Summer 2015

Priming Middle School Females' Engagement in Engineering and Technology

Mary Bonk Isaac
Old Dominion University

PRIMING MIDDLE SCHOOL FEMALES' ENGAGEMENT IN ENGINEERING AND TECHNOLOGY

By

Mary Bonk Isaac
B.S. May 1982, Union College
M.A.T. June 2011, North Carolina A & T

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY IN EDUCATION

OCCUPATIONAL AND TECHNICAL STUDIES

OLD DOMINION UNIVERSITY
August 2015

Approved by:

_____

C. Tomovic (Co-Chair)

_____

D. Draper (Co-Chair)

_____

T. Paredes. (Member)

ABSTRACT

PRIMING MIDDLE SCHOOL FEMALES' ENGAGEMENT IN ENGINEERING AND TECHNOLOGY

Mary Bonk Isaac
Old Dominion University, 2015
Co-Chairs: Dr. Cynthia Tomovic
Dr. Darryl Draper

The purpose of this study was to validate data collected between 2009-2013 via an instrument used to assess middle school girls' identification with and interest in engineering and technology following the Society of Women Engineers' WOW events. Recognizing the importance of measuring the impact of such mentored E & T activities on young females' attitudes about and interest in STEM, and more specifically, engineering and technology, SWE outreach experts developed a participant survey based on event objectives and domain knowledge, but never rigorously examined whether it produced statistically valid data that could be used to draw inferences about girls' propensity to enter the field, essentially a priming factor for future engagement.

Using exploratory and confirmatory factor analysis, the survey data evaluated in this study (N = 332) successfully validated a four factor latent construct, albeit not precisely as proposed with respect to item loading. Longitudinal use of such a validated tool could provide reliable data to better predict female engagement. It also establishes a jumping-off point for additional discourse and research on the effects of mentored E & T activities on female engagement in male-dominated career fields such as engineering and technology.

This dissertation is dedicated to my dad, Gary Bonk, who instilled in me a curiosity for life and learning, an intolerance for willful ignorance, and a joy for designing, building and fixing things that never cry. Without you pushing me to beat my own path in the world, I probably never would have become a machinist in the first place, leading me back to your field of engineering. In my mind's eye, I can see you and mom shaking your head when I told you I was going to start my doctorate at age 56, much as you did when I told you I was going to attend engineering college fulltime as a single mom at age 24. I love that you never told me what I should do, instead helping to make it easier along the way. You will forever be an important part of who I am.

XOXOXOXO

Mimi

ACKNOWLEDGEMENTS

This work is a labor of passion and persistence that began with a desire to teach and is ending with a solid love for research. Thanks to the many friends, family, and colleagues who supported me along the way and who will continue to support me as I continue to stretch my scholarly wings afterwards, especially my husband Nick without whom I would not have been able to read, write, and analyze for days at a time while he brought me water and food, massaged my shoulders, and walked the dogs.

Thanks also to Dr. Vince Childress at North Carolina A & T University who made me believe that I could do research, which helped convince me to keep going for my doctorate; special thanks to the late Dr. Mara Wasburn of Purdue University whose untimely death at age 71 revealed through her obituary that she had earned her doctorate at age 57, inspiring me to apply at just about the same age.

My gratitude is endless towards the Society of Women Engineers, especially Randy Freedman for sharing the secondary data that enabled us to take the first step towards validating our outreach efforts with middle school girls and my SWEsters for offering me a community where I can indulge my passion of drawing more young women into engineering and technology careers.

Finally, I want to thank my dissertation committee, especially Dr. Cindy Tomovic, for keeping me focused and offering the right balance of encouragement and pushback to enable me to carve something huge down into something manageable. I could not have done it without you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

At 4.6% of the management and professional workforce, less than 2% of all female workers, and less than 20% of the engineering and technology (E & T) workforce, female engineers and technologists remain underrepresented in the United States, according to data from the U. S. Bureau of Labor Statistics' *Women in the Workforce: A Databook* (2013). This has not changed substantively over the last 30-plus years (Hill, Corbett, & St. Rose, 2010) in spite of prolific research and interventions to change that.

These interventions have included modifying pedagogy specifically for females, such as curriculum prepared for the Department of Public Instruction in North Carolina (Childress, 2003) or recommendations to move away from "smash and crash" projects in technology and engineering classrooms (McCarthy, 2009). Sheryl Sorby introduced a first year undergrad engineering course focused on spatial skills to work at improving low-performing incoming freshmen who were mainly female (2009). At one of Finland's teaching schools, all in-service teachers are required to take a course in technology education in which teachers are familiarized with gender neutral projects intended to engage both, but especially females who are also relatively underrepresented in male-dominant occupations in Finland (Virtanen & Ikonen, 2009).

Post-secondary research examining engagement and persistence of post-secondary STEM academia students and faculty focuses on undergraduates (Hug, Jurow, & Chi, 2011; Holmes, Redmond, & Thomas, 2012; Buday, Stake, & Peterson, 2012), graduate students (Fried & MacCleave, 2009), and faculty (Ceci, Ginther, Kahn, & Williams, 2014; Opare, 2012). Although research into career mentoring as an

intervention really began with Kathy Kram's *Mentoring at Work* in 1985, the focus was on individuals already in their careers (Scandura, 1992).

Substantial research on the effects of mentoring females prior to choosing an E & T career (undergraduate level and below) did not begin appearing until well after 2000 (Betz & Sekaquaptewa, 2012; Bystydzienski & Brown, 2012; Crisp & Cruz, 2010; Weber K. , 2011; Ware & Stein, 2013; Poor & Brown, 2011; Marshall, Lawrence, & Peugh, 2013; Brand & Kasarda, 2014), with a primary emphasis on undergraduates. Concurrently, research on mentoring in E & T careers accelerated due to federal funding opportunities associated with broadening participation in engineering (Amelink, 2008; Dworkin, Maurer, & Schipani, 2012; Eby, Allen, Evans, Ng, & DuBois, 2008; Kram & Ragins, 2008).

By the middle of the first decade of the new century, in-school K-12 engineering and technology programs such as Project Lead the Way (founded in 1986) were well-entrenched in many high schools across the United States and on their way into lower grades (Project Lead The Way, Inc., 2014). Such programs offer opportunities to better prepare students for post-secondary engineering and technology curriculum by sharing domain knowledge in authentic design and construct applications, which has been demonstrated to drive female engagement (Bystydzienski & Brown, 2012). Other secondary in-school opportunities for such exposure include non-PLTW curriculum such as technology education, introductory engineering or physics, which are often elective.

The rigidity of curriculum structure within the formal classroom setting, however, limits E & T exposure to those students who have already demonstrated some initiative in signing up for the particular course and excludes those who do not. Assuming the gender

ratio is near the same as it will be in E & T colleges and the work force, the class makeup is likely 75-80% male, 20-25% female. What about the other 75-80% females? Where else can they get non-judgmental exposure to expert engineering and technology domain knowledge in authentic design and construct applications? Engineering and technology involve 'things', which can neither be nurtured nor persuaded, and are inherently stereotyped as masculine, which translates most often to "male" (Bem, 1981; Spence & Helmreich, 1978).

One avenue includes structured extra-curricular after school programs such as Odyssey of the Mind, FIRST robotics, or Science Olympiads, which generally receive financial and human resources from the school and community via teachers and parents. National organizations which focus on less formal, recurring, out-of-school programs for females include Girl Scouts of the USA, Girls, Inc., Techbridge, and SWE, among others. Since 2007, the Society of Women Engineers (SWE) has co-sponsored nearly two dozen large (> 100 participants) one-day mentored events known as *Wow! That's Engineering!?* in many communities across the United States for middle school girls and the adults who influence them.

Society of Women Engineers

Founded in 1950 with less than 100 female engineers and technologists, the professional society now numbers greater than 30,000 members (Society of Women Engineers, 2014), with 75 corporate partners (SWE Corporate Partnership Council, 2014), and numerous professional alliances in the domain of advocacy for females in

engineering and technology. Not all members are engineers. Many are STEM professionals committed to the mission of engaging more females.

Although they have only the occupation 'Engineer' in their name and mission statement, they inclusively recognize that engineers and technologists are often two integrally connected sides of the same coin, and have strong alliances with organizations such as the National Coalition of Women in Technology (NCWIT). Future member recruitment to this blended community of practice of female engineers and technologists is fostered through the engineering and technology-related activities and programs developed by SWE's K-12 outreach community, based on best practices drawn from various local chapters or communities known as *sections*.

SWE outreach. Community outreach is an important facet of SWE's portfolio of member services (as for most professional societies) because it ties intimately to their mission of advocacy – for girls who can and will become engineers and technologists and 21st century women who understand their important role as "passers of the baton", who want to make a difference to future girls. Advocating in their communities in support of SWE's mission engages the largest single segment of SWE members, nearly one in five (Society of Women Engineers, 2013). Through SWE's mission of advocacy, female engineers and technologists can find a sense of community and purpose in sharing their knowledge about being a female engineer or technologist, with a heavy emphasis on *doing* engineering and technology.

Role model resources. Local SWE outreach events introduce young females to engineering and technology-related concepts and tasks under the guidance of predominantly female engineers and technologists from local SWE sections, at various

stages of their careers from high school through to retirement. The power of these role models is that they diverge from sex-stereotyped norms, which further challenges girls' cognitive development (Bandura & Bussey, 1999; Hill, Corbett, & St. Rose, 2010) and helps them transition from a "pink frilly dress" (Halim, Ruble, & Amodio, 2011, p. 935) perspective of female-ness to something less stereotyped.

