduate studies. Thank you for the opportunity to be a part of your lab as well as work and study under your guidance.

I also wish to acknowledge the support and love of my family who encouraged me to work hard and aspire to succeed. To my Dad, thank you for teaching me dedication and perseverance, and always showing me how to enjoy the experiences of life. To my mom, thank you for your constant reassurance in times of doubt and being the powerful woman I strive to become. To my brother Mathew, thank you for always being my greatest supporter.

Table of Contents

A	bstractii			
Sı	ummary for Lay Audienceiii			
A	Acknowledgmentsiv			
Table of Contentsv				
List of Tables				
List of Figuresix				
List of Appendicesx				
In	troduction1			
1	Rationale1			
2	What is STEM?			
3	STEM Attitudes and Social Cognitive Career Theory			
4	Measuring STEM Attitudes10			
5	Assessing Quality of Instruments			
Μ	lethod17			
6	Search Strategy			
7	Inclusion and Exclusion Criteria17			
8	Title/Abstract Review17			
9	Full-Text Review			
1() Variable Coding/Data Extraction			
	10.1 Target Age Group19			
	10.2Number of Items			
	10.3Format of Measure19			
	10.4Completion Time			

10.5 Administration				
10.6STEM Subject Area Coverage & Categorization20				
10.7Attitude Construct Coverage21				
10.8Quality Appraisal: Psychometric Grading Framework22				
Results				
11 All Identified STEM Attitude Measures				
11.1Completion Time24				
11.2Administration24				
11.3Number of Items24				
11.4Content Coverage/Subject Areas of Measures25				
11.5Attitude Construct Coverage25				
11.6Evaluation of Psychometric Quality25				
11.6.1 Reliability25				
•				
11.6.2 Validity27				
11.6.2 Validity				
11.6.2 Validity 27 12 Determining Articles for Psychometric Grading 29 12.1 Group 1: Measures Cited Within One Article 29				
11.6.2 Validity 27 12 Determining Articles for Psychometric Grading 29 12.1Group 1: Measures Cited Within One Article 29 12.2Group 2: Measures Cited More than Once with only Reliability Information 29				
11.6.2 Validity 27 12 Determining Articles for Psychometric Grading 29 12.1 Group 1: Measures Cited Within One Article 29 12.2 Group 2: Measures Cited More than Once with only Reliability Information 29 13 Group 3 Measures 30				
11.6.2 Validity 27 12 Determining Articles for Psychometric Grading 29 12.1 Group 1: Measures Cited Within One Article 29 12.2 Group 2: Measures Cited More than Once with only Reliability Information 29 13 Group 3 Measures 30 13.1 Content coverage/subject Areas of Measures 30				
11.6.2 Validity 27 12 Determining Articles for Psychometric Grading 29 12.1 Group 1: Measures Cited Within One Article 29 12.2 Group 2: Measures Cited More than Once with only Reliability Information 29 13 Group 3 Measures 30 13.1 Content coverage/subject Areas of Measures 30 13.2 Attitude Construct Coverage of Measures 31				
11.6.2 Validity2712 Determining Articles for Psychometric Grading2912.1 Group 1: Measures Cited Within One Article2912.2 Group 2: Measures Cited More than Once with only Reliability Information2913 Group 3 Measures3013.1 Content coverage/subject Areas of Measures3013.2 Attitude Construct Coverage of Measures3113.3 Evaluation of the Psychometric Quality of Measures31				
11.6.2 Validity 27 12 Determining Articles for Psychometric Grading 29 12.1 Group 1: Measures Cited Within One Article 29 12.2 Group 2: Measures Cited More than Once with only Reliability Information 29 13 Group 3 Measures 30 13.1 Content coverage/subject Areas of Measures 30 13.2 Attitude Construct Coverage of Measures 31 13.3 Evaluation of the Psychometric Quality of Measures 31 Discussion 34				
11.6.2 Validity 27 12 Determining Articles for Psychometric Grading 29 12.1 Group 1: Measures Cited Within One Article 29 12.2 Group 2: Measures Cited More than Once with only Reliability Information 29 13 Group 3 Measures 30 13.1 Content coverage/subject Areas of Measures 30 13.2 Attitude Construct Coverage of Measures 31 13.3 Evaluation of the Psychometric Quality of Measures 31 Discussion 34 14 STEM Content Coverage 35				
11.6.2 Validity2712 Determining Articles for Psychometric Grading2912.1Group 1: Measures Cited Within One Article2912.2Group 2: Measures Cited More than Once with only Reliability Information2913 Group 3 Measures3013.1Content coverage/subject Areas of Measures3013.2Attitude Construct Coverage of Measures3113.3Evaluation of the Psychometric Quality of Measures31Discussion3414 STEM Content Coverage3515 Attitude Construct Coverage36				
11.6.2 Validity 27 12 Determining Articles for Psychometric Grading 29 12.1 Group 1: Measures Cited Within One Article 29 12.2 Group 2: Measures Cited More than Once with only Reliability Information 29 13 Group 3 Measures 30 13.1 Content coverage/subject Areas of Measures 30 13.2 Attitude Construct Coverage of Measures 31 13.3 Evaluation of the Psychometric Quality of Measures 31 Discussion 34 14 STEM Content Coverage 36 15 Attitude Construct Coverage 36 16 Evaluation of the Psychometric Quality of Measures 37				

18 Adapted Measures	42
19 Limitations of the Psychometric Grading Framework	44
Limitations	46
Implications	48
References	63
Curriculum Vitae	93

List of Tables

Table 1. Search Terms Used to Select Applicable Research Articles	3
Table 2. Inclusion Criteria for Title/Abstract and Full Text Review 5	4
Table 3. Overview of Attitude Construct Coverage within Measures 5	5
Table 4. Group 3 Instrument Characteristics (n= 15)	6
Table 5. Reliability, validity, and psychometric strength of STEM affinity instruments5	8
Table 6. Recommended measures given participant population and target attitude construct	
	1

List of Figures

List of Appendices

Appendix A. Articles Included in Review	
---	--

Introduction

1 Rationale

In the 20th century, major advancements were made in the fields of technology, science, and engineering leading to the number of jobs within these fields to increase three times faster than in any other field (Friday Institute for Educational Innovation, 2012). Despite the demand for science, technology, engineering, and mathematics (STEM) competent employees, there has been a decrease in student enrollment in post-secondary STEM courses (Kennedy et al., 2016) resulting in challenges finding qualified employees (Friday Institute for Educational Innovation, 2012). The likelihood that students will enroll in STEM courses and participate in the STEM workforce depends, in part, on their attitudes toward STEM and their interest in STEM careers (Kennedy et al., 2016). Students in elementary school demonstrate high interest in STEM subjects such as mathematics and science, however, attitudes towards STEM begin to decline in adolescence (Frenzel et al., 2010; Savelsbergh et al., 2016).

As a result, researchers and educators have developed programs for K-12 students promoting positive attitudes towards STEM and STEM careers. For example, Project Hope was a school-based intervention program, developed in the United States, to introduce and promote science and health care related occupations to students (Ali et al., 2017). This program showed promising results, with participating students displaying an increase in science and math self-efficacy beliefs from pre to post assessment. Other intervention programs can take place in more informal learning environments, such as a summer camp program. The Summer Ventures in Science and Mathematics (SVSM) program was a four-week program developed for high-school students, and yielded effective results showing an increase in STEM attitudes of participants from pre to post assessment as well (Binns et al., 2016).

Accurately measuring the impact of these STEM programs on student attitudes is critical for making decisions relevant to the design and implementation of future programs. Despite this, there is a lack of comprehensive, validated, STEM-related measurement instruments. The main objectives of this review are to: (1) Evaluate the content coverage/subject areas of instruments assessing STEM attitudes, (2) Evaluate the attitude construct coverage of instruments assessing STEM attitudes using a hierarchical framework, and (3) Evaluate the psychometric quality of instruments assessing STEM attitudes for K-12 students.

2 What is STEM?

STEM is an acronym for the subject areas of science, technology, engineering, and mathematics (McComas & Burgin, 2020). The term was first introduced in 2001, by biologist and former director of the U.S. National Science Foundation, Judith Ramaley (Teaching Institute for Excellence in STEM, 2010). However, even though the term STEM was not introduced until the 21st century, science and technology education was emphasized in Western civilization for the first time in the 1950s as a response to Soviet technological advances, such as the Sputnik satellite (Mohr-Schroeder et al., 2015). It was believed that more needed to be done to train the next generation of scientists and engineers to compete with the technological advances of other nations. In this same decade, the National Aeronautics and Space Administration (NASA) was formed and served as a leader in the push for greater science, technology, engineering, and mathematics awareness and education (Mohr-Schroeder et al., 2015). Several published international reports in the early 2000s, such as the Program for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS), shed light on the need for US and Canadian students to increase their competence in STEM related fields due to them trailing behind competitor countries (Mohr-Schroeder et al., 2015).

In the years following, national organizations, such as the Natural Sciences and Engineering Research Council of Canada (2010) and the National Science Foundation (2008) in the United States, have encouraged the creation and implementation of initiatives to promote STEM interests and skills within students. Since, the 1950's there has been a slight increase in the number of post-secondary graduates obtaining a degree in a STEM-related program. However, there is a gap between the number of graduates and the number of individuals entering the STEM workforce. This gap remains as there are still more STEM jobs than qualified professionals to fill these positions.

Since its conception, the definition of what encompasses the field of STEM has become unclear (Marrero et al., 2014). For many, STEM refers to the instruction and professions within mathematics and the hard sciences, while others include social sciences such as psychology, sociology, and political science (Marrero et al., 2014; Green, 2007). Koonce and colleagues (2011) compared over fifty definitions of STEM from both education and professional organizations (Koonce et al., 2011). The most commonly included subject areas were chemistry, computer science, biological sciences, mathematics, physics, geometric analysis and engineering disciplines, mapping on directly to the traditional topics of science, technology, engineering, and math. Most researchers did not include agricultural studies, psychology, or social sciences within their definitions of STEM (Koonce et al., 2011). Therefore, for the purposes of the current review, STEM will be defined by the four core disciplines of science (i.e., biology, chemistry, physics, earth science and astronomy), technology, engineering, and mathematics and not include psychology and other social sciences.

More specifically, science encompasses the field of study that attempts to explore the structure and actions of the physical and natural world that surrounds us through observation and systematic experimentation (Yata et al., 2020). The field of science can be further subdivided into the physical (e.g., physics chemistry, astronomy, and earth science) and biological sciences (e.g., biology and medicine). Technology is the application of scientific knowledge for practical purposes such as the advancement of industry (McComas & Burgin, 2020). Within the field of technology, machines and equipment are developed to enhance the human environment (Yata et al., 2020). Engineering is the application of science, technology, and mathematics to design and build machines, engines, and structures to advance mankind. (Yata et al., 2020). Finally, mathematics is defined as the study of structure, order, and relation that has evolved from the basic practices of counting and measuring (McComas & Burgin, 2020). Mathematics can be studied independently (pure mathematics) or it may be applied to a variety of other disciplines such as technology, engineering or physics (applied mathematics) (Yata et al., 2020).

All four core disciplines within STEM can be considered independently, however are frequently utilized and studied simultaneously in today's technologically advanced society. This causes further confusion regarding the definition of STEM and what should encompass STEM-education due to the disparity between professional opinions on the independence and/or integration of the four core subject areas. Additionally, there is some confusion regarding how many of the four core subject areas need to be integrated to be considered a unified approach to STEM. Some professionals state the amalgamation of two core subject areas is sufficient, while others argue several, if not all core disciplines, should be combined for an integrative approach (McComas & Burgin, 2020). McComas and Burgin (2020) have proposed that the independent instruction of any of the four-core subject areas is to be considered STEM education, while the partial (two) or full integration of disciplines within instruction should be referred to as I-STEM (McComas & Burgin, 2020). This distinction helps to clarify the definition of STEM and allows for the term to be more than just a slogan to identify science, technology, engineering, and/or mathematics, which has been the case since the term's inception. As the current researchers take an integrative approach to understanding STEM attitudes and education, when referring to STEM we are following McComas and Burgin's I-STEM definition.

3 STEM Attitudes and Social Cognitive Career Theory

Extensive research effort has focused on understanding internal, social, and educational factors that influence a student's entry into a STEM post-secondary program and their subsequent career choices (Wang, 2013). There is extensive literature supporting the role of positive student attitudes and perceived competence in STEM subjects on an individual's behaviour and academic outcomes (Akey, 2006). Additionally, STEM attitudes and self-efficacy beliefs have shown to impact student's intentions to pursue a STEM-related post-secondary degree (van Aalderen-Smeets et al., 2019).

STEM attitudes are often ambiguously or poorly defined, leading to criticism from experts (Blalock et al., 2008). In general, they consist of affective, cognitive, and behavioural components and positive or negative dispositions toward STEM education overall as well as specific subject areas (Potvin & Hasni, 2014). For some researchers, STEM attitudes consist of multiple subconstructs, while others use only a single factor to define their construct (Kennedy et al., 2016). For example, some definitions are very specific (e.g., capturing only an individual's career aspirations) while others combine attitudes, interest, and motivation to form a general and broad construct (Potvin & Hasni, 2014).

Savelsbergh and colleagues (2016) have attempted to address these discrepancies by creating a multidimensional framework of STEM attitudes. The framework includes four general ways of thinking and/or feeling about STEM including relevance, interest, self-efficacy and one's perception of the normality and usefulness of scientists within the community. Relevance is the perceived importance of the field of STEM to one personally or society (Savelsbergh et al., 2016). Interest is a positive or negative emotional response towards STEM subjects, which results from exposure and experience with the subject areas within a classroom or informal/ leisure activities (Savelsbergh et al., 2016). Interests can be further subcategorized as affinity towards STEM within the classroom, towards STEM leisure activities, and/or towards future careers within a STEM-related field (Hidi & Renninger, 2006). Self-efficacy is an individual's belief of their abilities to complete tasks and display behaviours to produce desired outcomes, such as succeeding within a STEM classroom or job (Bandura, 1977). Subjective norms of scientists describe the perceptions of whether society supports and/or values scientists as a profession within the community (Savelsbergh et al., 2016). Savelsbergh et al. (2016)'s multidimensional model of STEM attitudes is based on extensive review of the literature and incorporates multiple definitions of attitudes present, making it an acceptable model for use in the current review.

Researchers have highlighted that the decision to pursue STEM related coursework or career is partially influenced by attitudes toward STEM (Wang, 2013). Building on Bandura's (1986) social cognitive theory, social cognitive career theory (SCCT) describes the variables (environmental, individual, and behavioural) that affect one's academic development and later career choice (Lent et al., 1994; Lent & Brown, 2006). This theory illustrates the dynamic interplay between an individual's environment and their subsequent behaviour and beliefs. It has been applied to understand interest and career choice in STEM. According to SCCT, self-efficacy, or the belief that an individual is capable of mastering events, is the most influential cognitive component of goal setting. Outcome expectations also affect interest. For example, if a student believes they have the ability to get good grades in a STEM class, they have high self-efficacy. If the student feels their grades in STEM will make their parents proud (i.e. outcome expectations), they may study harder. The resultant success within STEM will then also increase future interest in the area.

Studies have used social cognitive career theory as a predictive model for interest in STEM fields (Fouad & Santana, 2017). For example, Wang (2013) used the SCCT framework to examine which factors affected students' attitudes towards entering a STEM field and actual entrance into STEM programs by post-secondary students. Wang (2013) reported that math attitudes in early high school were strong predictors of later math self-efficacy beliefs, achievement, as well as exposure to math and science courses. Further, students' positive attitudes towards STEM were a better predictor of later enrollment than achievement in math and science courses. The link between early attitudes and later career choice was also examined; entrance into a STEM field in post-secondary education increased due to the relationship between early exposure to math and science courses in high school and a student's subsequent intent to major in a STEM field.

Similarly, Nugent and colleagues (2015) found that STEM interests in adolescents were a strong predictor of career orientation and achievement through its direct effect on self-efficacy and career expectancy. Chachashvili-Bolotin's and colleagues (2016) identified a positive relationship between students' interest in STEM, subsequent math achievement and later enrollment in advanced STEM courses. Similarly, Inda-Caro and collaborators (2016) conducted a path analysis to determine SCCT's validity within an international sample of Spanish high-school students. Researchers found that student's technological self-efficacy and attitudes within high-school determined outcome expectations such as interest towards pursuing a career within the field of technology (Inda-Caro et al., 2016). These results support SCCT theory and highlight the importance of intrinsic motivational factors, such as attitudes and interest in predicting STEM engagement and later entrance into a STEM-related field. These research studies utilize surveys, such as the S-STEM (Zhou et al., 2019), to measure aspects of STEM attitudes and predict outcomes. Self-report tools are commonly used due to their ease of

administration and ability to track changes in a respondent's attitudes between preintervention and post-assessment.

Criticisms of this theory point to the limited environmental components considered within social cognitive career theory, which are primarily occurring within the classroom or informal learning context. Other important environmental influences (e.g., parental opinions, gender stereotypes, cultural norms, financial instability) are critical components that affect an individual and their career-related behaviour, however, are not incorporated within SCCT (Ambriz, 2016). For example, a key barrier to minorities obtaining a career in a STEM-related field has been termed the "leaky STEM pipeline" (Sheltzer & Smith, 2014). The STEM pipeline metaphor describes the educational pathway that begins in early secondary education and progresses to post-secondary graduation and the workforce (Blickenstaff, 2005). Most minority or female students enrolled in a STEM field within college or university switch to a non-STEM major or drop out during undergraduate education, with the metaphor of a leaky pipeline describing this attrition. Leaks have been identified at key career stages including the bachelors to PhD stage, employment selection (Morgan et al., 2013), promotion (Ong et al., 2011) and retention. Factors contributing to the attrition of minority individuals include experiences with microaggressions, incivility and ostracism within the field (Carter-Sowell & Zimmerman, 2015). Therefore, there are multiple contributing factors that determine whether an individual, especially women of colour, will enter and pursue a STEM-related career that goes beyond initial experiences within the classroom.

4 Measuring STEM Attitudes

A multitude of instruments have been developed to assess attitudes towards science, math, and technology for use in both descriptive and intervention studies (Blalock et al., 2008; Yáñez-Marquina & Villardón-Gallego, 2016). For example, Chachashvili-Bolotin's and colleagues (2016) used a single tool, the Attitudes about Academic Education Questionnaire, within their research design, which was a self-report survey aimed to evaluate student's attitudes towards future enrollment in STEM courses and achievement. In contrast, Inda-Caro and colleagues (2016) combined and adapted four independent survey tools (Technology Grade Self-Efficacy Scale, Sources of Technology Self-Efficacy Scale, Technology Interests Scale, and Technology Intentions and Goals Scale) within their research study to target specific outcome variables.

When assessing STEM attitudes, researchers and educators need measures that address multiple core subject areas, have a clear STEM attitude definition, and are reliable and valid. Unfortunately, many measures fail to do this by only targeting one or two core subject areas, lacking an integrated view of STEM. Most commonly, measures target science and/or mathematics. This is congruent with approaches to education where a single subject is taught at a time, rather than integrating knowledge across seemingly disparate subjects (National Academy of Engineering and National Research Council, 2014). As STEM education moves towards a more holistic discipline, measurement tools designed to evaluate attitudes must follow suit.

The construct of STEM attitudes is typically defined and operationalized as an interest towards the core subject areas (Staus et al., 2020) or attitudes towards pursuing a career in a STEM-related field (Mau et al., 2019). It is common for researchers to define

a construct based on specific research questions or theories used within a particular study. If researchers define attitudes differently based on the common themes within the literature, they may be creating instruments that are all capturing different aspects or definitions of STEM attitudes. This issue results in difficulty when evaluating the appropriateness of a scale as well as when comparing the effects of intervention across multiple studies. Therefore, a review of STEM attitude measures that evaluates and compares reported operational definitions following a validated framework will shine light onto what subconstructs are overly represented within the literature.

Ensuring the reliability and accuracy of attitude measures is imperative for drawing conclusions about the relationships between interventions and STEM attitude outcomes. Many authors create STEM attitude measures that serve to answer specific questions for their individual studies. When these new independent measures are created, little data are reported on the instrument's reliability and validity (Blalock et al., 2008). Without this information, readers must be cautious when interpreting results. If unsure of a measure's reliability and validity, it draws into question whether it can accurately track changes in attitudes as a result of the implementation of an intervention. If the measure is invalid, then the observed relationship between variables (e.g. intervention and STEM attitudes) may also be invalid, thereby undermining the results (Blalock et al., 2008). It is crucial to ensure all measurement tools used in a research project have adequate psychometric properties to ensure that accurate and reliable relationships between independent and dependent variables are captured and that the subsequent conclusions are valid. The goal of this review, therefore, is to identify STEM attitude measures and their psychometric quality, supporting researchers using these instruments as primary outcome measures for intervention programs looking to change attitudes in students.

Previous reviews of science attitude measures have identified limitations such as minimal subject/content coverage, insufficient reported psychometrics, as well as inconsistent construct definitions of attitudes across measures. Blalock et al. (2008) reviewed existing science attitude instruments and identified a multitude of measures reporting minimal psychometric information. Prior, the only large-scale review of STEM attitude measures was conducted by Hugh Munby in 1983, who also only evaluated science attitude measures. Blalock and colleagues (2008) identified 66 tools, evaluating each based on published psychometric data using a rubric created by the authors. The rubric was used to assign scores to a total of five content areas of importance for each tool; theory, reliability, validity, dimensionality, as well as development and usage. Of the 66 total measures, 28 were missing fundamental psychometric data; either reliability or validity or in some cases both. Additionally, a total of 37 measures were only used in a single study with no follow-up information, which demonstrates researchers' preference for creating their own science attitude tools. Only a minority of measures (e.g. the Test of Science Related Attitudes and the Scientific Attitude Inventory) were associated with multiple studies and extensive psychometric data. Blalock and colleagues identified a major lack of reliable and validated measures of science attitudes to be used within the scientific literature. Blalock et al. (2008) suggested researchers stop the creation of new instruments and promote the validation of promising existing measures.

More recently, Kennedy and colleagues (2016) conducted a review of published measures of science attitudes, aiming to identify the subconstructs within overall attitudes

commonly utilized by researchers as well as creating a new web-based instrument. Researchers examined measurement items for common themes relevant to the formation of positive attitudes towards science. Identified themes included student's perceived enjoyableness of the topic, believed difficulty of school science, self-efficacy, the relevance of science to a student's everyday life, as well as perceived usefulness of science to one's future career. Their evaluation of identified measures was limited to qualitative theme analysis and did not use quantitative psychometric information. Additionally, Kennedy et al (2016) only included measures addressing school science attitudes. Unlike Blalock and colleagues, Kennedy did not include psychometric data to aid in their evaluation of identified science attitude measures.

5 Assessing Quality of Instruments

Building on the work of Blalock and colleagues (2008), this study will assess the psychometric data of identified measures to determine the measurement precision, reliability, of a specific tool (Smelser & Baltes, 2001). Reliability addresses a psychological tests ability to consistently produce similar outcomes when used on the same individual at multiple time points in different settings (Smelser & Baltes, 2001). Therefore, reliability is measured based on the consistency between two sets of scores as well as the amount of true and observed variance within each set (Smelser & Baltes, 2001). Common statistical tests of reliability include test-retest reliability, internal consistency, and interrater reliability (Leung et al., 2012). Test-retest reliability requires the administration of a particular measure to the same group of individuals at two-time points. If the construct measured by a tool is stable in an individual, the differences in scores at two-time points are due to unsystematic error (Leung et al., 2012). Statistical

analysis used to describe test-retest reliability include Kappa's coefficient, probability values and Pearson correlations (Leung et al., 2012). Internal consistency addresses correlations between items within a measure to ensure all items are addressing similar constructs if intended to. If a measure contains multiple subconstructs, internal consistency will ensure items meant to assess different variables do not positively correlate to one another (Leung et al., 2012). Spearman-Brown, Kuder-Richardson, and/or Coefficient alphas can be used to measure internal consistency (Leung et al., 2012). Finally, interrater reliability measurements inquire about the agreement between two independent data collectors administering the same tool and can be assessed using various statistic tests including Kappa's coefficient and Pearson correlations (Leung et al., 2012; McHugh, 2012).

To determine if a tool is accurately assessing the intended construct, validity measurements are used (Smelser & Baltes, 2001). A psychological instrument is considered to be valid only if it measures what it is supposed to be measuring (Smelser & Baltes, 2001). For example, a STEM attitude measure is, in fact, measuring a student's attitudes towards the field of STEM. Similar to reliability, there are multiple different methods for evaluating the validity of a measure including; content, construct, and criterion validity (Leung et al., 2012). Content validity provides information on the representation of target constructs within questions, items, and/or tasks in an assessment (McDowell, 2006). A common method of evaluating content validity is through the judgment of an expert panel and/or content relevance ratings yielding a content validity index score (Leung et al., 2012). Construct validity addresses how well a psychological tool correlates with predictions from the theoretical model that was used to create the

measure (McDowell, 2006). Construct validity can be evaluated using convergent and/or discriminant evidence (Leung et al., 2012). Common statistical tests that are used to uncover convergent and discriminant validity include sample t-tests, analysis of variance (ANOVA), Pearson correlations, and/or factor analysis (Leung et al., 2012). Finally, criterion validity concerns itself with how well a measure is an accurate predictor of a particular criterion, either at the same time of administration (e.g. how well IQ scales predict current academic success in school) or in the future (e.g. how well IQ scales predict admission into competitive post-secondary institutions). These two subcategories of criterion validity are referred to as concurrent and predictive validity (Leung et al., 2012). Statistical analysis that can evaluate criterion validity include t-tests, ANOVA's, Pearson correlations, positive and/or negative likelihood rations, as well as probability values (Leung et al., 2012).

For the current review of STEM attitude measures, authors used the psychometric grading framework (PGF) to evaluate the strength of published tools (Leung et al., 2012). This framework has proven its usefulness while evaluating attitude and belief scales in health care settings such as hospitals and clinics (Leung et al., 2012). The psychometric grading framework uses common psychometric data to assign a grade representative to the strength of each published measure allowing for their comparison (Leung et al., 2012). Since Blalock and colleagues (2008) review, there has been an increase in the understanding by researchers on the importance of reporting reliability and validity data, hopefully leading to increases in the publication of psychometric data. Therefore, having the ability to utilize this data and evaluate the strength of a measure, will allow for a reliable comparison of a tool's usefulness based on psychometric data. Using the PGF

will allow for the current review to build on the results of previous reviews and provide recommendations of the most reliable and valid measures to be used in future research studies. Additionally, considering attitude construct and STEM subject coverage of all published tools may provide insight into which measures are appropriate for particular research objectives.

Method

Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) standards are followed in this report.

6 Search Strategy

In June 2020, the following databases were searched: ERIC, PsycINFO, Web of Science, SCOPUS. Publication year was not restricted, but the search was restricted to peer-reviewed articles. Search terms are listed in Table 1. All terms were searched within the title and topic areas in each database. Each search term in each group listed in Table 1 was paired with every word from every other group by the word *and*. The search was limited to publications in English. The reference lists of selected studies were also hand-searched to identify original sources of included measures and were subsequently added directly to full-text screening. Each measure evaluated within included articles was subsequently obtained from their original source.

7 Inclusion and Exclusion Criteria

Inclusion and exclusion criteria used for this review is displayed in Table 2. The titles, abstracts and full texts of each article were evaluated against the inclusion criteria.

8 Title/Abstract Review

All references were exported to Covidence where duplicates were removed. Initially, two authors (NN, MK) screened a randomly selected portion (20%) of titles and abstracts against the inclusion/exclusion criteria and achieved 85% agreement; any discrepancies were then discussed until consensus was reached and criteria was subsequently refined for clarity. Authors then independently screened all titles and abstracts, and any disagreements were discussed until consensus. If there was insufficient information in the title or abstract to evaluate all inclusion and exclusion criteria, the full text of that document was reviewed.

9 Full-Text Review

The full text of all articles considered for inclusion was independently reviewed by researchers (NN, MK, KL) against the inclusion criteria. Each article was reviewed by two researchers and disagreements on inclusion or exclusion were resolved via discussion until reaching consensus. Following full-text review, included articles were retained for data extraction and psychometric grading.

10 Variable Coding/Data Extraction

Following inclusion, data on study information (e.g., participant population, psychometric data) was extracted from each article. Data on survey content (e.g., number of items) was extracted directly from copies of the measure. Data extraction forms were developed by authors and piloted on a portion (20%) of randomly selected studies by researchers (MK, KL). Two forms were created, one to gather information on the research study and one for the identified measure. Specific content areas selected to be extracted were modeled after Blalock et al. (2008). The following variables were extracted from each research article: participant sampling method used, total number of participants, administration method of measure, and reported psychometric data. Additionally, the following variables were extracted for each identified measure: target age group, total number of items, format of measure, time to complete measure, STEM

subjects addressed, and reported attitude constructs. Following data extraction, the psychometric grading framework was used to evaluate the quality of each measure using collected psychometric data.

10.1 Target Age Group

Information regarding a measure's intended age group was obtained from the measure itself. Each measure's target population was classified into one of three groups: grades 1-6 (early elementary), grades 7-9 (middle-school), grades 9-12 (high-school). If data on a measure's target age group could not be found within the tool's description or title, its original source was searched.

10.2 Number of Items

The reported number of total items of a given measure was collected either from a direct statement from researchers found within the methods section of the article or by counting the total number of items within the measure. Data extractors counted also the number of items within a measure that targeted the subject areas of science, technology, engineering, and/or mathematics.

10.3 Format of Measure

Only self-report measures were included for consideration and were categorized as either a Likert scale or semantic differential. Likert scale items typically use 3-, 5- or 7-point scales to allow a respondent to report how much they agree or disagree with a particular statement (Likert, 1932). Semantic differential items ask participants to rate a particular statement, object, or event based on contrasting adjectives (e.g., good-bad, pleasant-unpleasant etc.; Heise, 1969). Data on the format of a measure was extracted by reviewing items on the measure or via an explicit statement in the methodology of the research article including said measure.

10.4 Completion Time

Data on the time to complete a measure was extracted from statements within the methodology of included articles.

10.5 Administration

Data on means of administration of a particular measure was extracted when explicitly stated within an article. There are two main methods of delivery for self-report measures, pencil and paper or computer administration. The administration method of a given measure is chosen by individual researchers to meet their own research goals and restrictions. Therefore, a single measure may be administered via different methods across research studies. Data on all possible administration techniques was extracted from included studies within the current review.