Signature events. Over the years, the Society of Women Engineers has introduced many programmatic outreach activities for girls, predominantly to share knowledge and facilitate its members doing outreach to engage more females as engineers and technologists. In the last 10 years, several programs have been "branded" by SWE, that is, a marketing plan was developed and implemented, dedicated funding for corporate donors was earmarked, and in general, resources were aligned and prioritized to support the programs, including program curriculum and assessment. These are known as signature events or programs; they provide SWE brand recognition.

Wow! That's Engineering?!. One such event, known as Wow! That's Engineering?! (WOW), began in 2007 as a replicable middle school immersion event where local cadres of role models and volunteers used templates and earmarked funding to plan and deliver 6-8 hours of engineering and technology activities and guidance to large groups of females, tentatively 100 or more. The templates standardized processes and resources for the program while allowing flexibility in activities, venue, etc. Over the years, SWE has also added a concurrent session for parents and educators.

Sections submit applications to host such an event and are selected using a criterion-based rubric that is evaluated annually by the SWE outreach community. They receive funding, planning, and administrative assistance from the SWE outreach

community beginning 3-6 months before the event. Typically, one to three sections are awarded grants annually to host the event. Nearly two dozen such events have been conducted since the program's creation in 2007, with an average attendance of 150 middle school girls, 100 parents and educators, and 40-50 volunteers and role models. A large percentage of volunteers serve as role models and are also members of SWE's outreach community.

Program assessment. Although well-over a dozen WOW events have been conducted since the program's inception, back in 2007 there was little practical research on the connection between doing K-12 outreach and girls' intent to pursue E & T or how to measure it. With four years between middle school and college, and another four until career entry, a lot can happen in eight or more years – anything career-related that a 10-14 year old might say they intended to do in 2009 does not necessarily have good predictive value for 2017 (Allwin & Krosnick, 1991; Bandura A. , 2006; Gottfredson, 1981), so what to measure needed to be thoughtfully determined.

Extra-curricular programs are generally categorized as low-stakes enrichment programs, which are at the whim of private donors or members, so program evaluation generally is not held to as high a standard as the formal program assessments demanded by No Child Left Behind legislation. Recognizing that it would take at least eight to ten years before participants would ultimately select their careers, SWE still needed a way to assess impact on program participants that might also be able to predict girls' follow through on their intent down the line.

Minimally, they knew it was important to use the events as an opportunity to open a dialogue with young females (and the adults who influence them) that would provide

enough information about their experiences for SWE to assess program effectiveness and direct training resources for role models and other volunteers.

Recognizing the importance of assessing the impact of such mentored E & T activities on young females' attitudes about and interest in STEM, and more specifically, engineering and technology, SWE developed a participant survey based on event objectives and domain knowledge within the SWE outreach community when it introduced the WOW program, but never rigorously examined whether it produced statistically valid data that could be used to draw inferences about girls' propensity to enter the field, essentially a priming factor. This study serves as that validation using extant data.

Purpose of the Study

The purpose of this study was to validate data collected between 2009-2013 via an instrument used to assess middle school girls' identification with and interest in engineering and technology following the Society of Women Engineers' WOW events. The survey (Appendix A) is intended to measure middle school girls' identification with and interest in engineering and technology as a precursor to their intent to pursue careers in engineering and technology, which aligns with the event's objectives.

Survey Origination

WOW Boston was the first event to use a survey to evaluate student interest and intent in 2009, using both a pre- (N = 78) and post-survey (N = 145) derived from a questionnaire developed and recommended for secondary students by Assessing Women and Men in Engineering (AWE, 2008), which was modified by volunteers for the SWE event to accommodate the event's objectives and pre-teen girls' attention spans (seven

pages to three-and-a-half). This version was used at four subsequent events with nearly 700 responses: Pittsburgh (*N* = 116), Cleveland (*N* = 33), San Diego (*N* = 145), and Arizona State University (N = *159*).

Current Version

Further revisions were made to the survey in 2011 in which several items were revamped, combined, or deleted in their entirety to better address the event's objectives. The current survey has been used at eight events, although data is only available for three, with 332 responses at a 99% response rate. The current survey contains a total of 24 items deemed relevant to the study, with variables as defined in Table 1.

Other Survey Usage

This survey is also the model for the participant survey used in a newer, larger, annual immersion event for middle school girls introduced in 2010 called *Invent it. Build it.* (IIBI). The data collection and analysis is administered by an external evaluation consultant under SWE's direction. The consultant did not complete any formal validation of the instrument at the time of adoption since the survey was for program evaluation only and had been in use for several years. They have been analyzing the annual data since 2011 to provide evaluative reports for internal review of the annual collaboration between industry (Exxon Mobil), girl-serving organizations (Girl Scouts of the USA), volunteers (SWE and affiliates), and content providers (WGBH-Design Squad) (Paulsen, 2014).

Table 1

Variables from Survey Items

Variables				Survey Item
Description	**Name**	**Type**	**Values**	
Gender	SEX	Categorical	1 = Female 2 = Male	*[a]
Race/ethnicity	ETH	Categorical	1 = White/Causasian 2 = Hispanic 3 = African American 4 = Asian American 5 = Native American/Alaskan 6 = Pacific Islander 7 = Mixed 8 = Did not report	1
Age	AGE	Interval	Select one 8-15, Other	2
Future work type importance (6 items)	FWK	Interval	1 - Very unimportant 2 - Unimportant 3 - Not sure 4 - Important 5 - Very important	3a-f
Knowledge of what engineers do (6 items)	EDO	Interval	1 = Strongly disagree 2 = Disagree 3 = Not sure 4 - Agree 5 - Strongly agree	4a-f
Interest level (2 items)	INT	pre/post-INT Interval	1 = Strongly disagree 2 = Disagree 3 = Neither agree nor disagree 4 - Agree 5 - Strongly agree	5a, b
Change in knowledge level (2 items)	KNO	pre/post-KNO Interval	1 = Strongly disagree 2 = Disagree 3 = Neither agree nor disagree 4 - Agree 5 - Strongly agree	5c, d
Psychoscial aspects (2 items)	PSA	Interval	1 = Strongly disagree 2 = Disagree 3 = Not sure 4 - Agree 5 - Strongly agree	5e, f
E&T self-efficacy change (4 items)	ESE	Interval	1 = Got worse 2 = I don't know 3 = Stayed the same 4 = Improved	6a-d[b]
Would you attend again?	REP	Categorical	1 = Yes 2 = No 3 = Maybe	7
Recommend to friends?	REC	Categorical	1 = Yes 2 = No 3 = Maybe	8

Note: Items were added, modified or deleted between 2008 and 2013, however the final items examined reflected only those common to all.

[a] Gender was not collected for all events but presumed to be female if not reported. [b] The survey version used after 2014 has five items but no data was collectded for pre-2014 events, so this item is not included.

Statement of the Problem

The survey used to collect participant feedback from WOW events was reported to have been designed to measure how well the events met their objectives, however, the extent to which the survey supports those objectives had never been rigorously validated. A secondary purpose of the survey was to provide an indication of the impact that such events might have on middle school females' interest in engineering and technology.

However, before that discussion could take place, the survey first had to be shown to produce reliable and valid data that supported SWE's objectives. Since the subject survey is also being used for other outreach events within and outside SWE that are intended to prime young females' interest in the engineering and technology fields, it is strategically important to know whether the event that the survey was developed for has been shown to make a difference in girls' longer term decision-making – why continue to measure, fund, or replicate if they are not valuable? – however, the construct of the survey is what this study examined in preparation for that.

Without the validation, the survey might be acceptable for low stakes program evaluation, but drawing inferences from, or making predictions based on un-validated data would be imprudent and not very credible. Notwithstanding that, middle school females' intent to engage in E & T careers would also require as a minimum two data sets in time, one at the beginning and one after respondents finally entered their chosen careers, which was not the case here. The problem to be evaluated in this study is whether the sample data validated the construct and measurement model built into the instrument used to collect the sample.

Research Questions

As a validation study, the primary research question asked was the degree to which data collected via the survey accurately and consistently measured the event objectives. Specifically,

1. To what extent did the survey provide an accurate and consistent measure of how WOW events facilitate a change in attitude about careers in engineering and technology?
2. To what extent did the survey provide an accurate and consistent measure of evidence that participants connected priorities about wants in their future work and life to the experiences of the volunteer role models they were interacting with?
3. To what extent did the survey provide an accurate and consistent measure of how the event fostered an expanded sense of community and a deeper understanding of what engineers do?
4. To what extent did the survey provide an accurate and consistent measure of participants' self-confidence and critical thinking skills following E & T-related activities?

The primary goal of the WOW immersion events are to improve middle school females' engagement in E & T. The event objectives were translated into survey objectives by the original survey developers, with items and scales contributing to the assessment of how well the event met those objectives. Individual survey results presume to indicate some level of engagement, with higher scores potentially implying higher propensity for engagement beyond the event itself and into future career.

The objectives that the Society of Women Engineers posit predict longer term engagement in E & T must first be accurately and reliably represented by items from the instrument for the data collected via the survey to be deemed valid. Based on the literature and theory, those objectives and their relationship to the items and scales defined in the survey itself can be represented in a factorial model or construct as shown in Figure 1.