10.6 STEM Subject Area Coverage & Categorization

To evaluate the content coverage/subject areas, individual items on each measure were categorized for targeting core subject areas of STEM (science, technology, engineering, and/or mathematics). If a measure contained at least one item addressing a particular STEM subject, it was classified as such. Subject areas (e.g., science, technology, engineering, and mathematics, or general STEM) were typically explicitly identified by authors and therefore can be reliably classified. If the subject area was not clearly defined by creators, authors of the current review discussed what content was included for each individual item on the measure to determine all subject areas covered. If items assessed multiple subject areas within STEM, all were noted. For example, the Self-Efficacy Scale Towards Science and Technology contains items such as "I can accomplish science and technology projects successfully" which addresses both science and technology. If an item addresses all four core STEM subject areas, or refers to STEM generally (e.g., Science, technology, engineering, and mathematics are very important in life; Unfried, 2015) the item was classified under general STEM. All individual items were used to classify each measure.

Similarly, a single measure may address multiple STEM fields and therefore was included in multiple categories (e.g., the STEM Career Interest Scale addresses all four subject areas of science, technology, engineering, and mathematics). As many previous reviews have not addressed subject areas within STEM attitude measures, it is important to identify which content areas (e.g. mathematics and/or science) are overly represented and others that are not adequately addressed within tools.

10.7 Attitude Construct Coverage

Each identified measure was categorized according to Savelsbergh and colleagues' (2016) hierarchical framework of attitudes as assessing relevance (personal and societal), interest (classroom, leisure, career), self-efficacy, and/or subjective norms/normality of scientists. Constructs were defined using Savelsbergh's (2016) model. Relevance was defined as perceived importance of the field of STEM to one personally or society. Interest addresses a positive or negative emotional response towards STEM taught within one's classroom, activities completed during leisure time, or future careers in STEM. Self-efficacy focuses on an individual's belief of their abilities to complete tasks and display behaviours to produce desired outcomes within STEM.

Finally, subjective norms of scientists describe perceptions of whether society supports and/or values scientists as a profession within the community (Savelsbergh et al., 2016). Items were categorized as assessing one of these constructs and the total number of items addressing each construct was tallied. All reported construct/s were recorded and assessed within individual items and used to categorize the overall measure.

10.8 Quality Appraisal: Psychometric Grading Framework

Identified measures were evaluated using the psychometric grading framework (PGF) created by Leung and colleagues in 2012. The PGF consists of two scales: Scale 1 is a matrix for assigning a level (A–D) to six psychometric properties (content validity, construct validity, criterion validity, internal consistency, test-retest reliability, and interrater reliability) based on the strength of any measures reported. Levels were only assigned for reported psychometric properties. All psychometric information was extracted and individually assigned a grade based on guidelines outlined within the PGF (Leung et al., 2012). Grades (A-D) were assigned to reported psychometric co-efficients based on their comparison to pre-determined criteria. Once all co-efficients were assigned a grade based on strength, the overall measure can be evaluated, using Scale 2, by combining all individual grades. Scale 2 grades the overall psychometric strength of the instrument (Good – Very weak) by combining the number and level of psychometric measures arising from Scale 1.

To determine which tools to evaluate, all identified measures were categorized into one of three groups; (1) Measures with psychometric data from one article; (2) Measures with psychometric data from more than one article with only reliability information reported, and (3) Measures with psychometric data from more than one article with reliability and validity information reported. Only measures in group 3 were evaluated using the psychometric grading framework. This categorization aided in identifying measures that have been validated across multiple participant samples and have enough published psychometric data to be accurately evaluated using PGF.

Results

A total of 144 relevant peer reviewed articles were identified for data extraction. Within these 144 articles, 104 STEM affinity measures were identified and evaluated. Peer-reviewed articles that were included for data extraction, as well as the number of publications excluded is displayed in Figure 1.

11 All Identified STEM Attitude Measures

11.1 Completion Time

Completion time was reported for a total of 16 measures (15%), ranging from 10 minutes (STEM Career Interest Scale; Kier et al., 2014) to 50 minutes (ROSE - Science Interest Scale; Schreiner and Sjøberg 2004).

11.2 Administration

Included measures were administered via paper-and-pencil (n = 46, 32%) or computer tasks (n = 22, 13%). Unfortunately, many (n = 76, 53%) articles did not explicitly state the mode of administration used within their research protocols, and therefore no data was extracted.

11.3 Number of Items

Measures varied in length from containing as few as 3 items (Ability Self-Concept in Physics and Chemistry Questionnaire; Wang et al., 2017) to 338 items (Math, Science, and Technology Questionnaire; Rice et al., 2013). However, the majority (n = 87, 83%) of included measures contained fewer than 100 items.

11.4 Content Coverage/Subject Areas of Measures

The most common subject area addressed within all identified measures was science (n = 89; 86%), followed by mathematics (n = 57; 55%), technology (n = 49; 47%), engineering (n = 34; 33%), and general STEM (n = 13; 13%). Most measures targeted only two fields (n = 66; 63%), with fewer targeting three (n = 4; 4%) or all four STEM fields (n = 34; 33%).

11.5 Attitude Construct Coverage

The most commonly addressed subconstruct was Interest, present within 57 measures (45%), followed by Self-Efficacy (n = 31, 29%), Attitude (n = 32; 31%), Relevance (n = 1; <1%) and Subjective Normality of Scientists (n = 1, <1%). Most measures focused on only one construct area (n = 86, 83%), while others addressed two (n = 17, 16%) or three constructs (n = 1, <1%). There were no included measures that assessed more than three constructs. Table 3 displays attitude construct coverage amongst extracted measures.

11.6 Evaluation of Psychometric Quality

11.6.1 Reliability. There are multiple ways to evaluate the reliability of a given measure. Within the PGF, internal consistency, test-retest reliability, and inter-rater reliability are considered. Across measures reporting reliability

information, internal consistency was the most frequently assessed. Inter-rater and test-retest reliability values were also reported, however less frequently.

- 11.6.1.1 Internal Consistency. The internal consistency of a measure refers to the stability of results between items within a given measure or scale (McDowell 2006). It is most commonly evaluated by an α value. The reliability of measures was most commonly reported as a Cronbach's alpha (α), with values reported for 95 measures (91%). The only other internal consistency metric reported was a Kuder-Richardson 20, for a single measure (Career Interest Survey- CIS; Donovan et al., 1985). Internal-consistency values were provided for instrument sub scales and/or the overall measure. Coefficients ranged in value from .42 .98 for subscales within a measure, and .59 .98 for the entire instrument.
- 11.6.1.2 Test-retest and Inter-rater Reliability. Test-retest reliability is concerned with the stability of a given construct that is not expected to change over time (McDowell 2006). Test-retest reliability coefficients were reported for 6 (5%) included measures as a Pearson correlation (*r*) or a Intraclass correlation coefficient (ICC). Test-retest coefficients ranged from .60 .87 at below marginal to acceptable (McDowell 2006). Inter-rater reliability is a value of agreement between independent raters who are using the same measure to obtain results (McDowell 2006). Inter-rater coefficients were not provided for any included measures, as expected, as all
measures were self-report questionnaires and not behavioural observation tools.

- 11.6.2 Validity. The validity of a given measure can be evaluated via multiple techniques. Within the PGF, content, construct, and criterion validity are considered within the overall evaluation. Unfortunately, many measures (n = 44, 42%) did not report any validity information at all, limiting an evaluation of psychometric properties.
- **11.6.2.1 Content Validity.** Content validity addresses the extent to which all facets of a construct are represented within a given measure (McDowell 2006). Content validity was the most commonly reported, being assessed for a total of 40 measures (38%). The most common technique used to determine content validity included the review of a measure by an expert panel, reported for 36 measures (35%). Many researchers also conducted a literature review (n = 15, 14%) and/or feedback was collected from participants on their understanding of instrument's content (n = 16; 15%).
- **11.6.2.2 Construct Validity**. Construct validity measures the relationship between a given construct and expected predictions based on a theoretical model (McDowell, 2006). The most common measurement of construct validity reported was percent variance (n = 26, 25%) following a confirmatory or exploratory factor analysis. Values for overall scale variance ranged from, poor to good at 36-97%. A few measures (n = 3, 3%) did not calculate overall percentage variance explained, but individual

subscale variance instead. For example, an article outlining the development and validation of the Student Interest and Choice in STEM (SIC-STEM; Roller et al., 2020) Survey, conducted a confirmatory factor analysis for each subject subscale instead of the measure as a whole. However, the psychometric grading framework being utilized within this review does not consider subscale percent variance scores, so these values were not used to evaluate the overall strength of a measure. Kaiser-Meyer-Olkin (KMO) values were the second most common metric of construct validity reported (n = 15, 14%), with values ranging from .68 - .96. KMO coefficients indicate the proportion of variance within a sample that might be the result of underlying factors and are used for determining the suitability of a factor analysis on a given dataset (Leung et al., 2012).

11.6.2.3 Criterion Validity. Criterion validity evaluates how well a measure correlates with a gold standard tool or an indicator of the real situation (McDowell 2006). Criterion validity was not frequently evaluated by researchers, present for only 5% of extracted measures (n = 5). Pearson product correlations evaluating concurrent validity of a measure were the most commonly reported, with extracted values ranging from .42 - .75. These coefficients were calculated by correlating a given measures results (e.g., STEM Career Interest Survey; STEM-CIS) with a pre-validated measure of a related construct (e.g. The Interest in Science Scale, ISS; Bozdogan, 2007).

12 Determining Articles for Psychometric Grading

The 104 included measures were organized into three groups: (1) Measures with psychometric data from one article; (2) Measures with psychometric data from more than one article with only reliability information reported, and (3) Measures with psychometric data from more than one article with reliability and validity information reported. Only measures in group 3 were evaluated using the psychometric grading framework within the current review.

12.1 Group 1: Measures Cited Within One Article

Of the included measures, 83 (80%) were only cited within one included peerreviewed article. Most of these measures were uniquely developed by researchers for the purpose of their individual research goals and questions (e.g., Self-Efficacy in Engineering, Programming, and Circuitry Questionnaire; Nugent et al., 2019). While others contained measures that were adapted from other validated tools by the addition of relevant items and/or replacement of question wording to match the individual study (e.g., Science Motivation Questionnaire II – Adapted; Glynn et al., 2011).

12.2 Group 2: Measures Cited More than Once with only Reliability Information

Multiple measures (n = 6; 6%) were cited in more than one included peer-reviewed article, but articles only presented reliability information. The most common measure of reliability was internal consistency (n = 6, 100%) which was reported for all measures contained within Group 2. The most common internal consistency coefficient reported was Cronbach's alpha, which was published for all six measures within Group 2 (e.g., Attitudes Towards Math and STEM; Prieto et al., 2011).

13 Group 3 Measures

A total of 15 measures (14%) were cited in multiple included peer-review articles and both reliability and validity information was published. These measures and their general characteristics can be found within Table 4. All collected psychometric information for each measure, as well as the final grade can be found within Table 5. The measures were primarily designed to target students in grades 9-12 (n = 7, 47%), also referred to as high-school students, as well as middle-school students in grades 7-8 (n =6,40%). Measures designed for elementary aged students, in grades 4-6, were present, however less common (n = 3, 20%). There were no measures within group 3 that were designed for students in grades 1-3. Additionally, identified measures contained a range in the number of contained items, from 8 (PISA 2006 – Science Interest Scale; OECD, 2009b) to 108 (Relevance of Science Education; Schreiner and Sjøberg 2004) items. A total of 14 (93%) identified measures used a Linkert-style format, with a single tool (STEM Semantics Survey; Tyler-Wood et al., 2010) utilizing a semantic differential format. Within the 15 identified measures, only 5 (33%) explicitly reported a completion time, which ranged from 10 (e.g., STEM Career Interest Scale; Kier et al., 2014) to 50 minutes (Relevance of Science Education; Schreiner and Sjøberg, 2004).

13.1 Content coverage/subject Areas of Measures

Of the 15 measures identified within Group 3, a total of 7 (47%) addressed all four core STEM subject areas (e.g., STEM Attitude Survey; Guzey et al., 2014). The most common subject addressed was science, which was included within almost all identified measures (n = 14, 93%), followed by technology (n = 11, 73%), mathematics (n = 9, 60%), engineering (n = 47%), and general STEM (n = 5, 33%). A few identified measures (n = 4, 27%), further broke down science items to address the specific subjects of biology, chemistry, physics, earth science, and/or astronomy (e.g., Trends in International Mathematics and Science Study 2007; Olson et al., 2008). Information on the specific subject coverage for each identified measure can be found within Table 4.

13.2 Attitude Construct Coverage of Measures

Two identified measures (13%), of the 15 within group 3, included more than one attitude construct, addressing self-efficacy and interests (STEM-CIS; Kier et al., 2014, Measure of STEM Interest in Adolescents; Falk et al., 2016). The remaining thirteen measures (87%) focused on a single STEM attitude construct. The most common construct was interest (n = 7, 47%) followed by general attitudes (n = 6, 40%), and self-efficacy (n = 4, 27%). Relevance and subjective norms of scientists were not common attitude constructs assessed within measures, as they were not addressed within any of the fifteen identified measures. The specific construct coverage of each identified measure can be seen within Table 4.

13.3 Evaluation of the Psychometric Quality of Measures

All 15 identified measures reported reliability and validity data, with extracted coefficients displayed in Table 5. A total of 11 measures within Group 3 (73%) received the highest overall grade of Good. The remaining 4 measures within Group 3 (27%) received the second highest grade of Adequate (Educational and Career Interest Scale; Oh et al., 2013, Mathematics and Technology Attitude Scale; Pierce et al., 2007, Science Self-Efficacy Questionnaire; Smist, 1993, STEM Project Based Learning Questionnaire;

Han, 2017). There were no identified measures within Group 3 that received a lower overall grade of Weak or Very Weak.

Of the 11 identified measures receiving an overall grade of Good, many were evaluated using specific and/or international samples. It is important to utilize measurement tools within population samples that they have been previously validated with, to ensure accuracy of obtained results. Both the Science and Technology Attitude Scale (Nuhoğlu, 2008) and the Self-Efficacy for Science and Technology (Tatar et al., 2009) measure were validated using middle-school aged participants within the country of Turkey. As there is no other information regarding the validity of these measures on an international or western sample, it would only be appropriate to utilize these tools on a similar sample population. In contrast, PISA's Science Interest Scale (OECD, 2009b) as well as the Trends in International Mathematics and Science Study (Olson et al., 2008) measures have had psychometric evaluations conducted with international samples of thousands of students and have been translated into several different languages. It would be appropriate for the use of these measures world-wide. However, it is important to note that the Science Interest Scale (OECD, 2009b), was developed for use with 15-year-old participants, and has not been evaluated within other age groups. This age restriction limits the measures use to same age participants.

The STEM Career Interest Survey (Kier et al., 2014), STEM Semantics Survey (Tyler-Wood et al., 2010), STEM Attitude Survey (Guzey et al., 2014), and the Upper-Elementary S-STEM (Unfried, 2015) have been validated in several North American (e.g., United States of America and Canada) and Asian countries (e.g., China, Taiwan, Malaysia, and Turkey). These measures can be appropriately utilized within these communities and administrators can be confident on the validity of obtained results. The Measure of STEM Interest in Adolescence (Falk et al., 2016) and the Middle/High School S-STEM (Unfried, 2015) have only been validated on communities solely within the United States (US). Due to similarities between the US and Canadian populations, these measures can be utilized within both neighboring countries. However, they cannot be reliably used within international samples until further psychometric evaluation is conducted. Finally, the Relevance of Science Education Measure (Schreiner and Sjøberg 2004) has been validated within Asian (e.g., Taiwan and China) and European countries (e.g., Finland) and can accurately evaluate the STEM attitudes of students within these communities. Therefore, if used within North American regions, obtained results may not be valid.

Discussion

The current review builds on previous work evaluating STEM attitude measurement tools used with elementary and/or high school aged students. Using a holistic view of STEM, we searched for tools that assessed more than one core subject area. Savelsbergh and colleague's (2016) hierarchical framework of attitudes was used to categorize identified measures based on the attitude construct definition used. Additionally, using a validated framework from Leung and colleagues (2012), published psychometric data was evaluated to yield a grade of relative strength for identified measures within our review. This allowing the current researchers to identify measures based on reliability and validity data.

The current review yielded a total of 15 unique measures that were cited within multiple included peer-reviewed studies and contained both reliability and validity psychometric data. Many of the identified measures took a holistic approach to evaluating STEM attitudes as they targeted all four core STEM subject areas. Additionally, with regards to attitude construct coverage within measures, many used a global perspective addressing general attitudes. While others used a more tailored perspective by evaluating student's interests and/or self-efficacy beliefs to conceptualize attitudes. Within the 15 identified measures within Group 3, 11 received the highest overall psychometric grade of Good, with the remaining 4 measures awarded a grade of Adequate. Within and across these measures there was a variety of STEM subject area and attitude construct coverage.

Studies commonly evaluated the attitudes of middle and high-school aged participants, with limited focus on the attitudes of younger elementary aged students.

This may be due to older student's experiences with STEM subjects and a better ability to understand and voice opinions on attitudes towards science, technology, engineering, and mathematics (Lesseig, Slavit, Nelson, 2017).

14 STEM Content Coverage

The results of our review highlighted the abundance of science attitude measures within the literature. Many measures targeted two or more STEM subject areas, with science present across almost all measures (89%). Measures addressing two STEM subjects primarily included science and mathematics or science and technology (e.g., Science and Technology Attitude Scale; Nuhoğlu, 2008). Many measures divided the subject of science into subtopics of biology, chemistry, and physics. It is likely that science is the most common STEM subject addressed within measures due to its importance within elementary and high-school curriculum (Herschbach, 2011). Additionally, science may be the most common core subject addressed due to the age of targeted respondents and the relative newness of the concept of STEM. All four core STEM subject areas are well known within post-secondary education, however younger students do not have as much exposure to engineering and technology courses, leading to little need for evaluating student attitudes in said subjects, and a lack of tools to do so.

Of the 15 measures included in Group 3, seven were published between the years 1990 – 2010. Of these seven, only one measure addressed all four core subject areas of STEM (STEM Semantics Survey; Tyler-Wood et al., 2010), with the remaining six covering technology, mathematics, and/or science (e.g., Science Self-Efficacy Questionnaire; Smist, 1993). The term STEM was not coined until 2001 by Judith Ramaley, and not widely adopted until years later, this may explain the lack of focus on attitudes towards technology, mathematics, and engineering in earlier literature (Teaching Institute for Excellence in STEM, 2010). The other eight identified measures included for evaluation within Group 3 were published between 2011 – 2017, with all addressing all four subject areas and/or use of the term STEM within items. Within the past decade, the integration and perceived importance of science, technology, engineering, and mathematics has grown substantially, as indicated through the published attitude scales of this time period.

15 Attitude Construct Coverage

The most common construct identified was interest followed by general attitudes, and self-efficacy. Relevance and subjective norms of scientists were not assessed within the fifteen group 3 measures. This suggests that current researchers prefer to evaluate attitudes via an individual's interests or feelings of self-efficacy. Both constructs, interest and self-efficacy, can be considered automatic associative affects. Through an individual's repeated exposure to relevant contexts, over time, the affect is paired with the context (Lamb et al., 2015). For example, following multiple positive experiences with STEM subjects (e.g., doing well on a science or math test), positive interests begin to form and become automatically associated with the subject/context. Also, self-efficacy and interest may be the two most common constructs within attitude measures due to their ability to predict behavioural changes (Lamb et al., 2015). These results are consistent with published literature and social cognitive career theory, which highlighted the importance of self-efficacy beliefs and interest for goal setting, achievement, and outcome expectations within STEM (Lent & Brown, 2006, Nugent et al., 2015)

As the attitudes of students are being targeted within intervention programs in order to increase student engagement and persistence within the field of STEM, choosing a construct that is closely associated with behaviours is more likely to show intervention/treatment effects. Relevance and subjective norms are less associated with behavioural change and not as common an antecedent to behaviour as interest and selfefficacy, potentially leading to a decreased number of attitude measures utilizing the constructs (Lamb et al., 2015). Results gathered outlining the specific construct included within a given measure is critical to inform researchers of what they are actually evaluating (e.g. self-efficacy as opposed to general attitudes) and ensure accurate research conclusions are being made.

16 Evaluation of the Psychometric Quality of Measures

Only measures reporting both reliability and validity information across multiple peer-reviewed studies were included for evaluation within the current review. The fifteen identified STEM attitude measures reported diverse psychometric coefficients, most commonly evaluating the internal consistency of a tool with a Cronbach's alpha coefficient and the content validity via literature review and/or expert panel review. These coefficients were used to assign an overall grade of strength to each tool. The STEM attitude measures receiving the highest overall grade of "Good" (e.g., STEM Career Interest Survey; Kier et al., 2014) reported the most psychometric information across multiple research studies supporting the reliability and validity of the measure. Several included measures received a lower grade of "Adequate", indicating there is minimal psychometric information supporting the use of the tools, however, more testing and examination may be necessary.

It is important to note, that the psychometric properties of a tool are influenced by the sample on which it was tested. Reliability and validity evaluations can be affected by the sample size, variability of a construct within and across samples, and limited representation of a population within a chosen sample (Standards for Educational and Psychological Testing, 2014). A small sample size limits the statistical power of a given psychometric calculation as well as limits the representation of a population within said sample. Additionally, limited representation of a particular population (e.g., cultural group, gender identity, specific socioeconomic status, and/or level of education achieved) limits the generalizability to non-included groups. For example, the Middle/High-School S-STEM was tested on a sample of 67 participants located within the southeastern United States (Binns et al., 2016). Participants were academically talented students, with 40% identifying as Caucasian, 30% as Asian, 18% as Black, 2% as Hispanic, and 9% as other. Due to restrictions based on educational background of participants, the sample does not represent the general population of the southeastern United States, which has a cultural composition of 62% Caucasian, 22% Black, 3% Asian, 9% Hispanic, 1% Alaskan Native, and 2% of individuals identifying with two or more racial groups (United States Consensus Bureau, 2019). The reliability and validity of the Middle/High School S-STEM from this sample may vary when used with other groups within the general population. It is important for multiple evaluations on multiple diverse samples to ensure generalizability of a measure across samples.

Standards for Educational and Psychological Testing have been put forth by the American Psychological Association and National Council on Measurement in Education (Standards for Educational and Psychological Testing, 2014). They have outlined the ethical standards and criteria necessary for the accurate interpretation of test scores and guidelines for assessing reliability, validity, and intended use of a measurement tool. More specifically, standard 1.1 has defined the test developer's role to clearly state the intended population(s) a measure was designed for as well as the construct(s) the measure is intended to assess (Standards for Educational and Psychological Testing, 2014). As no measure permits use with all populations in all situations, it is the test developers' duty to distinguish targeted population groups to ensure accuracy of results obtained. Within the included articles, test developers indicated an intended age group (e.g., elementary, middle, or high school aged students) for specific measures. Additionally, studies included in the current review, typically described samples using age, gender, and ethnicity, however, lacked description of other relevant factors such as exclusion criteria, educational background, admissions policies for a particular school, socioeconomic status, disability status, dominate language spoken and linguistic ability within a community (Standards for Educational and Psychological Testing, 2014). Conclusions within the current review will be specific to the provided information within included research articles regarding participant populations targeted.

With respect to validating a measure, researchers aim to show that the tool assess the intended construct it was developed to evaluate. This can be proven within a target population using criterion or construct validity assessments, however as a construct may present differently within different populations of individuals, proving validity within one cannot generalize to another without further testing (Standards for Educational and Psychological Testing, 2014). For example, if a STEM attitude measure was consistently validated using a group of English-speaking Caucasian high-school students from the United States, it may not be able to accurately assess the attitudes of Hispanic middle school females due to different conceptualization and presentation of attitudes between these groups. It is imperative that researchers describe in detail the composition of any sample due to the influence it may have on the validity statistics, which was lacking within identified articles in the current review.

17 Comparison to Previous Reviews

Similar results were obtained to previous reviews, with respect to the number of single-use measurement tools identified in the literature, and dissimilar with respect to recommended STEM attitude measures. Blalock and colleagues (2008) conducted a large-scale evaluation of pre-existing science attitude measures published between 1935-2005. As the term STEM was not coined until 2001 (Teaching Institute for Excellence in STEM, 2010), Blalock did not include attitude measures addressing subject areas other than science. There was no overlap in measures included in Blalock and colleagues' 2008 review and the current analysis, and many measures were developed following the Blalock review. This is most likely due to our two subject minimum inclusion criteria, which eliminated many science attitude measures cited by Blalock. The majority (87%) of measures included for psychometric evaluation within Group 3 of our review were published after 2005.

Blalock and colleagues (2008) review yielded a total of 66 tools, with 28 missing fundamental psychometric data and 37 cited within only a single study (56%). The current review found similar results with 80% of included measures only being cited within a single article, which shows an increase in single-use measures over the past 13 years. Despite Blalock et al.'s (2008) suggestions, researchers have continued to create single-use measures that meet the specific goals of a given project and/or intervention program. (e.g., Test of Science Related Attitudes – Adapted; Koul et al., 2018). However, researchers are developing measures that are strikingly similar to previously published tools. Following creation of these new or adapted scales, little reliability and validity assessment is done to ensure appropriateness of a measure. The use of an unvalidated measure inhibits a researcher's ability to say with certainty that they are measuring an intended construct. For example, an adapted measure may aim to assess an individual's positive and/or negative attitudes towards school subjects such as science and mathematics, but the items are worded in such a way that the cognitive and affective underpinnings that constitute an attitude are not what is being evaluated. Instead, disparate items are put together due to a lack of a definition of the underlying construct. Items may follow a common theme (e.g., attitudes towards science and mathematics) but not a common construct (e.g., attitudes towards completing science and math activities in school vs attitudes towards wanting to become a engineer). These discrepant constructs may result in very different attitude ratings within a single measure attempting to capture student attitudes (Kind et al., 2007). If this measure was being used within a research study to gauge the effectiveness of an intervention aimed to increase student attitudes, scores collected at both pre- and post-assessment would be inaccurate and therefore subsequent conclusions drawn about the intervention are also inaccurate. Authors who routinely adapt measures to meet their own research needs appear to lack or at minimum undervalue the theoretical and methodological rigor that goes into creating a scale and defining a given construct. Similar conclusion's must still be drawn within the current

review, as improvements in the prevalence of single-use measures have not been made since Blalock's 2008 recommendations.

Kennedy, Quin and Taylor (2016) completed a similar review of science attitude measures in their process of creating a new school science attitude measure. There was minor overlap of reported measures with the inclusion of the Relevance of Science Education (ROSE; Schreiner and Sjøberg, 2004) scale within both Kennedy and colleagues review and the present one. This measure was included within the current review as it addresses multiple core subject areas of science including biology, chemistry, physics, earth science, and astronomy as well as technology. With the exception of the ROSE scale, there was no additional overlap in evaluated instruments, mostly due to Kennedy and colleagues focus on science attitude measures, and no other subject areas within STEM. Additionally, Kennedy utilized different construct definitions of attitudes when categorizing measures, however, they addressed self-efficacy, interest towards a future career in science, and personal relevance within their framework, similarly to Savelsberg's (2016) framework used within the current review. Kennedy's (2016) review lacked an evaluation of included measures psychometric information, limiting the current comparison of results.

18 Adapted Measures

The current review yielded adapted measures from well-known and previously validated tools (e.g., Test of Science Related Attitudes – Adapted, Science Motivation Questionnaire – Adapted, Self-Efficacy and Anxiety Questionnaire – Adapted). Many researchers looking to address STEM attitudes modified a measure of science or mathematics attitudes with the addition of items to address other core subject areas.

42

Identical items were commonly used with the substitution of the term "science" or "mathematics" with another STEM target subject (Science Motivation Questionnaire II – Adapted; Ardura & Perez-Bitrian, 2018). Researchers trust this single word swap maintained the reliability and validity of the original scale while addressing new subject attitudes (Fennema-Sherman Attitude Scale -Adapted; Kager & Foley, 2017). Many conducted a single reliability and/or validity analysis of adapted measures, yielding suitable Cronbach's alpha values or factor analysis results to support the use of the new scale. However, it was common for the original measures to be validated using different target populations (e.g., undergrad students or adults) than the adapted measures were used for (e.g., elementary or high-school students). It becomes difficult to compare and/or combine psychometric information of a given scale in order to determine its overall utility when items and intended population are different between administrations due to adaptions.

As there were many published measures addressing science and math, and a limited number of scales addressing STEM attitudes as a whole, researchers may have been compelled to adapt a scale due to a lack of alternatives. However, as this review highlights, there are now a multitude of measures that have been published within the past decade that report strong psychometric information and were designed to address multiple STEM subject areas. These measures are more suitable for use than an adapted measure due to the amount of research and time that has been put into ensuring their reliability, validity, and appropriateness for use with target participant populations.

19 Limitations of the Psychometric Grading Framework

Leung and colleagues psychometric grading framework was used to evaluate the strength of included measures. The framework was user-friendly and allowed for the synthesis of psychometric coefficients to yield an overall evaluation of a given measure. However, through its use, researchers made note of a few limitations and shortcomings that may contribute to a discrepancy within results. For example, the framework is limited to the amount of psychometric information provided for a given measure. As an overall grade is determined based on a minimum number of strong coefficients to be awarded a grade of A or B, a measure that reports a minimum of three "B-worthy" coefficients in conjunction to multiple poor psychometric values will receive a grade of "Good" similarly to a measure reporting 10 "A-worthy" values. Therefore, two measures with large differences in psychometric grading framework appears to take the stance of "good-enough" with regards to measure evaluation.

Additionally, the framework does not require both reliability and validity coefficients to be present in order to receive a strong grade. As at least three A or B grades are required to be awarded an overall strength of "good", a given measure could report three high Cronbach's alphas evaluating internal consistency and are rated as "good". The framework does not outline that at least one A or B must be from a reliability and validity coefficients, or even that the coefficients need to be from different psychometric tests. This allowing for limited information to yield a higher grade for a given measure than it may deserve due to only reporting reliability or validity

information. However, if used appropriately and with proper discretion, the PGF allows for a comprehensive evaluation of survey tools.

Limitations

A common limitation of systematic reviews are possible inconsistencies between researcher ratings across title/abstract and full text reviews. Attempts to minimize errors in ratings were used in the current review, such as training together, discussing disagreements as a team until reaching consensus, and developing a structured data extraction form to ensure the accurate collection of data. This did not appear to be a concern due to consistency between reviewer ratings and data extraction across raters.

Results also may be limited by the initial databases (ERIC, PsycINFO, Web of Science, SCOPUS) searched. Researchers choose the well-known databases within the psychological community, however, there may be relevant peer-reviewed articles within other databases not searched such as STEM specific databases. To minimize this effect, reference lists of included articles were searched by researchers to find any other related articles containing identified STEM attitude measures that may be housed on different databases or missed in the initial search. This search resulted in 19 additional relevant articles included for data extraction. Despite attempts to include all pertinent peer-review articles and STEM attitude measures, it would be inaccurate to believe the current review captures all published material.