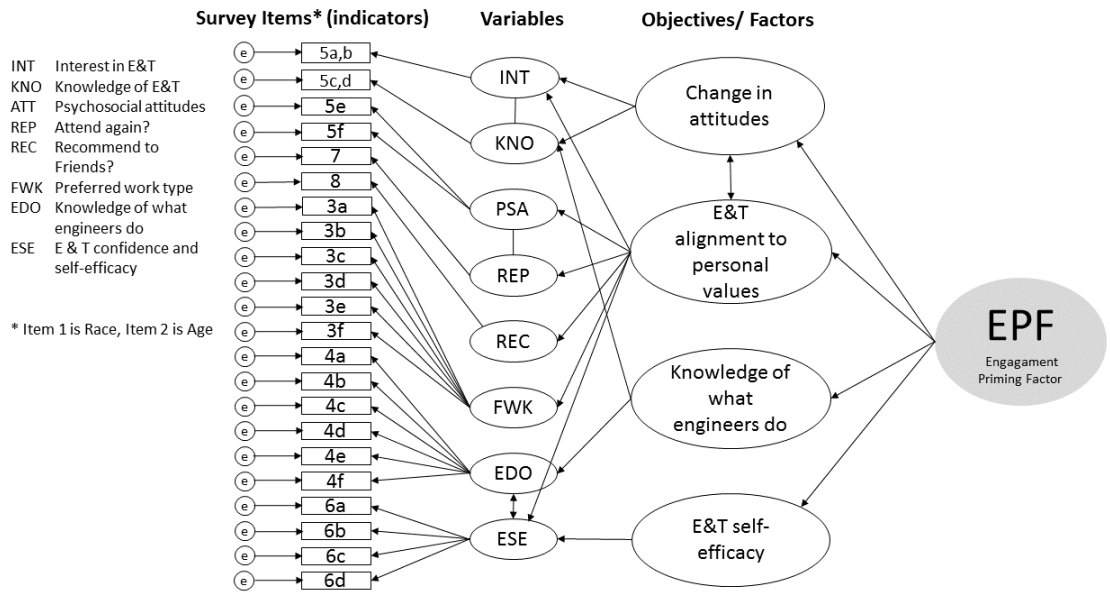

Figure 1. Preliminary structural/factorial model of Wow! That's Engineering?! survey instrument with proposed higher order latent factor engagement priming factor (EPF). Items/indicators represent responses to the statements in the survey. Variables are defined further in Table 1 (p. 9).

The goal of this study is to validate whether the collected data reliably confirms the proposed model which is based on the research questions stemming from the event objectives. Several tests can be performed without examining the data itself much beyond record counts, such as reviewing the survey itself for content validity, item format, and other observational notes that normally take place when first developing an assessment instrument. Any recommended changes to items or scales that would improve the validity of data collected in the future from this survey may or may not be implemented by the Society of Women Engineers, however, if they were to be implemented, a follow-on confirmatory factor analysis (CFA) would be needed to re-validate data collected by the modified instrument. In that sense, scale validation is a continuous process.

Significance of the Study

Doing engineering and technology with young females means giving them opportunities to solve problems in a hands-on way that involves the engineering design process and use of multiple technologies, the same as for males. The biggest difference with girls is the relative importance of same-sex role models for careers in which females are underrepresented. As a minority in the world of E & T, being able to work side by side with knowledgeable “experts” who look a lot more like them than most of their science or technology education teachers provides new data points in their perception of gender roles and where they fit in society.

In middle school, children’s sense of gender is still quite elastic (Gottfredson, 1981), so exposure to females in cross-gender occupational roles stretches them cognitively by opening their mind to alternative career possibilities. The role models who

volunteer at single day immersion events that feature *doing* engineering may never see some of these young females again, yet the event would likely be less effective in the ultimate mission of engaging more females without their presence and participation during these activities to inspire and pique the interest of middle school girls.

Validating data collected from an instrument designed to measure the degree to which mentored engineering and technology activities engage middle school females provides a robust assessment standard and tool for programs, curriculum, and training intended to encourage females to consider engineering and technology careers. Longitudinal use of such a tool could provide data that could be used to better predict female engagement.

Definition of Terms

The following definitions are provided to ensure uniformity and understanding of these terms throughout the study. The researcher developed all definitions not accompanied by a citation.

Confirmatory Factor Analysis (CFA) – A statistical method employed "when the goal is to test the validity of a hypothesized model of factors and those factors ' relationships to a set of observed variables" (Dimitrov, 2012, p. 95).

Cross-Gender Occupation – Any work-related role which is less socially acceptable for the opposite sex, for instance, engineer for a female or nurse for a male.

E & T – Engineering and technology

Exploratory Factor Analysis (EFA) – A statistical method that is "typically used when researchers do not have enough theoretical or empirical information to hypothesize how many factors underlie the set of observable variables and which

variables form which factor" (Dimitrov, 2012, p. 70).

Factor – "Unobservable variable that influences more than one observed measure and that accounts for correlations among these observed measures" (Brown, 2006, p. 13).

Femininity – A gender construct which ascribes certain behavioral and cognitive traits as attributable to females. Those traits embody expressivity and communality.

Gender Role Orientation – "The extent to which [a person] possesses characteristics of masculinity, femininity, or androgyny (separate from sexual orientation)" (Gurung, 2009).

Hands-on Activities – Design and build activities which involve three dimensional manipulation of materials or resources.

Immersion Event – An event that fully *immerses* the participants in the event's community domain, typically for more than just an instant. For this study, the event is Wow! That's Engineering?!, the community is female engineering and technology role models mentoring middle school females across the United States, and the domain is doing hands-on engineering and technology activities over a 4-8 hour period.

Instrument – Short for data collection instrument. Also known as a scale. The data collection instrument in this study is a paper-and-pencil survey.

Latent Variable or Factor – "In statistics, latent variables or hidden variables (from Latin: present participle of lateo ('lie hidden'), as opposed to observable variables), are variables that are not directly observed but are rather inferred (through a mathematical model) from other variables that are observed (directly

measured)" (Wikipedia, 2014).

Masculinity – A gender construct which ascribes certain behavioral and cognitive traits as attributable to males. Those traits embody instrumentality and agency.

Mentor – "Someone who teaches or gives help and advice to a less experienced and often younger person" (Merriam-Webster, 2014).

Role Model – A type of mentor who may passively or actively provide guidance and inspiration to mentees, typically in non-recurrent mentoring instances.

Scale – Another term for measurement assessment instrument, often used in psychological and sociological research. Scales may have sub-scales imbedded within them, particularly if there are latent factors in the theoretical construct being measured.

Structural Equation Modeling (SEM) – A statistical technique for building and testing statistical models, which are often causal models. Factor analysis is a special case of SEM.

Sex-typed Occupation - Any work-related role which is socially acceptable for the same sex, for instance, engineer for a male or nurse for a female.

STEM – An acronym often used in education and careers that encompasses science, technology, engineering, and math fields.

SWE – Society of Women Engineers

WOW – Wow! That's Engineering?!, a SWE-branded E & T event for middle school females

Assumptions

1. In developing the original student survey used to collect the primary data, the

Society of Women Engineers used a sound assessment design approach based on a literature review coupled with practical experience.

2. All of the secondary data in the study were kept completely anonymous and confidential.
3. All primary data collection waivers or consents are appropriately and securely maintained by the Society of Women Engineers.

Limitations

The limitations of this study were:

1. The survey was already operationalized and little was known about its development other than anecdotal accountings.
2. Data were self-reported assessments, which, while potentially problematic for high stakes evaluations, have acceptable fidelity for low stakes assessments. Self-reflection is *low stakes* assessment in that, typically, the only one who sees the results is the participant, while a college entrance assessment would be considered *high stakes* evaluation wherein participants may be more inclined to inflate their self-worth because it is of higher value (Allwin & Krosnick, 1991).
3. The study participants were predominantly females aged 10-14 (96%).
4. Primary data was collected via paper survey for students with data entry by contract into Survey Monkey. Secondary data is drawn from raw primary data. Errors may have been made during data entry.
5. Sampling was not totally random: event venues went through an application and selection process that limited participation to specific geographical areas

Delimitations

The delimitations of this study were:

1. The events in which primary data respondents participated were advertised locally and limited to 250 middle school females in specific geographical areas, so the sample may not be representative of subjects who were not able to register in time, or did not live in the geographic area.
2. Although the study began with a predominantly white female sample population, due to critical modifications made to the survey over the five year period, available data sets were reduced from eight to three, and the racial mix shifted.

Summary

Female engineers function in a world where females are 12 out of 100 people in the room (Hill, Corbett, & St. Rose, 2010) and that can be challenging to those who have a high need for belonging. Some are successful, some are not. Many of them share themselves, their coping mechanisms, and their experiences with younger women in hopes of encouraging them to enter the same field. Some have been doing it for a long time, some have just started. What they all have in common, however, is that they are communing about careers in a field that is dominated by men, where most have learned how to get along in a world where they are members of an out-group during the day, and expected to morph into in-group (including family, friends, etc.) by night.

Mentored engineering and technology activities provide opportunities for experienced mentors and role models to interact with middle school females in a sharing of knowledge that has an impact on the girls' attitudes about and interest in engineering and technology.

Chapter II provides a review of literature relevant to middle school female engagement in engineering and technology, including a detailed conceptual framework based on social, gender, and community of practice theories. Exploring the topic of mentoring females in cross-gender occupations begins to peel back layers related to stereotypes, gender role orientations, and bias, and how they all relate to engineering identity development and performance.

Chapter III describes the methodology and procedures used during the study. It details the validation process, from evaluation of the survey construct, content, and sampling methods to analysis of internal data consistency and overall reliability.

Chapter IV presents the findings of the study where the researcher addresses how well the data supported the four objectives: to 1) facilitate engineering and technology attitude change, 2) improve E & T efficacy, 3) encourage community and a deeper understanding of what engineers and technologists do, and 4) cognitively align future work priorities with E & T possibilities.

Chapter V discusses the summary and conclusions. Recommendations for future studies and next steps is provided, including suggestions to address obstacles to female engagement and appropriate interventions.

CHAPTER II

Literature Review

Based on the continued underrepresentation of females in the society-critical occupations of engineering and technology, it would be reasonable to assume that such interventions as SWE's WOW events or, Invent it. Build it., the U.S. Science and Engineering Festival, and For Inspiration and Recognition of Science and Technology (FIRST), among others, have had limited success over the last 30 years.

However, such an assumption is not supported by historical research that extolls the benefits of events which provide opportunities for girls to experiment with spatial, typically hands-on, activities (McCarthy, 2009; Weber, 2011) while receiving some form of beneficial mentorship from role models and speakers (Kram & Ragins, 2008; Ghosh & Reio, 2013; Eby, Allen, Evans, Ng, & DuBois, 2008; Dworkin, Maurer, & Schipani, 2012; Marshall, Lawrence, & Peugh, 2013).

Learning and practicing new skills in a supportive community with members who they can identify with is particularly relevant to females in traditionally male occupations, where they represent less than one in 20 undergraduate students and less than one in 12 employees in technology and engineering-based venues (Brand & Kasarda, 2014; Kekelis, Ancheta, & Countryman, 2005; Scandura, 1992; Stout, Dasgupta, Hunsinger, & McManus, 2011; Eisenhart, Bystydzienski, & Bruning, 2010). Engineering and technology involve 'things', which can neither be nurtured nor persuaded, and are inherently stereotyped as masculine, which translates most often to "male" (Bem, 1981; Spence & Helmreich, 1978). It is through social interactions with real-world role models who eschew or reject gender stereotypes that they are able to

resolve concerns they may have about fitting in to a culture where they are significantly in the minority.

Conceptual Framework

In the quest to validate a career-related survey used with pre-adolescent females, this study grounds itself in several different but related concepts or theories, most of which are based on social learning theory (Bandura, 1971). While not a theory per se, the concept that hands-on activities benefit females in particular (because of their relatively poorer spatial skills) is an integral part of SWE's outreach event objectives (Paulsen, 2014; Phelps, 2012).

Mentoring is a form of guidance (Kram, 1985) that springs from social learning theory and informs communities of practice (CoP) theory (Lave & Wenger, 1991), while mentoring females in cross-gender roles also incorporates various gender theories such as gender development (Bandura & Bussey, 1999), gender role (Eccles, 1987), gender schema (Bem & Lenney, 1976), and social role (Eagly, 1987). Eccles' achievement expectancy theory (Eccles, 2011) and Bandura's self-efficacy theory (Bandura, 2006) round out the theoretical foundations of this study examining an instrument that measures student engagement.

Social theories. In general, human living is a process of socialization, first within the small family unit, then, expanding outwards into the broader world and workforce with normal cognitive development and maturity. Most people have the capability to achieve a high enough level of competence at socialization to enable them to lead

fulfilling lives in pursuit of knowledge and satisfaction with classmates, family, and work colleagues.

Researchers who try to understand individual behaviors work from a psychological framework, while those who study group behaviors work from a sociological perspective. Those who bridge the two perspectives to understand how individual behaviors translate to collective behaviors, like Albert Bandura, are psycho-sociologists. They recognize the continuous interplay between learner, model, and contextual conditions.

This study of female engagement in a field stereotypically assigned to males draws heavily on Bandura's social learning theory (1971), social cognitive theory (1986), and self-efficacy theory (2006; 2006); Eccles' model of achievement-related choices (1994); and Lavé and Wenger's community of practice theory (1998). Bandura and Bussey's gender development and differentiation work (1999; 2004), along with social / gender role theory is covered under gender theories.

Social learning. How much students learn is a function of the quality of the instruction or guidance (Allen & Eby, 2003). Many researchers concur with Bandura that "most of the behaviors that people display are learned, either deliberately or inadvertently, through the influence of example" (1971, p. 5). He stresses the often subtle impact of role models who may not even be aware that they are being emulated and of learners who are not conscious that they are learning. It is Bandura's basic work on

understanding how humans develop and learn to self-regulate that is the basis for social learning.

In social learning theory, there are four key modeling processes that occur to effect learning: 1) attentional, 2) retentional, 3) motoric reproduction, and 4) reinforcement or motivational (1971, p. 8).

Attentional. Attentional processes include association, identification, and appeal. The first step towards engagement is attention. Association and identification include head nodding and verbalization from observers when listening to a speaker, both signs of a captive audience. Both male and female human babies observe attractive animal faces longer than unattractive ones (Quinn, Kelly, Lee, Pascalis, & Slater, 2008). Media (the fourth estate) and marketing are powerful examples of fields where gaining and keeping consumers' attention is the ultimate goal.

Retentional. Retentional processes are important because the observed behavior must be repeated in order to ensure it is learned. Retention is aided by modeling, wherein,

> by relying on the actions of knowledgeable models, a novice can behave appropriately in synagogues, in mosques, in saloons, at wedding ceremonies, [in school, at work,] and in countless other situations, without having to discover the acceptable conduct through shocked or pleased reactions to his unguided performances. (Bandura, 1971, p. 18)

This is particularly relevant when learning something new or when interacting in a different cultural milieu, such as being female in predominantly male work environment.

The task or concept to be learned must be encoded in such a way as to either leave a visual memory or image of what is being modeled or transmitted through verbal coding

or instructions. This was thought by Bandura to be most effective when coupled, because the verbal memory enables the learner to model what is written, while confirming their own performance against the image of an expert's product or performance. He also suggests that anticipated reinforcement or punishment can aid or harm ability to retain information or knowledge, which is also demonstrated in Steele's concept of stereotype threat as applicable both to gender and race (Spencer, Steele, & Quinn, 1999; Steele, 2010). Being explicitly told the stereotype that men are better mathematicians than women frequently introduces enough negative reinforcement stress to trigger avoidance or minimally inhibit performance for fear of fitting the stereotype.

Motoric reproduction. Motoric reproduction involves transferring thought to action, memory to operations. Practice makes perfect. Motoric reproduction builds muscle memory and is a critical piece to the modeling process – learners must have an opportunity to demonstrate what they learned authentically. This modeling process step equates to the hands-on building portion of the WOW events, and, like physical play, reinforces learning and advances cognitive development (Bjorklund & Brown, 1998).

Reinforcing and motivational. According to Bandura, a person "can acquire, retain, and possess the capabilities for skillful execution of modeled behavior, but the learning may rarely be activated into overt performance if it is negatively sanctioned or otherwise unfavorably received" (1971, p. 8). How others perceive behavior influences an individual's behavioral response both positively and negatively, depending in large part on the individuals themselves and the context of the interaction. Ultimately, the best form of reinforcement for social learning to occur is self-reinforcement, and Bandura

indicates that “the highest level of autonomy is achieved when individuals regulate their own behavior by self-evaluative and other self-produced consequences” (1971, p. 24).

Bandura (1995) also discovered that extrinsic rewards and reinforcement, while of some value, do not predict learning, and that

> individuals who had been exposed to models favoring lenient standards of self-reinforcement were highly self-rewarding and self-approving for comparatively mediocre performances; conversely, persons who observed models adhering to stringent performance demands displayed self-denial and self-dissatisfaction for objectively identical accomplishments. (p. 30)

In other words, models with mediocre standards and indulgent reinforcement do not inspire high performance or autonomy, that is “human behavior is regulated to a large extent by *anticipated* consequences of prospective actions” (p. 36).

Bandura also proposes that “lack of matching behavior following exposure to modeling influences may result from either failure to observe the relevant activities, inadequate coding of modeled events for memory representation, retention decrements, motoric deficiencies, or inadequate conditions of reinforcement” (1971, p. 8), so it is important to ensure that scales provide opportunities to isolate the failure mode.

Social cognition. Social cognition models are similar theories, each of which “specifies a small number of cognitive and affective factors (`beliefs and attitudes') as the proximal determinants of behavior” (Sutton, 2002, p. 2). Perhaps the best known of these is social cognitive theory (Bandura, 1986). In social cognitive theory,

> sociostructural factors operate through psychological mechanisms of the self system to produce behavioral effects. Thus, for example, economic conditions,

> socioeconomic status, and educational and family structures affect behavior largely through their impact on people's aspirations, sense of efficacy, personal standards, affective states, and other self-regulatory influences, rather than directly. (2001, p. 15)

Through a triadic and bi-directional causal interplay between environment, behaviors, and the personal makeup of the individuals involved, social cognitive theory aids in understanding human functioning in a social setting. As social cognitive theory is a refinement of social learning theory, self-efficacy theory is a subset of social cognitive theory.

Self-efficacy. Self-efficacy is a form of self-regulation that has roots in agency, or the capacity to act. Bandura describes it as the belief one has in their ability to accomplish something (2006). It has little to do with intentions or actual competence, rather, it is a product of individuals interacting in and with the environment around them.