The current review limited its scope to quantitative self-report measures of STEM attitudes. Qualitative measures, encompassing drawings, focus groups, or interviews, were not included which may limit results obtained. The Draw-a-Scientist Test (DAST; Chambers, 1983) is a frequently cited measure within included articles of this review and appeared as an effective tool to evaluate younger elementary students (K-5) attitudes towards scientists. Our exclusion of qualitative measures may contribute to the lack of

identified measures that were appropriate for use with younger elementary students (grade 1-3) within the current review (see Table 5). As self-report measures require a minimum reading level and agency by examinees, their use with young students may be inappropriate, prompting the use of interviews and drawing exercises to evaluate attitudes. However, when evaluating the attitudes of a large number of participants in order to assess the impact of a particular research intervention, self-report measures are the most feasible data-collection technique. Therefore, when choosing a STEM attitude measure, educators, administrators, and researchers must consider the age of examinees, instruments with the strongest psychometric information, and the intended application of the tool.

Implications

A core objective of the current review was to identify and evaluate measures of STEM attitudes with strong psychometric evidence to be recommended and used within psychological research. This would allow researchers the opportunity to use a reliable and validated tool to accurately evaluate participants STEM attitudes and/or track treatment changes. Measures receiving the highest overall grade of "Good" using the PGF signifies the measure contains a multitude of psychometric information, with strong coefficients reported. Recommendations will be made for measures receiving the highest possible grade within the framework as their grade indicated strong psychometric data. As previously mentioned, multiple measures were identified as good and can be recommended by the current researchers for use within a variety of projects to address a multitude of research questions. However, the final decision on an appropriate measure is placed on individual researchers in order to choose a tool that fits best given the target participant population and operationalization of the construct of attitude.

To determine an appropriate measure, a given researchers intended participant population should be considered and identified. This is important to ensure the measure is being used with participant populations with which it has already been validated. Measures of the current review can be subdivided into intended use for early elementary (grades 4-6), middle school (grades 7-8) and high-school aged students (grades 9-12). If looking to evaluate STEM attitudes of early elementary students, the STEM Attitude Survey (Guzey, Harwell, Moore, 2014), STEM Career Interest Survey (STEM – CIS; Kier et al., 2014), or the Upper Elementary S-STEM (Unfried, 2015) would be an appropriate choice.

For researchers targeting middle-school aged students the Measure of STEM Interest for Adolescents (Falk et al., 2016), Science and Technology Attitude Scale (Nuhoğlu, 2008), Self-Efficacy for Science and Technology (SESST; Tatar et al., 2009), Trends in International Mathematics and Science Study 2007 (Olson et al., 2008), and the Middle/High S-STEM (Unfried, 2015) would be fitting choices. Finally, researchers evaluating the attitudes of high-school aged students can use the Middle/High S-STEM (Unfried, 2015), PISA 2006 - Science Interest Scale (OECD, 2009b), ROSE - Science Interest Scale (Schreiner and Sjøberg 2004), or the STEM Semantics Survey (Tyler-wood et al., 2010).

Additionally, a measure can be chosen based on the specific attitude construct that it addresses. Recall, within Savelsbergh and colleague's model of general attitudes there are multiple subconstructs (interest, relevance, self-efficacy and subjective norms of scientists). As most measures focused on a single construct, a tool can be identified based on a researcher's specific construct operational definition of attitudes. For example, if looking to target participants self-efficacy the Self-Efficacy for Science and Technology (SESST; Tatar et al., 2009) would be an appropriate choice. The strongest measure identified to address career interests within participants was the STEM Career Interest Scale (Kier et al., 2014). More general interests can be evaluated using the Measure of STEM Interest in Adolescents (Falk et al. 2016). Also, general overarching attitudes can be evaluating using the STEM Attitude Survey (Guzey et al., 2014) or the Upper Elementary or Middle/High School S-STEM (Unfried, 2015).

Finally, researchers, practitioners, and other professionals looking to utilize a specific measure must also consider the characteristics of their participant population in comparison to those of the sample used to gather initial psychometric data and validate a tool. This will ensure that the measure can reliably and accurately assess STEM attitudes within a particular population of students as it is comparable to the sample used in the psychometric evaluation. A measure is only valid for the sample population on which it was tested, and results cannot be generalized to all populations. For example, the Middle/High School S-STEM was only evaluated on students within the southeastern United States, who were primarily Caucasian. In comparison, the STEM Career Interest Survey was administered to students in Turkey, Malaysia, Taiwan, and various regions of the United States. The STEM Career Interest Survey can be appropriately applied to a diverse group of students, while the Middle/High School S-STEM should only be used with samples within the United States and/or Canada due to their population similarities. Information regarding the population each included measure was validated with can be found within Table 5.

Finally, the current review aimed to evaluate STEM attitude self-report measurement tools for the use within research and program evaluation studies. However, these measures can be used for a multitude of other purposes such as for self-evaluation of student's interests and/or as a tool for career counsellors. The core purpose of these tools is to evaluate the attitudes of elementary and high school students towards the field of STEM. This construct would be informative to evaluate within upper high school students contemplating whether they should pursue a STEM-related post-secondary program or career or post-secondary students looking to switch specializations, as a quantitative evaluation of their attitude levels may aid in this decision. Career and guidance counsellors utilize a multitude of self-assessment tools for narrowing a professional field of interest and developing self-awareness within a student such as the Myers-Briggs Type Indicator. From the current evaluation of STEM attitude measures, the Middle/High S-STEM (Unfried, 2015) or STEM Career Interest Survey (STEM – CIS; Kier et al., 2014) would be appropriate measures, with US or Canadian students. These measures received the highest overall evaluation of good and would be helpful in evaluating a student's interest towards STEM. Even though the STEM-CIS has been classified as a tool appropriate for elementary aged students, it has been validated within high school samples (Çevik, 2018; Mau et al., 2019).

Conclusions

Within the literature, a multitude of STEM attitude measures have been published in the past decade. This has been in conjunction with the push by educators and stakeholders to increase post-secondary participation within STEM courses, and eventual entry into a STEM-related field. Research intervention programs aiming to promote STEM attitudes and literacy within elementary and high-school aged students have increased over the past decade as well, leading to an influx of newly published measures aimed to evaluate treatment effects. The broad STEM subject coverage of these measures, with most addressing all four core subjects, speaks to the pivot towards a holistic and interdisciplinary view of STEM education and attitudes. Additionally, while evaluating specific construct coverage used by researchers within STEM attitude measures, it became apparent that most tools operationalize attitudes as an interest towards a specific topic and/or self-efficacy beliefs. The use of Savelsbergh's theoretical framework of STEM attitudes allowed for the accurate categorization and evaluation of measures based on specific construct definitions used.

As previous reviews, as well as the current review have uncovered, many of the measures identified are single-use tools, lacking reliability and validity information. We echo recommendations made by previous researchers, and advocate for the use of well validated measures within psychological research to ensure proper conclusions are drawn on treatment effects. Multiple reliable and validated tools were identified within the current review using a psychometric grading framework. We recommend their use to accurately evaluate student STEM attitudes.

Group	Terms	
STEM	STEM OR Science OR Technology OR Engineering OR Math*	
Attitude	Affinity OR Attitude* OR Perception* OR Interest OR Belief* OR Opinion OR Motivation OR career* OR enjoyment OR engagement OR attainment OR self- confidence OR self-efficacy	
Measurement	Measur* OR instrument* OR interview OR scale OR questionnaire OR assessment OR inventory	
Filters	• NOT (undergraduate OR non-experimental OR preschool OR pre-school OR adulthood OR pre-service OR preservice)	
	• In the Web of science subject area was limited to Education, Psychology, or Educational research	
	Language limited to English	

Table 1: Search Terms Used to Select Applicable Research Articles

Criterion	Accept	Reject
Science Technology Engineering and Math (STEM)	Two or more STEM subject areas addressed including; Science, Technology, Mathematics, Engineering, Chemistry, Biology, Computer Science, Earth Science	Only one subject area addressed (e.g. Mathematics), Economics, Social science, Agriculture, Medicine/Health (exception: if health is used as context for any of the included subjects)
Primary or Secondary students	Elementary/primary school, Secondary/high school, Gender-specific (for example girls only)	Pre-elementary, tertiary education/college/university, adult education, pre-service teacher/teacher students
STEM attitude measured as a quantitative dependent variable	Attitude toward STEM (lessons), STEM attitudes, Relevance of STEM, Interest in STEM (lessons), Career interest/intentions for STEM, Interest in STEM as a leisure activity, Self-efficacy/self- confidence for STEM (lessons), Normality of scientist	Attitudes measured with open questions, attitudes towards computer learning, collaborative learning, dissection, environment/social issues, or health/disease / safe sex / healthy food, student's perception of usefulness/learning value of the intervention
Assessment/Instrument/Sc ale Properties - The study must present information on the measurement properties of the scale	Reliability, test-retest reliability, internal consistency, standard error of measurement, validity, content validity, convergent validity, discriminant validity, factor analysis, construct validity, face validity	No assessment/instrument/scale properties reported
Self-Report	The study must describe the use of a self-report questionnaire or an instrument that is completed with a student.	Studies which use only observational methods of recording outcomes (e.g. event recording) or interviews, respondents are not students (for example teachers)
Empirical Article - Publications must present original empirical data	Group designs, randomized controlled trials, survey studies, single-case experimental research designs	Case reports or secondary sources/reviews

Table 2: Inclusion Criteria for Title/Abstract and Full Text Review

Category	Number of Measures	Highest Scoring Measures Based on PGF
Overall Attitude	24	STEM Attitude Survey (Guzey, Harwell, Moore, 2014)
Relevance (Personal and/or Societal)	0	
Interest (classroom, leisure, and/or career)	41	PISA 2006 – Science Interest Scale
		(OECD, 2009b)
Self-Efficacy	20	Self-Efficacy for Science and Technology (SESST; Tatar et al., 2009)
Subjective Norm/Normality of Scientists	1	Conception of Professionals in STEM – Adapted (DeWitt et al., 2013)
Multiple Categories	18	STEM Career Interest Scale (Kier et al., 2014)
All Categories	0	
Total	104	

Table 3: Overview of STEM Attitude Construct Coverage within Measures

Instrument	Number of Items & Format	Intended Age Group	Attitude Construct Coverage	STEM Subject Area(s) Covered (n = number of items per subject area)
Educational and Career Interest Scale (Oh et al., 2013)	9; Multiple-point response	Grade 9-12	Interest (classroom, leisure, career)	Science (3), Technology/Engineering (3), Math (3)
Mathematics and Technology Attitude Scale (MTAS; Pierce et al., 2007)	20; Multiple-point response	Grade 7-8	Attitude	Technology (8), Math (12)
Measure of STEM Interest in Adolescents (Falk et al, 2016)	32; Multiple-point response	Grade 7-8	Interest (classroom, leisure, career), Self-Efficacy	Science (16), STEM (16)
PISA 2006 – Science Interest Scale (OECD, 2009b)	8; Multiple-point response	Grade 9-12	Interest (classroom, leisure, career)	Biology (2), Chemistry (1), Physics (1), Astronomy (1), Earth Science (1), Science (2)
Relevance of Science Education (ROSE; Schreiner and Sjøberg 2004)	108; Multiple-point response	Grade 9-12	Interest (classroom, leisure, career)	Biology (40), Chemistry (10), Physics (20), Earth Science (9), Technology (10), Astronomy (11), Science (8)
Science and Technology Attitude Scale (Nuhoğlu, 2008)	20; Multiple-point response	Grade 7-8	Attitude	Science (10), Technology (10)
Science Self-Efficacy Questionnaire(SSEQ; Smist, 1993)	27; Multiple-point response	Grade 9-12	Self-Efficacy	Biology (8), Chemistry (8), Physics (8), Science (3)

Table 4: Group 3 Instrument Characteristics (n= 15)

Self-Efficacy for Science and Technology (SESST; Tatar et al., 2009)	27; Multiple-point response	Grade 7-8	Self-Efficacy	Science (27), Technology (27)
STEM Attitude Survey (Guzey et al, 2014)	28; Multiple-point response	Grade 4-6	Attitude	Science (5), Technology (5), Engineering (6), Math (5), STEM (7)
STEM Career Interest Survey (STEM-CIS; Kier et	44; Multiple-point response	Grade 4-6	Interest (classroom, leisure, career)	Science (11), Technology (11), Engineering (11), Math
al., 2014)			Self-Efficacy	(11)
STEM Project Based Learning Questionnaire (STEM PBL; Han, 2017)	51; Multiple-point response	Grade 9-12	Attitude	Science (8), Technology (11), Engineering (1), Math (1), STEM (30)
STEM Semantics Survey (Tyler-Wood et al., 2010)	25; Semantic Differential	Grade 9-12	Interest (classroom, leisure, career)	Science (5), Technology (5), Engineering (5), Math (5), STEM (5)
Upper Elementary S- STEM (Unfried, 2015)	43; Multiple-point response	Grade 4-6	Attitude	Science (9), Technology/ Engineering (11), Math (10), STEM (13)
Middle/High School S- STEM (Unfried, 2015)	43; Multiple-point response	Grade 7-12	Attitude	Science (9), Technology/ Engineering (11), Math (10), STEM (13)
Trends in International Mathematics and Science Study 2007 (TIMSS 2007; Olson et al., 2008)	24; Multiple-point response	Grade 7-8	Interest (classroom, leisure, career)	Science (4), Math (4), Biology (4), Earth Science (4), Chemistry (4), Physics (4)

Instrument	Sample	Reliability	Validity	Overall Psychometric Grade
Educational and Career Interest Scale (Oh et al., 2013)	High-school students (<i>n</i> = 658 [USA; 102], 702 [Turkey; 67])	<i>Internal Consistency</i> (α): .88 [102], .88[67]	<i>Content:</i> expert panel review, focus group feedback, literature review [102] <i>Construct:</i> CFA yielded 4-factor subscale % variance (49%, 38%, 7%, 4%) [67]	Adequate
Mathematics and Technology Attitude Scale (MTAS; Pierce et al., 2007)	High-school students (<i>n</i> = 305 [Australia; 10], 1068 [Greece; 11])	<i>Internal Consistency</i> (α): .68 92 [11], .6589 [106]	<i>Content</i> : literature review [106] <i>Construct:</i> KMO of .679 [11], 5-factor 65% of total variance [106]	Adequate
Measure of STEM Interest in Adolescents (Falk et al., 2016)	Middle-school students (<i>n</i> = 811 [USA: 121], 249 [USA; 45], 106 [USA; 120])	<i>Internal Consistency (</i> α): .87, [121], .85 [121], .6989 [45], .5889 [120]	<i>Content:</i> Literature review, expert panel review, participant feedback [121,45] <i>Construct:</i> $r = .94$, .89, .92 [121], PCA yielded 8-factor subscale % variance (9.84%, 8.46%, 40.09%, 9.67%, 37.16%, 5.58%, 7.97%, 13.59%) [45]	Good
PISA 2006 – Science Interest Scale (OECD, 2009b)	15-year-old aged students (<i>n</i> = 25,476 [International; 40], 401 [Finland; 72], 40,000 [International; 101], 4714 [Finland; 74])	<i>Internal Consistency</i> (α): .87 [72], .83 [101], .85 [74], .64 - .78 [40]	<i>Content:</i> Expert panel review, participant feedback [101] Criterion: <i>r</i> = .75 (JOYSCIE), .72 (SCIEFUT) [101]	Good
Relevance of Science Education (ROSE; Schreiner and Sjøberg 2004)	High-school students (<i>n</i> = 942 [Taiwan; 28], 1591[China; 31], 3615 [Finland; 132])	<i>Internal Consistency</i> (α): .97 [28], .6798 [31], .5588 [132]	<i>Content</i> : Expert panel review, participant feedback [115,31] <i>Construct</i> : KMO of .902 [132], EFA yielded 3-factor subscale % variance (34.4%, 12.7%, 9.2%) [132]	Good