> Perceived efficacy plays a key role in human functioning because it affects behavior not only directly, but by its impact on other determinants such as goals and aspirations, outcome expectations, affective proclivities, and perception of impediments and opportunities in the social environment (Bandura, 1995, 1997). (Bandura, 2006, p. 309)

Individual perceived efficacy can be transformed into collective efficacy, as seen in the causes of Gandhi and Nelson Mandela. It can also be affected by social and cultural

constraints, whether real or mandated (Fernández-Ballesteros, Díez-Nicolás, Caprara, Barbaranelli, & Bandura, 2002).

Some studies suggest that highly efficacious people are more satisfied with and successful at following through on the learning process (Bandura, 2006; Hackett & Betz, 1981; Schunk, 1981; Zimmerman, 2000) and fear the unknown less (Bandura, 2006). People with high self-efficacy are open to more career experiences, prepare themselves better for more than one career option, and persist in their aspirations (Lent, Brown, & Hackett, 1994).

Expectancy value. Jacqueline Eccles and her colleagues constructed a model in the 1980s to explain the motivational factors underlying females' and males' educational and vocational decisions. Her research focuses on understanding gender-related differences in performance and achievement, particularly in fields where females are outnumbered. Her studies focus on choices, not deficits, and are related to Bandura's social cognitive theory that links personal factors with milieu and with behaviors in what she describes as "social roles and socialization processes in multiple contexts as key influences on the ontogeny of individual and group differences in expectancies and task values" (2011, p. 511).

Eccles' concept of subjective task value comprises "several . . . . other motivational beliefs, including anticipated interest and enjoyment likely to be experienced in; the attainment and utility value of; and the anticipated psychological, economic, and social costs of various possible task or activity choices" and assumes that "hierarchies of the subjective task values of various options are directly influenced by the immediate social context and the developmental stage of the individual making the

choice" (2011, p. 511). Decisions that girls make early on about the career possibilities open to them can narrow or broaden their options from a career perspective; mentors' attitudes about themselves and their roles as engineers in a predominantly male environment can also have a positive or negative motivational influence on their mentees' career decisions.

Eccles' research today examines ways in which to "change young women's beliefs that [physical, computer, and engineering sciences] fields are less appropriate for them or less compatible with their life goals and both their personal and social identities" (2011, p. 514), and is primarily focused on studies of self-efficacy and achievement, both considered masculine traits. Bona and Kelly did a study of n = 141 collegiate females and their results

> indicated that women who were pursuing careers in a male-dominated field had significantly higher confidence and interest in math and science and also had parents who modeled less-traditional gender roles than women who were aspiring to enter traditional gender careers. (2010, p. 123)

Fixed versus growth mindset. In their study of why females opt out of mathematics and occupations that are math-dependent, Good and Dweck (2012) learned that

> undergraduate students' perceptions of two factors in their math environment—the message that math ability is a fixed trait and the stereotype that women have less of this ability than men—worked together to erode women's, but not men's, sense of belonging in math. (p. 700)

The messaging about math competence as an innate ability that you either have or do not is representative of entity theory (Blackwell, Trzesniewski, & Dweck, 2007), or a *fixed mindset*, while incremental or *growth theory* proposes that competence can be developed and mediated through exposure to role models, stereotype awareness interventions, and usage. Like spatial abilities, mathematical competency can be developed, so that there is little difference between male and female performance. Having role models who encourage effort is critical.

Community of practice. Community of practice (CoP) theory (Wenger, 1998) informs the intrinsic motivation of females in engineering and technology to form and maintain communities which help them sustain a sense of belonging, such as the Society of Women Engineers. In addition to fulfilling a sense of belonging, communities of practice are collectives of individuals willingly doing the community's explicit and implicit work, which in this study is engaging more female engineers and technologists. In Bandura's words:

> People's beliefs that they can effect social change by working together is, to a large extent, grounded in the perceived self-efficacy of its members. One cannot easily create a strong collective force from members who are plagued by a profound sense of self-doubt. People of low efficacy see little point in attempting to exercise control or, if they try, they easily convince themselves of the futility of further effort should they encounter difficult obstacles. (Fernández-Ballesteros, Díez-Nicolás, Caprara, Barbaranelli, & Bandura, 2002, p. 112)

More experienced members of the CoP of female engineers and technologists coach or train less-experienced members of the CoP, who in turn reach out to others,

thereby deepening all members' sense of belonging to the larger community of practice of female engineers and technologists (Jackson, 2013). Community of practice theory is operationalized in informal mentoring communities such as the Society of Women Engineers, where more experienced members mentor other females, who in turn, recruit and coach new community members, moving them from legitimate peripheral participants (Lave & Wenger, 1991) at events similar to the one being studied here, to full and active community members.

In this study of the community of practice of female engineers and technologists, legitimacy is derived simply from being a female interested in engineering and technology, while peripherality enables aspiring or new members (middle school females at WOW events) to observe the female engineers in action, that is, for them to safely dabble in the domain's practices somewhat vicariously through its non-peripheral members with little risk and much reward. In order to accomplish that,

> peripheral participation must provide access to all three dimensions of practice: to mutual engagement with other members, to their actions and their negotiation of the enterprise, and to the repertoire in use. No matter how the peripherality of initial participation is achieved, it must engage newcomers and provide a sense of how the community operates. (Wenger, 1998, p. 100)

Gender theories. Since the focus of this study is a survey intended to measure female engagement in traditionally male occupations in engineering and technology, it is important for the reader to understand why, even though biological sex has been rendered inconsequential to performing engineering and technology work, gender and biological

sex continue to be synonymous and continue to be at issue in the engineering and technology workplace.

Gendered attitudes and behaviors are those conscious and unconscious attributes that proscribe social or cultural constraints on the activities, roles, and behaviors of one or more individuals because of their sex. In this sense, gender is a psychosocial concept rather than a biological categorization. Gendered attitudes are typically dyadic in nature, consistent with stereotypes. Highly gendered individuals view masculinity and femininity as a dichotomous measure of biological maleness and femaleness; they often do not understand that gender is a psychosocial construct that serves societal purposes, and as such, can be (and is often) manipulated, e.g. gender-differentiated product marketing strategies or gender discrimination, the latter also known as sexism. Gender theories encompass the concept of gender schemas, development and differentiation, identity, and social roles.

Gender schema. Sandra Bem introduced the concept of gender schema theory following her exploration of psychological androgyny in the mid to late 1970s (1981).

> Gender schema theory . . . . proposes that sex typing is derived, in part, from gender schematic processing, that is, from a readiness on the part of the individual to encode and to organize information—including information about the self—in terms of the cultural definitions of maleness and femaleness that constitute the society's gender schema. (p. 364)

Gender schema theory suggests that people fall into one of four categories with respect to their gender schema: either sex typed, that is, scores above the median for either masculine or feminine attributes, and below the median for the cross-sex

attributes; cross-sex typed, that is, scores above the median for cross-sex attributes and below the median for sex typed attributes; androgynous, with both sex-typed and cross-sex typed attribute scores above the median; and undifferentiated, with both scores below the median.

Bem asserted that schemas help individuals organize information about the world around them by functioning as "an anticipatory structure, a readiness to search for and to assimilate incoming information in schema-relevant terms" (1981, p. 355). She proposes that the child assimilates its gender schema, wherein the schema becomes a "prescriptive standard or guide (Kagan, 1964; Kohlberg, 1966), and self-esteem becomes its hostage" (p. 355). In fact, she implicates "heterosexuality subschemas" (p. 361) in driving gender-based schematic processing.

Gender differentiation and dichotomy is a critical component of the heterosexuality subschema, which is predominantly biologically triggered, but supported by social structures. Bem suggests that the strength of gender schema's effects relies on "its insistence upon the ubiquitous functional importance of the gender dichotomy" (1981, p. 363): if society deemed gender differences not functionally salient in any jurisdiction other than reproduction, gender-based stereotypes would soon lose their relevance and structural support.

Gender development and differentiation. Almost immediately following birth, children begin receiving information that helps them define who they are and, ultimately, who they want to be when they grow up. This information gets explicitly and implicitly transmitted through various vectors, starting with family, and continuing throughout their lives. By four or five years of age, children have typically begun to gender identify in

response to a need to understand the rules of the world around them and how they fit in, particularly as they expand their social network through school and play. They are typically unconscious to their role as a receiver in this process, as are the entities transmitting the gendered information often oblivious to themselves as transmitters. The child's gender development is somewhat involuntary, driven by biologics and genetics on one hand, and psychosocial constructs on the other.

Bandura and Bussey tell us that "gender development and functioning are socially situated, richly contextualized, and conditionally manifested. Entrenched institutional constraints, pervading normative structures, widespread symbolic modeling of gendered lifestyles, and intricate incentive systems are active players in the social construction of gender" (2004, p. 699). They also believed that gender development was a continually evolving psychosocial construct which follows the triadic reciprocal determinism model of social cognitive theory (1999). Children develop gender identities based on what they see going on in the world around them, particularly in families; adults further modify their constructs based on their personal experiences and motivations.

In these early years, parents and other caregivers provide the first gender models whose behaviors the child may choose to emulate or reject in their own behaviors. According to Bandura, "self-conceptions are formed through cognitive processing of gendered information conveyed by modeling, differential social evaluative reactions, and direct tutelage in interpersonal transactions within familial, educational, peer, mass media, and occupational social subsystems" (Bandura & Bussey, 2004, p. 695).