 Table 5: Reliability, validity, and psychometric strength of STEM affinity instruments (n= 15)

Science and Technology Attitude Scale (Nuhoğlu, 2008)	Middle-school students (<i>n</i> = 422 [Turkey; 100], 66 [Turkey; 41])	<i>Internal Consistency</i> (α): .87 [100], .82 [41]	<i>Content:</i> Expert panel review [100] <i>Construct:</i> KMO of .86 [100], Factor analysis yielded 5-factor structure with 56% variance explained [100]	Good
Science Self-Efficacy Questionnaire (SSEQ; Smist, 1993)	High-school students (<i>n</i> = 157 [USA; 82], 402 [USA; 93])	<i>Internal Consistency</i> (α): .94 [82], .95 [93]	<i>Construct:</i> EFA yielded 3-factor structure with 52.9% variance explained [93]	Adequate
Self-Efficacy for Science and Technology (SESST; Tatar et al., 2009)	Middle-school students (<i>n</i> = 400 [Turley; 123], 705 [Turkey; 131])	<i>Internal Consistency</i> (α): .93 [123], .93 [131]	<i>Content:</i> expert panel review, participant feedback, literature review [123] <i>Construct:</i> KMO of .95 [123], EFA yielded 3-factor structure with 51% variance explained [123]	Good
STEM Attitude Survey (Guzey et al., 2014)	Elementary school students (<i>n</i> = 662 [USA; 128], 732 [USA; 86], 40 [Turkey; 10])	Internal Consistency (α): .91 [128], .91 [86], .92 [10]	<i>Content:</i> expert panel review, literature review [128] <i>Construct:</i> KMO of .892 [128], general liner model yielded 36.11% variance explained [86]	Good
STEM Career Interest Survey (STEM-CIS; Kier et al., 2014)	Elementary school students (<i>n</i> = 165 [Turkey; 24], 892 [Turkey; 42],1061 [USA; 76], 1033 [Turkey; 134], 129 [Malaysia; 94], 590 [Taiwan; 92])	<i>Internal Consistency</i> (α): .92 [24], .88 [42], .93 [134], .98 [92], .7789 [76], .8586 [94] <i>Test-Retest:</i> r = .87 [134], .60 85 [92]	<i>Content:</i> expert panel review, participant feedback, literature review [76,134,92] <i>Construct:</i> KMO of .86 [24] <i>Criterion:</i> $r = .47$ to ISS [134]	Good
STEM Project Based Learning Questionnaire (STEM PBL; Han, 2017)	Middle-school students (<i>n</i> = 816 [Korea; 60], 785 [Korea; 61])	<i>Internal Consistency</i> (α): .87 [61], .8089 [60]	<i>Construct:</i> CFA yielded 5-factor structure with 56.97% variance explained [61]	Adequate

STEM Semantics Survey (Tyler-Wood et al., 2010)	High-school students (<i>n</i> = 174 [USA 111], 216 [China; 118] 360 [USA; 35], 174 [Hawaii; 130], 364 [USA; 34], 32[USA; 129])	<i>Internal Consistency</i> (α): .84 [118], .8493 [111], .8493 [130], .9094 [35], .8993 [34], .8493 [129]	<i>Content:</i> expert panel review [130] <i>Construct:</i> EFA yielded 1-factor structure with51.9% variance explained [118]	Good
Upper Elementary S- STEM (Unfried, 2015)	Elementary school students (<i>n</i> = 4232 [USA; 133], 242 [China; 143])	<i>Internal Consistency</i> (α): .90 [143], .91 [143], .8387 [133]	<i>Content:</i> expert panel review [133] <i>Construct:</i> EFA yielded 4-factor structure with 67.7% variance explained [133]	Good
Middle/High School S- STEM (Unfried, 2015)	Middle-school and High- school students (<i>n</i> = 17,485 [USA; 133], 67 [USA; 18])	<i>Internal Consistency (</i> α): .89 [18], .8990 [133]	<i>Content:</i> expert panel review [133] <i>Construct:</i> EFA yielded 4-factor structure with 73.9% variance explained [133]	Good
Trends in International Mathematics and Science Study (TIMSS2007; Olson et al., 2008)	Middle-school students (<i>n</i> = 40, 803 [International; 90], 160,922 [International; 103])	<i>Internal Consistency (</i> α): .72 [103], .6681 [90]	<i>Content:</i> expert panel review, participant feedback [103] <i>Construct (Convergent):</i> $r = .48, .45, .57, .69$ [90]	Good

Notes: USA = United States of America, PCA = Principal Components Analysis, CFA = Confirmatory Factor Analysis, EFA = Exploratory Factor Analysis, ISS = The Interest in Science Scale, JOYSCIE = Enjoyment of Science Scale, SCIEFUT = Future-Oriented Science Motivation Scale, n = number of participants within research study.

Intended Population	Attitude Construct	Recommended Measures
Elementary-School	Interest	STEM Career Interest Survey (Kier et al., 2014)
	Self-Efficacy	n/a
	Relevance	n/a
	Subjective Norm/Normality of Scientists	n/a
	General Attitudes	STEM Attitude Survey (Guzey et al., 2014) Upper Elementary S-STEM (Unfried, 2015)
Middle-School	Interest	Measure of STEM Interest for Adolescents (Falk et al., 2016)
		Trends in International Mathematics and Science Study 2007 (Olson et al., 2008)
	Self-Efficacy	Self-Efficacy for Science and Technology (Tatar et al., 2009
	Relevance	n/a
	Subjective Norm/Normality of Scientists	n/a
	General Attitudes	Science and Technology Attitude Scale (Nuhoğlu, 2008) Middle/High S-STEM(Unfried,2015)
High-School	Interest	PISA 2006 - Science Interest Scale (OECD, 2009b)
		ROSE (Schreiner and Sjøberg 2004) STEM Semantics Survey (Tyler-Wood et al., 2010)
	Self-Efficacy	n/a
	Relevance	n/a
	Subjective Norm/Normality of	n/a
	General Attitudes	Middle/High S-STEM(Unfried,2015)

Table 6. Recommended measures given participant population and target attitude

construct

Note: n/a = no measures identified



Figure 1. PRISMA 2009 Flow Diagram for Systemic Reviews and Meta-Analyses

(Moher, Liberati, Tetzlaff, Altman, The PRISMA Group, 2009)
References

- Akey, T. M. (2006). School context, student attitudes and behavior, and academic achievement: An exploratory analysis. MDRC.
- Ambriz, J. D. (2016). Social cognitive career theory (SCCT) and Mexican/Mexican-American youth career development, with a special focus on stem fields.Washington State University.
- American Educational Research Association. (2014). Standards for educational and psychological testing. American Educational Research Association.
- Augustine, N. R. (2005). *Ri*sing above the gathering storm: Energizing and employing America for a brighter economic future. Retrieved March, 19, 2008.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behaviour change. *Psychological Review*, 84(2), 191.
- Bandura, A. (1986) Social foundations of thoughts and actions: A social cognitive theory. Englewood Cliffs, NJ: Prentice Hall.
- Blalock, C.L., Lichtenstein, M. J., Owen, S., Pruski, L. Marshall, C., & Toepperwein, M.
 (2008) In Pursuit of Validity: A comprehensive review of science attitude instruments 1935–2005, *International Journal of Science Education*, 30:7, 961-977, DOI: 10.1080/09500690701344578

- Carter-Sowell, A. R., & Zimmerman, C. A. (2015). Hidden in plain sight: Locating, validating, and advocating the stigma experiences of women of color. *Sex Roles*, 73(9), 399-407.
- Chachashvili-Bolotin, S., Milner-Bolotin, M., Lissitsa, S. (2016). Examination of factors predicting secondary student's interest in tertiary STEM education. *International Journal of Science Education*, 38(3), 366-390.
- Fouad, N. A., & Santana, M. C. (2017). SCCT and underrepresented populations in STEM fields: Moving the needle. *Journal of Career Assessment*, 25(1), 24-39.
- Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, 20(2), 507-537.
- Friday Institute for Educational Innovation (2012). Middle/High School Student Attitudes toward STEM Survey. Raleigh, NC.
- Green, M. (2007). Science and engineering degrees: 1966-2004. (NSF 07–307). Arlington, VA: National Science Foundation.
- Heise, D. R. (1969). Some methodological issues in semantic differential research. *Psychological Bulletin*, 72(6), 406.
- Herschbach, D. R. (2011). The STEM initiative: Constraints and challenges. *Journal of* STEM Teacher Education, 48(1), 96-122.
- Hihi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational psychologist*, 41(2), 111–127.

- Inda-Caro, M., Rodríguez-Menéndez, C., & Peña-Calvo, J. V. (2016). Spanish high school students' interests in technology: Applying social cognitive career theory. *Journal of Career Development*, 43(4), 291-307.
- Lent, R. W., Brown, S. D., Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behaviour*, 45, 79-122.
- Lent, R. W., & Brown, S. D. (2006). Integrating person and situation perspectives on work satisfaction: A social-cognitive view. *Journal of Vocational Behavior*, 69(2), 236–247.
- Lent, R. W., Sheu, H., Gloster, C. S., & Wilkins, G. (2010). Longitudinal test of the social cognitive model of choice in engineering students at historically Black universities. *Journal of Vocational Behavior*, 76(3), 387–394.
- Lesseig, K., Slavit, D., & Nelson, T. H. (2017). Jumping on the STEM bandwagon: How middle grades students and teachers can benefit from STEM experiences. *Middle School Journal*, 48(3), 15-24.
- Leung K., Trevena L, Waters D (2012) Development of an appraisal tool to evaluate strength of an instrument or outcome measure. *Nurse Researcher*. 20, 2, 13-19.
- Kennedy, J.P., Quinn, F., Taylor, N. (2016) The school science attitude survey: a new instrument for measuring attitudes towards school science, *International Journal of Research & Method in Education*, 39:4, 422-445, DOI: 10.1080/1743727X.2016.1160046

- Kind, P. M., Jones, K., & Barmby, P. (2007). Developing attitudes towards science measures. *International Journal of Science Education*, 29(7), 871-893. https://doi.org/10.1080/09500690600909091
- Koonce, D. A., Zhou, J., Anderson, C. D., Hening, D. A., & Conley, V. M. (2011). AC2011-289: What is STEM. In American Society for Engineering Education.
- Marrero, M. E., Gunning, A. M., & Germain-Williams, T. (2014). What is STEM education. *Global Education Review*, 1(4).
- McComas, W. F., & Burgin, S. R. (2020). A Critique of "STEM" Education. Science & Education, 29(4), 805-829
- McDowell, I. (2006). Measuring health: a guide to rating scales and questionnaires. Third Edition. Oxford University Press. New York, NY.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, 22(3), 276-282.
- Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097
- Mohr-Schroeder, M. J., Cavalcanti, M., & Blyman, K. (2015). STEM education: Understanding the changing landscape. In A practice-based model of STEM teaching. *Brill Sense*. (pp. 3-14).

- Morgan, S. L., Gelbgiser, D., & Weeden, K. A. (2013). Feeding the pipeline: Gender, occupational plans, and college major selection. *Social Science Research*, 42(4), 989-1005.
- Munby, H. (1983). An investigation into the measurement of attitudes in science education. Columbus, OH: SMEAC Information Reference Center.
- National Academy of Engineering and National Research Council. (2014). STEM
 Integration in K-12 Education: Status, Prospects, and an Agenda for Research.
 Washington, DC. *The National Academies Press*.
- Natural Sciences and Engineering Research Council of Canada Corporate Planning and Policy Directorate. (2010). Women in science and engineering in Canada. *Natural Sciences and Engineering Research Council of Canada*. Ottawa, ON, Canada.
- National Science Foundation. (2008). Comments on outreach from the Mathematical and Physical Sciences Advisory Committee. Retrieved from http://www.nsf.gov/attachments/110624/public/CommentsonOutreach
- Nugent, G., Barker, B., Welch, G., Grandgenett, N., Wu, C., & Nelson, C. (2015) A model of factors contributing to STEM learning and career orientation. *International Journal of Science Education*, 37(7), 1067–1088.
- Ong, M., Wright, C., Espinosa, L., & Orfield, G. (2011). Inside the double bind: A synthesis of empirical research on undergraduate and graduate women of color in science, technology, engineering, and mathematics. *Harvard Educational Review*, 81(2), 172-209.

- Potvin, P., Abdelkrim, H. (2014) Interest, motivation and attitude towards science and technology at K-12 levels: a systematic review of 12 years of educational research, *Studies in Science Education*, 50:1, 85-129, DOI: 10.1080/03057267.2014.881626
- Savelsbergh, E. R., Prins, G. T., Rietbergen, C., Fechner, S., Vaessen, B. E., Draijer, J. M., & Bakker, A. (2016). Effects of innovative science and mathematics teaching on student attitudes and achievement: A meta-analytic study. *Educational Research Review*, 19, 158-172.
- Sheltzer, J. M., & Smith, J. C. (2014). Elite male faculty in the life sciences employ fewer women. Proceedings of the National Academy of Sciences, 111(28), 10107-10112.
- Smelser, N. J., & Baltes, P. B. (Eds.). (2001) International encyclopedia of social and behavioural sciences. Vol. 11. Amsterdam: Elsevier, 2001.
- Streiner, D.L., Norman, G.R. (2003) Health measurement scales: A practical guide to their development and use. Third edition. Oxford University Press, Oxford.
- United States Consensus Bureau. (2019) Quick Facts United States Results. U.S. Department of Commerce.

https://www.census.gov/quickfacts/fact/table/US/PST045219

van Aalderen-Smeets, S. I., Walma van der Molen, J. H., & Xenidou-Dervou, I. (2019).
Implicit STEM ability beliefs predict secondary school students' STEM
self-efficacy beliefs and their intention to opt for a STEM field career. *Journal of research in science teaching*, 56(4), 465-485.

- Wang, X. (2013). Why students choose STEM majors: Motivation, high school learning, and postsecondary context of support. *American Educational Research Journal*, 50(5), 1081-1121.
- Yáñez-Marquina, L., & Villardón-Gallego, L. (2016). Attitudes towards mathematics at secondary level: Development and structural validation of the scale for assessing attitudes towards mathematics in secondary education (SATMAS). *Electronic Journal of Research in Educational Psychology*, 14(3), 557-581.
- Yata, C., Ohtani, T., & Isobe, M. (2020). Conceptual framework of STEM based on Japanese subject principles. *International Journal of STEM Education*, 7(1), 1-10.

Appendix A. Articles Included in Review

- Ahmed, W., & Mudrey, R. R. (2019). The role of motivational factors in predicting STEM career aspirations. *International Journal of School & Educational Psychology*, 7(3), 201-214.
- Akpınar, E., Yıldız, E., Tatar, N., & Ergin, Ö. (2009). Students' attitudes toward science and technology: an investigation of gender, grade level, and academic achievement. *Procedia-Social and Behavioral Sciences*, 1(1), 2804-2808.
- 3. Ali, S. R., Brown, S. D., & Loh, Y. (2017). Project HOPE: Evaluation of health science career education programming for rural Latino and European American youth. *The Career Development Quarterly*, 65(1), 57-71.
- Alliman-Brissett, A. E., & Turner, S. L. (2010). Racism, parent support, and mathbased career interests, efficacy, and outcome expectations among African American adolescents. *Journal of Black Psychology*, 36(2), 197-225.
- 5. Ambriz, J. (2016). Social cognitive career theory (SCCT) and Mexican-American youth career development, with a special focus on stem fields.
- 6. Ardura, D., & Pérez-Bitrián, A. (2018). The effect of motivation on the choice of chemistry in secondary schools: adaptation and validation of the Science Motivation Questionnaire II to Spanish students. *Chemistry Education Research and Practice*, 19(3), 905-918.