It is these self-conceptions that contribute to the formation of a gender identity. In their words,

> gender differentiation takes on added importance because many of the attributes and roles selectively promoted in males and females tend to [also] be differentially valued, with those ascribed to males generally being regarded as more desirable, effectual, and of higher status (Berscheid, 1993). (Bandura & Bussey, 1999, p. 676)

Gender identity. Bem's gender schema theory formed the basis for Janet Spence's gender identity theory (1984) which has significant implications for females in cross-gender roles. In her studies of the construct of gender, Spence suggests that "gender is one of the earliest and most central components of the self-concept and serves as an organizing principle through which many experiences and perceptions of self and others are filtered" (1984, p. 64). Similarly, Bandura and Bussey reference Kohlberg's cognitive-developmental theory in which "gender identity is postulated as the basic organizer and regulator of children's gender learning" (2004, p. 677).

Psychological androgyny. Constantinople (1973) was really the first researcher to rattle the masculinity-femininity (M-F) cage with her notion that role identity was where the concepts of masculinity and femininity were the least defined. She, like Sandra Bem and Janet Spence later, did not support the psychoanalytical approach of conceptualizing masculinity and femininity as binary, dependent measures. Bem introduced the concept of *psychological androgyny* as a gender construct that reflected a well-balanced gender identity in her 1974 study "The Measurement of Psychological Androgyny".

Individuals who scored above both masculine and feminine scale medians on her sex role instrument were deemed *psychologically androgynous*. High masculine and low feminine scores on her instrument, or vice versa, represent *sex-typed* perspectives, while

scores below the median for both are categorized as *undifferentiated* or lacking average levels of either masculine or feminine traits. Spence felt that substituting the terms expressivity and communality for the term ‘femininity’, and instrumentality or agency for the concept of ‘masculinity’ removed some of the stigma associated with incongruence between gender identity and biological reality (Spence & Buckner, 2000).

Female engagement with E & T. Bystydzienski and Brown (2012) examined the gendered ways in which females engage in engineering in the early stages, along with some of the areas that 132 young females from the National Science Foundation-sponsored Female Recruits Explore Engineering (FREE) project identified as lacking in terms of knowledge and quality of communications at different venues, such as career fairs, exhibits, websites, and others. The authors state:

> gender dualism is strongly reinforced in engineering, where the emphasis on the technical (read: masculine) remains paramount, even though, as some have demonstrated (Devine 1992; Faulkner 2000; Jorgenson 2002), there is a great deal of overlap between the technical/abstract and social/concrete aspects of engineering in practice. (p. 3)

There was a significant tendency for females to gravitate towards “areas and practices of engineering perceived as female-friendly [such as biomedical, industrial, and environmental engineering] and express concerns about balancing work and family lives” (2012, p. 2). The real harm in a tendency such as this is “that women can become marginalized in certain engineering areas that come to be labeled as inferior or ‘not real’ engineering, while the male-dominated fields continue to garner social prestige and higher remuneration (Foor and Walden, 2009)” (Bystydzienski & Brown, 2012, p. 4).

They also tended to *do gender*, that is selectively employ gendered behaviors to gain acceptance in their male-dominated work communities, sometimes stereotypically feminine, "while at other times acting like 'one of the boys' or adapting an 'anti-woman' approach" (p. 3).

The "Queen Bee" (Derks, Ellemers, Laar, & Groot, 2011, p. 519) is an extreme exemplar of this dual identity phenomenon, where a successful female participates in maintaining the gendered constructs in male-dominated organizations, rejecting her feminine side while at work, while retaining a self-identity as female. Kilianski and Rudman might also argue that their equivocal egalitarians do gender in that they want gender equality but do not want to either share their current gender role or give up any of the benefits of paternal protectionism (Kilianski & Rudman, 1998). As noted in all three cases, females do gender because it has been effective for them under the current social structure.

Gender or social roles. Eagly was one of the first social psychologists to study gender role congruity as related to prejudice (1987). Her concept of social role theory proposes that, in social contexts, individuals rely on *correspondent inference*, which is a form of essentialism in which individuals ascribe what a certain person does or how they behave in a social role to represent what everyone within that person's social group will do or how they will behave. It is "the basic psychological process that produces stereotypes of social groups that mirror the qualities that they play out in their social roles" (Eagly, Wood, & Diekman, 2000, p. 137).

Bandura and Bussey (1999) theorize that this reality results "more from cultural design than from biological endowment" (p. 676). Appropriate gender roles begin to

resolve at very early ages, and continue to develop throughout life. Gender socialization plays an important role in female engagement in science, technology, engineering, and math (STEM) fields, as it might also for male engagement in nursing or elementary level teaching. Gender roles are often so woven into a culture that it is difficult to divorce biological sexes from gendered individuals and their behaviors. As Goldin and Rouse (2000) confirmed in their study of hiring female orchestra musicians, bias disappears with gender differentiators: blind auditions ensure the absence of stereotypes triggered through images, voice, or name, resulting in significantly more women being hired.

When they ignore or underestimate the effects of context on behaviors that have previously been cognitively etched in their mind as solely attributable to one gender, such as a female CEO of a construction company or a man changing a baby's diaper on the counter or floor in the men's room, individuals may experience significant role incongruity, which triggers prejudice, typically followed by *system justification* (Jost & Banaji, 1994): *I'll bet she slept her way to the top* or *His wife must be at sick or away*, rather than *Wow, she must be good!* or *Nice to see a dad bringing his kid to work!* It is easier to construct a story that is plausible to the unbelieving mind than to change a belief.

Eagly's more recent study with Fetterolf (2011) examined the gender role attitudes of undergraduate women in an experiment which involved them envisioning themselves as married parents of a pre-school children under three different possible future selves scenarios – working full-time, working part-time, and not working at all, and to rate different aspects of their expectations regarding different social roles in the family unit. They state that "both men and women may believe that any career sacrifices

to focus on family demands should be made by mothers, especially to the extent that they implicitly associate women and family" (p. 84).

Social role theory is rooted in division of labor and gender hierarchy is based on that division of labor. As Schmitt and Wirth explain: "in other words, the division of labor according to gender leads to stereotypes that rationalize the division of labor" (2009, p. 431). Social role theory infers that it is the growing numbers of women in professional roles that have changed gender stereotypes because of the masculine behaviors required by these roles, resulting in women now being seen as more instrumental or agentic, also characterized as self-efficacious.

Gender roles "consist of *injunctive norms*, which are expectations about what people ought to do or ideally would do, as well as . . . . *descriptive norms*, which are expectations about what people actually do" (Eagly, Wood, & Diekman, 2000, p. 131). It is gender roles themselves (and people's implicit associations with them) which cause stereotypical behaviors to proliferate. Expectancy confirmation, that is, performing the role in satisfaction of expectations, is at the heart of stereotypes and why they are so tenacious. More specifically, it is the autonomic reaction to deviation from expectations that is at the core of stereotypes, the implicit response to incongruence either internally or externally. Females who work in a male-dominated occupation such as engineering are considered deviant from a social role perspective, however, from a gender role orientation perspective, they are often exactly where they should be – in an agentic, directed, facilitative role that capitalizes on both their masculine and feminine gender traits.

Obstacles to Female Engagement in E & T

While some progress has been made over the last 50 years in terms of female penetration into higher-paying engineering and technology occupations, more is needed, and faster, to address the expected workforce demands in the areas of engineering and technology, especially in computer science. Engaging the other half of the world's population in occupations that help to solve critical, real, dynamic problems has proven to be challenging for several reasons.

How to measure engagement is the first obstacle to be concerned with, especially related to the validation that is the core of this study. Other challenges comprise dealing with gendered messaging, including sexism and stereotypes; females' limited early exposure to E & T and spatial activities; achieving a sense of belongingness and self-efficacy; biological or genetic limitations; and having and making career choices.

Engagement indicators. Many existing engagement indicators are either criterion-based achievement rather than affective assessments or have insufficient or unknown validity. Some K-12 organizations measure engagement effect through grades or other performance-related standards that students must achieve (Ferro, DeWit, Wells, Speechley, & Lipman, 2013; Eisenhart, Bystydzienski, & Bruning, 2010) which is an objective evaluation of presumed learning, but does not say much about whether they are interested and engaged enough to pursue it further. Others measure interest in the fields of engineering and technology such as the survey examined here, however, pre-adolescent and adolescent interest and intent are not always directly translatable to action (Bandura, 1986; 2006), especially for females (Bandura & Bussey, 1999; Lent, Brown, & Hackett, 1994) who have choices (Eccles, 2011; Wang, Eccles, & Kenny, 2013).

Gendered messaging. Gendered messaging addresses not only the explicit symbolism of media, but also the more subtle messaging found in stereotypes and sexism.

Media. Perhaps, as the National Academy of Engineering (NAE) suggested in *Changing the Conversation* (2008), the mediocre and gendered messaging about the possibilities of engineering contributes to lack of engagement for both boys and girls. Media certainly play a key role in the messages that children and adults continue to see and hear (Bakir & Palan, 2010), from textbooks to Facebook, billboards to blogs, catalogs to Amazon.com. The speed with which information is both generated and transmitted today makes it very challenging to know what to trust as Truth and what to ignore as a media construct.

Through attention to content and the power of social networking, though, individuals have now been given a voice regarding sex typing in media, as evidenced by the 2014 letter (Benjamin) from a seven year old girl challenging the absence of strong female LEGO characters in their toys, which resulted in a sellout within the first few days of the availability of the women in science set in mid-2014 (although notably absent in future production plans). This author contacted a well-known online re-seller about gendered ads in holiday gifts in December 2013 and by December 2014, the ads were replaced with gender neutral ones for the same toys.

Thanks to the Internet, what was once the bailiwick of marketing firms and publicists is now open to crowdsourcing and kick-start campaigns which can be accomplished from the comfort of home, like GoldiBlox (Marikar, 2014). Personal social efficacy, or drive to deliver social change locally (Fernández-Ballesteros, Díez-

Nicolás, Caprara, Barbaranelli, & Bandura, 2002), is the first step towards replacing society-level outdated and biased representations with socially equitable ones.

The issue that triggered the NAE report was not necessarily the absence of women from the E&T landscape as much as an economic need to prepare for the demands of a 21st century workforce with an adequate supply of engineers and technologists. Media can help or hurt the effort to maximize the numbers of both genders to pursue engineering and technology careers through their approach (Bystydzienski & Brown, 2012; Bakir & Palan, 2010; Bandura & Bussey, 1999), though society cannot afford to trust the messaging approach as the only solution to engaging more females and underrepresented minorities (URMs) in a world where stereotypes still abound.

Stereotypes. Claude Steele talks about "identity contingencies" (2010, p. 3) in his book *Whistling Vivaldi: How Stereotypes Affect Us and What We Can Do,* characterizing them as "the things you have to deal with in a situation because you have a given social identity" (p. 3). He goes on to say:

> The sense of having a given social identity arises from having to deal with important identity contingencies like negative stereotypes about your group, group segregation of one sort or another, discrimination and prejudice, and so on, all because you have a given characteristic. (p. 73)

Stereotypes serve as a form of short cut for assessing new situations. White and White aver that "even when objectively wrong, stereotypes simplify social perception and serve as guidelines for social interaction" (2006, p. 259). In many respects, learning itself is a cycle of developing a series of mini-stereotypes about how things work very early in life, then discarding invalid ones for new ones during maturation.

Stereotypes can apply to just about anything: *When dogs' tails are wagging, they are happy; all engineers are good at math*. Like identity contingencies, not all stereotypes are negative or hierarchy-enhancing. They are often related to residual experiences during development: *Mothers should stay home during their children's early years*. Some help keep us safe: *More trouble happens at night on deserted streets than during daylight*. Even if explicit stereotypes are rejected or cognitively abandoned, implicit stereotypes may still remain in the subconscious as "the continuing influence of past experience and learned associations" (White & White, 2006, p. 260).

More likely than not, negative stereotypes will be applied to anyone or thing which does not fit the perceiver's view of the world, whether from an individual viewpoint or group-think, that is, the social ideologies of the higher status in-group. As Pratto, et al. point out: "an organization's members help an institution perform its hierarchy role by endorsing legitimizing myths and adapting their SDO [Social Dominance Orientation] levels to the institution's norms" (1994, p. 758). If members can neither endorse legitimized myths nor adapt, their membership will be short-lived.

In male-dominant occupations, organizational SDO levels are directly related to the number of males (Pratto, Stallworth, Sidanius, & Siers, 1997). Females must thus walk a narrow balance beam between being perceived as too masculine, that is, self-promoting (Moss-Racusin & Rudman, 2010) and too radically feminist (Saunders & Kashubeck-West, 2006), or they are penalized (Parks-Stamm, Heilman, & Hearns, 2008). Pratto, et al. (1997) highlighted Eagly and Steffen's (1984) theory that gender stereotypes evolved from observing people in gendered social roles and echoed Eccles (1987) in their explanation of "how beliefs about gender could cause men and women to make

different career choices (through gender role socialization of personal and professional values) and how such beliefs will lead others to pressure males and females into different social roles [through stereotypes]" (Pratto, Stallworth, Sidanius, & Siers, 1997, p. 50).

Stereotype threat. Several researchers (Campbell & Collaer, 2009; Cheryan, Siy, Vichayapai, Drury, & Kim, 2011; Cheryan, Davies, Plaut, & Steele, 2009; Cheryan, Plaut, Handron, & Hudson, 2013; Nosek & Smyth, 2011) have discovered that stereotypes do not need to be explicit for them to be effective at maintaining the status quo – the mere possibility that a stereotype might be applied is enough to threaten performance in domains where the group is already marginalized, a phenomenon known as *stereotype threat*. Shapiro and Williams define it as "a concern or anxiety that one's performance or actions can be seen through the lens of a negative stereotype" (2012, p. 175).

Spencer, Steele, and Quinn examined stereotype threat for women in math domains, which they attribute as similar to physical sciences and engineering, and they caution: "at some point, continuously facing stereotype threat in these domains, women may disidentify with them and seek other domains on which to base their identity and esteem" (1999, p. 25). Many women have options (Wang, Eccles, & Kenny, 2013), and with typically lower social dominance orientation, are less competitive about salary and promotability, seeking instead flexibility, stability, and a sense of belonging (Ceci, Williams, Sumner, & DeFraine, 2012; Eccles, 2011; Fetterolf & Eagly, 2011). As in racism, Steele (2010) asserts that a certain amount of stereotype threat may actually result in an improvement of performance in an attempt to disprove the stereotype, however, the threat of stereotype never really disappears in environments where it might be salient,

such as women in male-dominant occupations, or males in nursing (Spencer, Steele, & Quinn, 1999), so the need for compensatory effort is chronic and may be psychologically and physiologically stressful (Evans & Steptoe, 2002).

Backlash avoidance. Since 1991, when Susan Faludi first published *Backlash*, some things have changed related to gender equality, but not all for the better. With respect to legal rape in marriage, every state now bans it, instead of only twenty, which is progress, but women are still underpaid for the same work as men (Corbett & Hill, 2012) and the "pink ghetto" (Stallard, Ehrenreich, & Sklar, 1983, p. 33) has now expanded to include psychologists, with females numbering over 50% in practice and nearly 70% of psychology doctorates (American Psychological Association, 2014). One in three law degrees now go to females, although less than one in five Fortune 1000 attorneys are female (American Bar Association Commission on Women in the Profession, 2013).

As in teaching, once females exceed 60% of the workforce in a particular domain, society may deem the role as female-dominant, which has historically justified lower pay and less prestige (Committee on Maximizing the Potential of Women in Academic Science and Engineering, 2007). It has been suggested that female occupations may receive lower pay because they are service-based (Corbett & Hill, 2012), where women fulfill nurturing roles, such as teacher, nurse, etc., characterized as stereotypically feminine roles. However, when the service-based role is female-dominant, say for physical therapists at near 70% or nurses at 95%, the evidence is that males generally still get paid more than females (Bullen, 2012).

To avoid being socially and economically penalized (backlash), it is not uncommon for women to either avoid or exit situations that expose them to negative bias

(Fletcher, 1999; Committee on Maximizing the Potential of Women in Academic Science and Engineering, 2007; Faludi, 1991; Moss-Racusin & Rudman, 2010). Alternatively or additionally, they may embrace protective paternalism and equivocal egalitarianism as practices of sexism (Becker & Swim, 2012; Glick & Fiske, 2011; Jackman, 1994; Kilianski & Rudman, 1998) or adopt a Queen Bee approach (Derks, Ellemers, Laar, & Groot, 2011; Staines, Tavris, & Jayaratne, 1974), either of which seems to provide sufficient individual benefit to them to warrant the denial of feminism and forfeiture of gender equality for the collective group.

Equivocal egalitarianism. There are many women who do not want to give up their favored positions in the lifeboat or throw away their little black dress under the dictum of feminism or equality. They want the benefits of equality, but do not want to give up the benefits of inequality. Having grown up in a role-traditional environment paved with protective paternalism, and having quashed any early questioning of the gender status quo in an effort to fit in, they can easily rationalize and justify inequities as not applicable to them (Jost & Kay, 2005), and may resent the threat to their own acceptable situations.

Kilianski and Rudman label these women “equivocal egalitarians” (1998, p. 335), that is, they support women’s rights when it suits their purpose, but otherwise, are very much camp followers for the male hegemony (Connell & Messerschmidt, 2005). They discount and rationalize benevolent sexism, which includes paternalism and innuendo, because it is the norm; they are acculturated. They may feel especially threatened by feminism in a male-dominant environment where they have been incrementally successful, because it may impact their good thing.

Queen Bee syndrome. Staines, et al. first described this antifeminist behavior in the mid-70s in *Psychology Today* as applying to working women whose "countermilitancy has its roots in their personal success within the system: both professional success (a high-status job with good pay) and social success (popularity with men, attractiveness, a good marriage)" (1974, p. 55). Coopting the first few of a minority when doors to an exclusive group are opened to them is one counterstrategy used by privileged groups when insubordination threatens, as in times of increasing militancy by minority groups to overthrow oppression.

An international study further examined this phenomenon to discern whether it was a consequence of organizational gender discrimination rather than an individual or personal characteristic that could be changed. Their analysis concluded that it was due to ingrained organizational and societal gender bias that certain women felt a need to associate more with the masculine ingroup and disassociate from the feminine outgroup as antithetical to their purposes. Staines, et al.'s Queen Bees also had a little of Kilianski and Rudman's equivocal egalitarian in them: they were not beyond using their femininity and wiles to get what they wanted, but only for themselves, "'looking so feminine' yet 'thinking like a man'" (1974, p. 57). The Queen Bee represents a form of tokenism in that she is accepted in the exclusive club, with the price of entry being subordination, and perhaps an implicit promise to assist in quelling any incipient resistance from the rest of the out-group. She is very averse to association with women's rights advocacy and will claim that hard work is what earned her professional and personal success. The comparable icon in racism is "Uncle Tom", who cooperates with slave owners in maintaining the subjugation of blacks (Steele, 2010).