- 7. Babarović, T., Dević, I., & Burušić, J. (2019). Fitting the STEM interests of middle school children into the RIASEC structural space. *International Journal for Educational and Vocational Guidance*, 19(1), 111-128.
- Baker, T. R. (2002). The effects of Geographic Information System (GIS) technologies on students' attitudes, self-efficacy, and achievement in middle school science classrooms (pp. 1-203). University of Kansas.
- Baker, T. R., & White, S. H. (2003). The effects of GIS on students' attitudes, selfefficacy, and achievement in middle school science classrooms. *Journal of Geography*, 102(6), 243-254.
- Baran, E., Canbazoglu Bilici, S., Mesutoglu, C., & Ocak, C. (2019). The impact of an out-of-school STEM education program on students' attitudes toward STEM and STEM careers. *School Science and Mathematics*, 119(4), 223-235.
- 11. Barkatsas, A. T., Kasimatis, K., & Gialamas, V. (2009). Learning secondary mathematics with technology: Exploring the complex interrelationship between students' attitudes, engagement, gender and achievement. *Computers & Education*, 52(3), 562-570.
- Barmby, P., Kind, P. M., & Jones, K. (2008). Examining changing attitudes in secondary school science. *International Journal of Science Education*, 30(8), 1075-1093.

- Bathgate, M. E., Schunn, C. D., & Correnti, R. (2014). Children's motivation toward science across contexts, manner of interaction, and topic. *Science Education*, 98(2), 189-215.
- 14. Beghetto, R. A., & Baxter, J. A. (2012). Exploring student beliefs and understanding in elementary science and mathematics. *Journal of Research in Science Teaching*, 49(7), 942-960.
- 15. Benek, I., & Akcay, B. (2019). Development of STEM attitude scale for secondary school students: Validity and reliability study. *International Journal of Education in Mathematics, Science and Technology*, 7(1), 32-52.
- 16. Bennett, J., Airey, J., Dunlop, L., & Turkenburg-van Diepen, M. (2020). The impact of human spaceflight on young people's attitudes to STEM subjects. *Research in Science & Technological Education*, 38(4), 417-438.
- 17. Benton, J. M. (2005). Self-concept and achievement: Student academic beliefs and self-concept as related to academic performance using the Repeat Third International Mathematics and Science Study. The University of Tulsa.
- Binns, I. C., Polly, D., Conrad, J., & Algozzine, B. (2016). Student perceptions of a summer ventures in science and mathematics camp experience. *School Science* and Mathematics, 116(8), 420-429.
- Blankenburg, J. S., Höffler, T. N., & Parchmann, I. (2016). Fostering today what is needed tomorrow: Investigating students' interest in science. *Science education*, 100(2), 364-391.

- 20. Bozdoğan, A. E., & Yalçın, N. (2009). Determining the influence of a science exhibition center training program on elementary pupils' interest and achievement in science. *Eurasia Journal of Mathematics, Science and Technology Education*, 5(1), 27-34.
- 21. Breakwell, G. M., & Robertson, T. (2001). The gender gap in science attitudes, parental and peer influences: Changes between 1987-88 and 1997-98. *Public Understanding of Science*, 10(1), 71-82.
- 22. Casas, Y., & Blanco-Blanco, Á. (2017). Testing Social Cognitive Career Theory in Colombian adolescent secondary students: a study in the field of mathematics and science. *Revista Complutense de Educación*, 28(4), 1173.
- 23. Cēdere, D., Jurgena, I., Helmane, I., Tiltiņa-Kapele, I., & Praulīte, G. (2015).
 Cognitive interest: problems and solutions in the acquisition of science and mathematics in schools of Latvia. *Journal of Baltic Science Education*, 14(4), 424.
- 24. Çevik, M. (2018). Impacts of the project based (PBL) science, technology, engineering and mathematics (STEM) education on academic achievement and career interests of vocational high school students.
- 25. Chachashvili-Bolotin, S., Milner-Bolotin, M., & Lissitsa, S. (2016). Examination of factors predicting secondary students' interest in tertiary STEM education. International *Journal of Science Education*, 38(3), 366-390.

- 26. Chandrasena, W. (2013). Seeding science success: relations of secondary students' science self-concepts and motivation with aspirations and achievement.
- 27. Chandrasena, W., Craven, R. G., Tracey, D., & Dillon, A. (2014). Seeding Science Success: Psychometric Properties of Secondary Science Questionnaire on Students' Self-Concept, Motivation, and Aspirations. *Australian Journal of Educational & Developmental Psychology*, 14, 186-201.
- 28. Chang, S. N., Yeung, Y. Y., & Cheng, M. H. (2009). Ninth graders' learning interests, life experiences and attitudes towards science & technology. *Journal of Science Education and Technology*, 18(5), 447-457.
- 29. Chen, Y.-F., Cannady, M. A., Schunn, C., & Dorph, R. (2017) Measures Technical Brief: Competency Beliefs in STEM. Retrieved from http:// www.activationlab.org/wp-content/uploads/2017/06/CompetencyBeliefs_STEM-Report_20170403.pdf
- 30. Cheema, J. R., & Skultety, L. S. (2017). Self-efficacy and literacy: A paired difference approach to estimation of over-/under-confidence in mathematics-and science-related tasks. *Educational Psychology*, 37(6), 652-665.
- 31. Chi, S., Wang, Z., Liu, X., & Zhu, L. (2017). Associations among attitudes, perceived difficulty of learning science, gender, parents' occupation and students' scientific competencies. *International Journal of Science Education*, 39(16), 2171-2188.

- 32. Choi, K., & Cho, H. H. (2002). Effects of teaching ethical issues on Korean school students' attitudes towards science. *Journal of Biological Education*, 37(1), 26-30.
- 33. Christidou, V. (2006). Greek students' science-related interests and experiences:
 Gender differences and correlations. *International Journal of Science Education*, 28(10), 1181-1199.
- 34. Christensen, R., Knezek, G., & Tyler-Wood, T. (2014). Student perceptions of science, technology, engineering and mathematics (STEM) content and careers. *Computers in Human Behavior*, 34, 173-186.
- 35. Christensen, R., Knezek, G., & Tyler-Wood, T. (2015). Alignment of hands-on STEM engagement activities with positive STEM dispositions in secondary school students. *Journal of Science Education and Technology*, 24(6), 898-909.
- 36. Demirtas, C., Arikan, S., & Mugaloglu, E. (2020). Adaptation of the Self-efficacy Beliefs in STEM Education Scale and Testing Measurement Invariance across Groups. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 1-20.
- 37. Dierks, P. O., Höffler, T. N., Blankenburg, J. S., Peters, H., & Parchmann, I. (2016). Interest in science: A RIASEC-based analysis of students' interests. *International Journal of Science Education*, 38(2), 238-258.
- 38. Donovan, E. P., Fronk, R. H., & Horton, P. B. (1985). A new science and engineering career interest survey for middle school students. *Journal of Research in Science Teaching*, 22(1), 19-30.

- 39. Dorph, R., Bathgate, M. E., Schunn, C. D., & Cannady, M. A. (2018). When I grow up: the relationship of science learning activation to STEM career preferences. International *Journal of Science Education*, 40(9), 1034-1057.
- 40. Drechsel, B., Carstensen, C., & Prenzel, M. (2011). The role of content and context in PISA interest scales: A study of the embedded interest items in the PISA 2006 science assessment. *International Journal of Science Education*, 33(1), 73-95.
- 41. Ercan, O. (2014). The effects of multimedia learning material on student's academic achievement and attitudes towards science courses. *Journal of Baltic Science Education*, 13(5).
- 42. Ergün, A. (2019). Identification of the interest of Turkish middle-school students in STEM careers: Gender and grade level differences. *Journal of Baltic Science Education*, 18(1), 90-104.
- 43. Falco, L. D., & Summers, J. J. (2019). Improving career decision self-efficacy and STEM self-efficacy in high school girls: Evaluation of an intervention. *Journal of Career Development*, 46(1), 62-76.
- 44. Falco, L. D., & Summers, J. J. (2021). Social persuasions in math and their prediction of stem courses self-efficacy in middle school. *The Journal of Experimental Education*, 89(2), 326-343.
- 45. Falk, J. H., Staus, N., Dierking, L. D., Penuel, W., Wyld, J., & Bailey, D. (2016).Understanding youth STEM interest pathways within a single community: The

Synergies project. *International Journal of Science Education*, Part B, 6(4), 369-384.

- 46. Finnerty, V. (2013). Can participation in a school science fair improve middle school students' attitudes toward science and interest in science careers? (Doctoral dissertation, University of Massachusetts Lowell).
- 47. Finson, K. D., & Enochs, L. G. (1987). Student attitudes toward science-technology-society resulting from visitation to a science-technology museum. *Journal of Research in Science Teaching*, 24(7), 593-609.
- 48. Fouad, N. A., Smith, P. L., & Enochs, L. (1997). Reliability and validity evidence for the middle school self-efficacy scale. *Measurement and Evaluation in Counseling and Development*, 30(1), 17-31.
- 49. Fredricks, J. A., Hofkens, T., Wang, M. T., Mortenson, E., & Scott, P. (2018).
 Supporting girls' and boys' engagement in math and science learning: A mixed methods study. *Journal of Research in Science Teaching*, 55(2), 271-298.
- 50. Garriott, P. O., Flores, L. Y., Prabhakar, B., Mazzotta, E. C., Liskov, A. C., & Shapiro, J. E. (2014). Parental support and underrepresented students' math/science interests: The mediating role of learning experiences. *Journal of Career Assessment*, 22(4), 627-641.
- 51. Garriott, P. O., Hultgren, K. M., & Frazier, J. (2017). STEM stereotypes and high school students' math/science career goals. *Journal of Career Assessment*, 25(4), 585-600.

- 52. Garriott, P. O., Raque-Bogdan, T. L., Zoma, L., Mackie-Hernandez, D., & Lavin, K. (2017). Social cognitive predictors of Mexican American high school students' math/science career goals. *Journal of Career Development*, 44(1), 77-90.
- 53. Gibbons, S. J., Hirsch, L. S., Kimmel, H., Rockland, R., & Bloom, J. (2004, October). Middle school students' attitudes to and knowledge about engineering. In International conference on engineering education, Gainesville, FL.
- 54. Griggs, M. S., Rimm-Kaufman, S. E., Merritt, E. G., & Patton, C. L. (2013). The Responsive Classroom approach and fifth grade students' math and science anxiety and self-efficacy. *School Psychology Quarterly*, 28(4), 360.
- 55. Grimmon, A. S., Cramer, J., Yazilitas, D., Smeets, I., & De Bruyckere, P. (2020). Interest in STEM among children with a low socio-economic status: further support for the STEM-CIS-instrument through the adapted Dutch STEM-LIT measuring instrument. *Cogent Education*, 7(1), 1745541.
- 56. Guzey, S. S., Harwell, M., & Moore, T. (2014). Development of an instrument to assess attitudes toward science, technology, engineering, and mathematics (STEM). School Science and Mathematics, 114(6), 271-279.
- 57. Halim, L., Abd Rahman, N., Wahab, N., & Mohtar, L. E. (2018). Factors influencing interest in STEM careers: An exploratory factor analysis. *Asia-Pacific Forum on Science Learning and Teaching* (Vol. 19, No. 2, pp. 1-34).
- 58. Hall-Lay, S. (2018). Gender Effects of Robotics Programs on STEM-Related Self-Efficacy of High School Students.

- 59. Hammack, R., Ivey, T. A., Utley, J., & High, K. A. (2015). Effect of an engineering camp on students' perceptions of engineering and technology. *Journal of Pre-College Engineering Education Research*, 5(2), 2.
- 60. Han, S. (2017). Korean students' attitudes toward STEM project-based learning and major selection. *Educational Sciences: Theory & Practice*, 17(2).
- 61. Han, S., & Carpenter, D. (2014). Construct validation of student attitude toward science, technology, engineering and mathematics project-based learning: The case of Korean middle grade students. *Middle Grades Research Journal*, 9(3), 27-41.
- 62. Hasni, A., & Potvin, P. (2015). Student's Interest in Science and Technology and Its Relationships with Teaching Methods, Family Context and Self-Efficacy. International Journal of Environmental and Science Education, 10(3), 337-366.
- 63. Hendley, D., Parkinson, J., Stables, A., & Tanner, H. (1995). Gender differences in pupil attitudes to the national curriculum foundation subjects of English, mathematics, science and technology in Key Stage 3 in South Wales. *Educational Studies*, 21(1), 85-97.
- 64. Hirsch, L. S., Carpinelli, J. D., Kimmel, H., Rockland, R., & Bloom, J. (2007, October). The differential effects of pre-engineering curricula on middle school students' attitudes to and knowledge of engineering careers. In 2007 37th Annual Frontiers in Education Conference-Global Engineering: Knowledge without Borders, Opportunities without Passports (pp. S2B-17). IEEE.

- 65. Hope, J. M. (2012). Exploring the nature of high school student engagement with science and technology as an outcome of participation in science journalism.
- 66. Huang, W. (2019). Examining the impact of head-mounted display virtual reality on the science self-efficacy of high schoolers. *Interactive Learning Environments*, 1-13.
- 67. İzzet Kurbanoğlu, N., & Arslan, S. (2015). High school students' educational and career interest (science-technology-mathematics) and career adaptabilities.
 Australian Journal of Career Development, 24(3), 166-172.
- 68. Jacobs, J. E., Finken, L. L., Griffin, N. L., & Wright, J. D. (1998). The career plans of science-talented rural adolescent girls. *American Educational Research Journal*, 35(4), 681-704.
- 69. Jacobowitz, T. (1983). Relationship of sex, achievement, and science self-concept to the science career preferences of black students. *Journal of Research in Science Teaching*, 20(7), 621-628.
- 70. Jansen, M., Schroeders, U., Lüdtke, O., & Marsh, H. W. (2019). The dimensional structure of students' self-concept and interest in science depends on course composition. *Learning and Instruction*, 60, 20-28.
- 71. Kager, E. (2015). Effects of participation in a STEM camp on STEM attitudes and anticipated career choices of middle school girls: A mixed methods study (Doctoral dissertation, Ohio University).

- Kang, J., Hense, J., Scheersoi, A., & Keinonen, T. (2019). Gender study on the relationships between science interest and future career perspectives.
 International Journal of Science Education, 41(1), 80-101.
- 73. Kang, J., & Keinonen, T. (2017). The effect of inquiry-based learning experiences on adolescents' science-related career aspiration in the Finnish context. *International Journal of Science Education*, 39(12), 1669-1689.
- 74. Kang, J., & Keinonen, T. (2018). The effect of student-centered approaches on students' interest and achievement in science: Relevant topic-based, open and guided inquiry-based, and discussion-based approaches. *Research in Science Education*, 48(4), 865-885.
- 75. Kaya, D., & Bozdag, H. C. (2016). Resources of Mathematics Self-Efficacy and Perception of Science Self-Efficacy as Predictors of Academic Achievement. *European Journal of Contemporary Education*, 18(4), 438-451.
- 76. Kier, M. W., Blanchard, M. R., Osborne, J. W., & Albert, J. L. (2014). The development of the STEM career interest survey (STEM-CIS). *Research in Science Education*, 44(3), 461-481.
- 77. Kiraz, A., Güneyli, A., Baysen, E., Gündüz, Ş., & Baysen, F. (2010). Effect of science and technology learning with foreign langauge on the attitude and success of students. *Procedia-Social and Behavioral Sciences*, 2(2), 4130-4136.
- 78. Koul, R. B., Fraser, B. J., Maynard, N., & Tade, M. (2018). Evaluation of engineering and technology activities in primary schools in terms of learning

environment, attitudes and understanding. *Learning Environments Research*, 21(2), 285-300.

- 79. Kutch, M. (2011). Integrating science and mathematics instruction in a middle school STEM course: The impact on attitudes, career aspirations and academic achievement in science and mathematics. Wilmington University (Delaware).
- 80. LaForce, M., Noble, E., & Blackwell, C. (2017). Problem-based learning (PBL) and student interest in STEM careers: The roles of motivation and ability beliefs. *Education Sciences*, 7(4), 92.
- 81. 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(5), 641-657.
- 82. Larry, T. A. (2017). Perceived Self-Efficacy and Student-Teacher Relationships Among Diverse Title I Students' Achievement in Science.
- 83. Leaper, C., Farkas, T., & Brown, C. S. (2012). Adolescent girls' experiences and gender-related beliefs in relation to their motivation in math/science and English. *Journal of Youth and Adolescence*, 41(3), 268-282.
- 84. Lee, E. (2017). Effects of South Korean high school students' motivation to learn science and technology on their concern related to engineering. *Educational Sciences: Theory & Practice*, 17(2).

- 85. Lent, R. W., Brown, S. D., Brenner, B., Chopra, S. B., Davis, T., Talleyrand, R., & Suthakaran, V. (2001). The role of contextual supports and barriers in the choice of math/science educational options: A test of social cognitive hypotheses. *Journal of Counseling Psychology*, 48(4), 474.
- 86. Lie, R., Guzey, S. S., & Moore, T. J. (2019). Implementing engineering in diverse upper elementary and middle school science classrooms: Student learning and attitudes. *Journal of Science Education and Technology*, 28(2), 104-117.
- 87. Liu, Y. H., Lou, S. J., & Shih, R. C. (2014). The investigation of STEM self-efficacy and professional commitment to engineering among female high school students. *South African Journal of Education*, 34(2).
- 88. Luo, T., Wang, J., Liu, X., & Zhou, J. (2019). Development and application of a scale to measure students' STEM continuing motivation. International Journal of Science Education, 41(14), 1885-1904.
- Maiorca, C., Roberts, T., Jackson, C., Bush, S., Delaney, A., Mohr-Schroeder, M. J., & Soledad, S. Y. (2020). Informal learning environments and impact on interest in STEM careers. *International Journal of Science and Mathematics Education*, 1-20.
- 90. Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J., Abdelfattah, F., Leung, K. C., ... & Parker, P. (2013). Factorial, convergent, and discriminant validity of timss math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Educational Psychology*, 105(1), 108.

- 91. Mativo, J. M., Womble, M. N., & Jones, K. H. (2013). Engineering and technology students' perceptions of courses. *International Journal of Technology and Design Education*, 23(1), 103-115.
- 92. Mau, W. C., Chen, S. J., & Lin, C. C. (2019). Assessing high school student's STEM career interests using a social cognitive framework. *Education Sciences*, 9(2), 151.
- 93. Miller, M. (2006). Science self-efficacy in tenth grade Hispanic female high school students.
- 94. Mohd Shahali, E. H., Halim, L., Rasul, M. S., Osman, K., & Mohamad Arsad, N. (2019). Students' interest towards STEM: a longitudinal study. *Research in Science & Technological Education*, 37(1), 71-89.
- 95. Mohtar, L. E., Halim, L., Rahman, N. A., Maat, S. M., Iksan, Z. H., & Osman, K. (2019). A model of interest in stem careers among secondary school students. *Journal of Baltic Science Education*, 18(3), 404.
- 96. Nam, Y., Lee, S. J., & Paik, S. H. (2016). The impact of engineering integrated science (EIS) curricula on first-year technical high school Students' attitudes toward science and perceptions of engineering. *EURASIA Journal of Mathematics, Science and Technology Education*, 12(7), 1881-1907.
- 97. Navarro, R. L., Flores, L. Y., & Worthington, R. L. (2007). Mexican American middle school students' goal intentions in mathematics and science: A test of social cognitive career theory. *Journal of Counseling Psychology*, 54(3), 320.

- 98. Novodvorsky, I. (1993). Development of an instrument to assess attitudes toward science.
- 99. Nugent, G., Barker, B., Lester, H., Grandgenett, N., & Valentine, D. (2019).
 Wearable textiles to support student STEM learning and attitudes. *Journal of Science Education and Technology*, 28(5), 470-479.
- 100. Nuhoğlu, H. (2008). The development of an attitude scale for science and technology course. *Elementary Education Online*, 7(3).
- 101. OECD (2009), PISA 2006 Technical Report, PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264048096-en.
- 102. Oh, Y. J., Jia, Y., Lorentson, M., & LaBanca, F. (2013). Development of the educational and career interest scale in science, technology, and mathematics for high school students. *Journal of Science Education and Technology*, 22(5), 780-790.
- 103. Olson, J.F., Martin, M.O., & Mullis, I.V.S. (2008). TIMSS 2007 Technical Report. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College
- 104. Özcan, H., & Koca, E. (2019). The impact of teaching the subject "pressure" with STEM approach on the academic achievements of the secondary school 7th grade students and their attitudes towards STEM. *Egitim ve Bilim*, 44(198).
- 105. Özcan, H., & Koca, E. (2019). Turkish adaptation of the attitude towards STEM scale: A validity and reliability study.

- 106. Pierce, R., Stacey, K., & Barkatsas, A. (2007). A scale for monitoring students' attitudes to learning mathematics with technology. *Computers & Education*, 48(2), 285-300.
- 107. Prieto, E., & Dugar, N. (2017). An enquiry into the influence of mathematics on students' choice of STEM careers. *International Journal of Science and Mathematics Education*, 15(8), 1501-1520.
- 108. Prieto-Rodriguez, E., Holbrook, A., Bourke, S., O'Connor, J., Page, A. (2015). Identification and Development of Strategies for Increasing Engineering Enrolments. 10.13140/RG.2.1.4861.5120.
- 109. Potvin, P., & Hasni, A. (2014). Analysis of the decline in interest towards school science and technology from grades 5 through 11. *Journal of Science Education* and Technology, 23(6), 784-802.
- 110. Rice, L., Barth, J. M., Guadagno, R. E., Smith, G. P., & McCallum, D. M. (2013).
 The role of social support in students' perceived abilities and attitudes toward math and science. *Journal of Youth and Adolescence*, 42(7), 1028-1040.
- 111. Richardson, S. S. (2016). The effect of an integrated STEM course on middle school students' interest and career aspirations in STEM Fields (Doctoral dissertation, University of Kansas).
- 112. Rochera, M. J., Merino, I., Oller, J., & Coll, C. (2019). Children's and Adolescents' Specific Interest in Science and Technology, Participation in Out-of-School

Activities and Inclination to Become Scientists. *Journal of Science Education and Technology*, 28(4), 399-413.

- 113. Roller, S. A., Lampley, S. A., Dillihunt, M. L., Benfield, M. P., Gholston, S. E., Turner, M. W., & Davis, A. M. (2020). Development and Initial Validation of the Student Interest and Choice in STEM (SIC-STEM) Survey 2.0 Instrument for Assessment of the Social Cognitive Career Theory Constructs. *Journal of Science Education and Technology*, 29(5), 646-657.
- 114. Romine, W., Sadler, T. D., Presley, M., & Klosterman, M. L. (2014). Student Interest in Technology and Science (SITS) survey: development, validation, and use of a new instrument. *International Journal of Science and Mathematics Education*, 12(2), 261-283.
- 115. Schreiner, C., & Sjøberg, S. (2004). Sowing the seeds of ROSE: background, rationale, questionnaire development and data collection for ROSE (The Relevance of Science Education): a comparative study of students' views of science and science education. http://urn. nb. no/URN: NBN: no-14449.
- 116. Shin, S., Rachmatullah, A., Roshayanti, F., Ha, M., & Lee, J. K. (2018). Career motivation of secondary students in STEM: a cross-cultural study between Korea and Indonesia. *International Journal for Educational and Vocational Guidance*, 18(2), 203-231.

- 117. Sırakaya, M., Alsancak Sırakaya, D., & Korkmaz, Ö. (2020). The impact of STEM attitude and thinking style on computational thinking determined via structural equation modeling. *Journal of Science Education and Technology*, 29, 561-572.
- 118. Smist, J.M. (1993). General chemistry and self-efficacy. Chicago, IL: National Meeting of the American Chemical Society.
- 119. So, W. W. M., Chen, Y., & Chow, S. C. F. (2020). Primary school students' interests in STEM careers: how conceptions of STEM professionals and gender moderation influence. *International Journal of Technology and Design Education*, 1-21.
- 120. Soobard, R., & Rannikmäe, M. (2014). Upper Secondary studentsself-perceptions of both their competence in problem solving, decision making and reasoning within science subjects and their future careers. *Journal of Baltic Science Education*, 13(4), 544.
- 121. Staus, N. L., Falk, J. H., Penuel, W., Dierking, L., Wyld, J., & Bailey, D. (2020). Interested, disinterested, or neutral: Exploring STEM interest profiles and pathways in a low-income urban community. *Eurasia Journal of Mathematics, Science and Technology Education*, 16(6), em1853.
- 122. Staus, N. L., Lesseig, K., Lamb, R., Falk, J., & Dierking, L. (2020). Validation of a measure of STEM interest for adolescents. *International Journal of Science and Mathematics Education*, 18(2), 279-293.

- 123. Suprapto, N. (2016). Students' attitudes towards STEM education: Voices from Indonesian junior high schools. *Journal of Turkish Science Education*, 13(special), 75-87.
- 124. Tatar, N., Yıldız, E., Akpınar, E., & Ergin, Ö. (2009). A Study on Developing a Self Efficacy Scale towards Science and Technology. *Eurasian Journal of Educational Research (EJER)*, (36).
- 125. Tran, Y. (2018). Computer programming effects in elementary: Perceptions and career aspirations in STEM. *Technology, Knowledge and Learning*, 23(2), 273-299.
- 126. Turner, S. L. (2000). Native American adolescent career development: Initial validation of a computerized career exploration and assessment instrument (pp. 1-107). University of Missouri-Columbia.
- 127. Turner, S. L., Joeng, J. R., Sims, M. D., Dade, S. N., & Reid, M. F. (2019). SES, gender, and STEM career interests, goals, and actions: A test of SCCT. *Journal* of Career Assessment, 27(1), 134-150.
- 128. Turner, S., & Lapan, R. T. (2002). Career self-efficacy and perceptions of parent support in adolescent career development. The career development quarterly, 51(1), 44-55.
- 129. Turner, S. L., & Lapan, R. T. (2003). The measurement of career interests among at-risk inner-city and middle-class suburban adolescents. *Journal of Career Assessment*, 11(4), 405-420.

- 130. Tyler-Wood, T., Ellison, A., Lim, O., & Periathiruvadi, S. (2012). Bringing up girls in science (BUGS): The effectiveness of an afterschool environmental science program for increasing female students' interest in science careers. *Journal of Science Education and Technology*, 21(1), 46-55.
- 131. Tyler-Wood, T., Knezek, G., & Christensen, R. (2010). Instruments for assessing interest in STEM content and careers. *Journal of Technology and Teacher Education*, 18(2), 345-368.
- 132. Uçak, E., & Bağ, H. (2012). Elementary school pupils' self efficacy towards science and technology lesson. *Journal of Baltic Science Education*, 11(3), 203.
- 133. Uitto, A., Juuti, K., Lavonen, J., Byman, R., & Meisalo, V. (2011). Secondary school students' interests, attitudes and values concerning school science related to environmental issues in Finland. *Environmental Education Research*, 17(2), 167-186.
- 134. Unfried, A., Faber, M., Stanhope, D. S., & Wiebe, E. (2015). The development and validation of a measure of student attitudes toward science, technology, engineering, and math (S-STEM). *Journal of Psychoeducational Assessment*, 33(7), 622-639.
- 135. Unlu, Z. K., Dokme, I., & Veli, U. (2016). Adaptation of the science, technology, engineering, and mathematics career interest survey (STEM-CIS) into Turkish. *Eurasian Journal of Educational Research*, 16(63).

- 136. Uysal, N. K., & Arıkan, Ç. A. (2018). Measurement invariance of science selfefficacy scale in PISA. *International Journal of Assessment Tools in Education*, 5(2), 325-338.
- 137. van Aalderen-Smeets, S. I., Walma van der Molen, J. H., & Xenidou-Dervou, I.
 (2019). Implicit STEM ability beliefs predict secondary school students' STEM self-efficacy beliefs and their intention to opt for a STEM field career. *Journal of Research in Science Teaching*, 56(4), 465-485.
- 138. Wang, M. T., Chow, A., Degol, J. L., & Eccles, J. S. (2017). Does everyone's motivational beliefs about physical science decline in secondary school?:
 Heterogeneity of adolescents' achievement motivation trajectories in physics and chemistry. *Journal of Youth and Adolescence*, 46(8), 1821-1838.
- 139. Wang, M. T., Fredricks, J. A., Ye, F., Hofkens, T. L., & Linn, J. S. (2016). The math and science engagement scales: Scale development, validation, and psychometric properties. *Learning and Instruction*, 43, 16-26.
- 140. Watt, H. M., Bucich, M., & Dacosta, L. (2019). Adolescents' motivational profiles in mathematics and science: associations with achievement striving, career aspirations and psychological wellbeing. *Frontiers in Psychology*, 10, 990.
- 142. Weinberg, A. E., Basile, C. G., & Albright, L. (2011). The effect of an experiential learning program on middle school students' motivation toward mathematics and science. *RMLE Online*, 35(3), 1-12.

- 142. White, J. L., & Mitchell, S. K. (2013). Career certainty and persisting interest in STEM: An analysis of underrepresented groups. *Journal of Women and Minorities in Science and Engineering*, 19(1).
- 143. Yaşar, Ş., & Anagün, Ş. S. (2009). Reliability and validity studies of the science and technology course scientific attitude scale. *Journal of Turkish Science Education*, 6(2), 43-54.
- 144. Zhou, S. N., Zeng, H., Xu, S. R., Chen, L. C., & Xiao, H. (2019). Exploring changes in primary students' attitudes towards science, technology, engineering and mathematics (STEM) across genders and grade levels. *Journal of Baltic Science Education*, 18(3), 466.

Curriculum Vitae

Name:	Meagan, Koufis
Post-secondary Education and Degrees:	University of Waterloo Waterloo, Ontario, Canada 2013-2018 B.Sc.
	The University of Western Ontario London, Ontario, Canada 2019-2021 M.A.
Honours and Awards:	SSHRC Canada Grad Scholarship 2020-2021
Related Research Experience	Graduate Research Coordinator University of Western Ontario 2019-2021
	Undergraduate Research Assistant Child & Adolescent Neuropsychology Lab, Waterloo, ON 2016-2019
	Cogmed Research Assistant Cambridge Memorial Hospital, Cambridge, ON 2016-2018
Related Work Experience	Psychology Intern Thames Valley District School Board, London, ON 2020 - 2021
	Intake Counsellor Homewood Health Inc., Guelph, ON 2019
	Child & Youth Worker Avalon Specialized Services, Guelph, ON 2017 - 2018