Sexism. There are many theories as to why sexism continues to prevail. Social dominance theory (Sidanius & Pratto, 1999) links sexism and all group oppression, particularly protective paternalism, to power distribution (Vescio, Gervais, Snyder, & Hoover, 2005), which is inevitably linked to economics (Corbett & Hill, 2012). Kilianski and Rudman's results suggested that women in particular are significant contributors to their own oppression from three different sources:

> 1) an implicit personality theory (Bruner & Taguiri, 1954) that construed hostile and benevolent sexism as unrelated or inversely related ' traits, (2) traditional life goals that were consistent with or were directly dependent upon benevolent sexism in men; and (3) a more positivist personal epistemology (Unger, Draper & Pendergrass, 1986); that is a set of convictions about reality that tends to support a belief in traditional gender roles as inevitable outcomes of a natural order, as opposed to socially constructed scripts and schemas. (1998, p. 336)

Privilege. Like racism, sexism is a negative bias of the privileged class towards the non-privileged, or oppressed, class (McIntosh, 1988), and the strength and proliferation of stereotypes is evidence of how entrenched bias is in a culture (Steele, 2010). As many researchers have demonstrated, bias is pretty egalitarian in that prejudice towards one out-group likely signifies negative biases towards any out-group (Aosved, Long, & Voller, 2009; Committee on Maximizing the Potential of Women in Academic Science and Engineering, 2007; Halim, Ruble, & Amodio, 2011; Jost & Kay, 2005).

Several researchers highlight that although explicit and blatant discriminatory practices such as sexual harassment may have diminished, more covert (Swim & Cohen, 1997; Becker & Swim, 2012) and ambivalent (Glick & Fiske, 2011) forms of sexism

have surfaced, in many cases undiscoverable except through tests of implicit bias (Goldin & Rouse, 2000; Halim, Ruble, & Amodio, 2011; Jost, et al., 2009; Lane, Goh, & Driver-Linn, 2011; Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012; Nosek & Smyth, 2011; Park, Smith, & Correll, 2010).

Implicit bias. In their seminal work on implicit social cognition, Greenwald and Banaji (1995) identified stereotypes as a form of implicit bias, along with halo effects, or attributing many positive traits to someone with some positive characteristic, such as attractiveness. Specifically, they say "the theorized ordinariness of implicit stereotyping is consistent with recent findings of discrimination by people who explicitly disavow prejudice" (1995, p. 4). Implicit bias presupposes autonomous response and the Implicit Assessment Test was designed to measure that response without the inherent measurement bias of self-reporting (Greenwald, Nosek, & Banaji, 2003).

Greenwald and Banaji (1995) identified three strategies to avoid unintended discrimination through implicit bias: 1) blinding, 2) consciousness raising, and 3) affirmative action. Blinding

> denies decision makers access to potentially biased information. . . .[while, alternatively, consciousness raising] encourages the decision maker to have heightened awareness of potential cues that could elicit discrimination. . . . [and affirmative action is where] an attribute that is known to be responsible for adverse discrimination is treated instead as if it were a positive qualification for the decision in question. (p. 19)

Golden and Rouse (2000) examined one of the strategies (blinding) used in interviews for predominantly male American orchestra openings and confirmed that

using a screen to block candidates from view, combined with using initials on paperwork, resulted in an increase in the number of females hired. Appearance, voice, and other salient details often cue implicit bias automatically and unconsciously. Denise Sekaquaptewa used her work on racism constructs to point out that the

> stereotype inoculation model and work on [implicit] stereotypic attribution bias suggests that [females'] choices can be [negatively] guided by stereotypes acting as an invisible hand that not only nudges them away from STEM but also may sweep their STEM successes under the rug. (2011, p. 294)

Moss-Racusin, et al.'s recent study on science faculty hiring bias (2012) confirmed that cues as innocuous as providing full name and gender on resumes was enough to prime implicit bias towards fictitious candidates whose curriculum vitae were identical except for name and sex. Both male and female faculty preferred the male candidate, and even recommended paying him $5,000 more salary than the female. They did find the female more likeable than the male candidate, but significantly less competent, which translated into poorer hireability and lower salary. The authors suggest:

> these findings underscore the point that faculty participants did not exhibit outright hostility or dislike toward female students, but were instead affected by pervasive gender stereotypes, unintentionally downgrading the competence, hireability, salary, and mentoring of a female student compared with an identical male. (p. 16477)

As in many studies of psychosocial responses, contrived experimental laboratory settings do not necessarily translate directly into the workplace, however, implicit social cognition, including implicit bias, is at play in many workplace hiring and promotional

practices being applied. Organizational values that emphasize self-promotion, risk-taking, and hierarchy, while sacrificing relational practices including collaboration and teamwork, may result in very masculine organizations due not only to barriers of entry, but also exodus due to hostile environments.

Devaluation and discrimination. Fletcher contends in *Disappearing Acts* (1999) that, since females have historically been better than males at relational practices such as nurturing, collaborating, teamwork, and other socialization skills, they are expected to unilaterally provide them as part of their gender role in any occupation. They are not recognized, and may even be devalued for these skills which may be perceived as stereotypically female behaviors, regardless that the skills are increasingly required in most positions in the incoming workforce; and that "conventional definitions of work are gendered because they reflect a splitting of the public and private domains of life along gender lines" (loc. 292).

Fletcher's study examined the social construction of gender in engineering (a male-dominant occupation) as gender role was salient in this environment in terms of socialization theories. Her qualitative study revealed that

> being expected and even relied upon to enact relational practice, while being dismissed or devalued for acting that way. . .embodies a. . .contradiction inherent in the public/private split that the power-knowledge system of patriarchy works to suppress: *relational activity is not needed and* [yet] *women must provide it* (loc. 1470).

Although hiring practices generally forbid the enactment of discriminatory practices, it is difficult to legislate or control what happens subconsciously, particularly in domains or

organizations where transparency is not a norm (which is most private corporations and many public).

Covert sexism. Sexism is less overt than in the past and has evolved into what Glick and Fiske term *ambivalent sexism*, where ambivalence arises because of the juxtaposition of overt, aggressively hostile sexism alongside more benevolent forms that glorify the feminine aspects of females while rejecting cross-gender role expression (Glick & Fiske, 2001). Contemporary sexism, also identified as *modern sexism* (Becker & Swim, 2012) or *neosexism* (Tougas, Brown, Beaton, & Joly, 1995), recognizes ambivalence and adds measures of sociopolitical pervasiveness and denial of oppression. All three have scales which measure their relevant elements and draw from modern racism research.

Ambivalent sexism. Glick and Fiske (1996) identified three generative sources of sexism: paternalism, gender differentiation, and heterosexual intimacy. The term ambivalent sexism connotes having mixed feelings about women and gender roles and is comprised of two elements which are not additive, but related: hostile and benevolent sexism. Dominative paternalism correlates with hostile sexism, while protective paternalism aligns with benevolent sexism. Competitive gender differentiation manifests as hostile sexism, while complementary gender differentiation is an element of benevolent sexism that points out dyadic ways that men and women complement each other: men are hard, women are soft, etcetera.

Finally, heterosexual hostility may occur when women challenge males for leadership, often resulting in hostile sexism, whereas, if they submit to subordinate roles that follow gender boundaries (heterosexual intimacy), they are less threatening to the

status quo and rewarded in subtle or visible ways through benevolent sexism, such as when female engineers are offered roles in staff (read: support) positions rather than ones that involve operations or highly technical expertise.

According to Becker and Swim (2012), benevolent sexism can hurt women in many ways:

> First, it legitimizes and maintains inequality by attributing traditional feminine characteristics to women and by offering the promise of protection that is, however, enacted only when women behave in accordance with their prescriptive gender role (Glick, Diebold, Bailey-Werner, & Zhu, 1997). Second, benevolent sexism can be effective in maintaining gender inequality by undermining women's resistance against discrimination (Jackman, 1994). (p. 127)

As in implicit bias, most ambivalent sexists would likely explicitly say they care for and/or respect women. Many of them are women and have no idea they are sexist. As in any behavior modification program, the first step lies in facing the unfavorable behavior (sexism) and understanding the depth of its hold at micro and macro levels so that interventions can be developed.

Modern sexism. Swim, et al. (1997) subdivided sexism into overt, deemed old-fashioned, and covert, labeled modern, sexism, also distinct but correlated. She and her colleagues concluded that those

> high in modern sexism are likely to have different perceptions of women's experiences in the workforce and are more likely to perceive greater equality in the workforce than actually exists; . . . are more likely to attribute sex segregation to individualistic causes rather than to discrimination or prejudice against

> women;. . . . [and] are likely to lead to less support for social and political changes designed to increase women's opportunities. (Swim, Aikin, Hall, & Hunter, 1995, p. 209)

In Becker and Swim's later studies, they further examined benevolent or modern sexism, discerning that prejudice or discrimination must not only be perceived to be harmful, but also pervasive before social action is likely to occur (2012).

Neosexism. Tougas, et al. define neosexism to be a "manifestation of a conflict between egalitarian values and residual negative feelings toward women" (1995, p. 843), similar to modern racism. In their study of Canadian male students (n = 130) and managers (n = 149), they too found distinctions between traditional sexism and neosexism, marked by "clear evidence that both neosexism and considerations of collective interest negatively affect the evaluation of the competence of those targeted by affirmative action" (1995, p. 847). One limitation to their study was that it was all male, leading to their recommendation that future studies were needed

> to focus in future research on neosexism among women. Achievement of equality for women in our society is partly contingent on the willingness of women to define themselves in relatively new terms, recognize their disadvantaged situation, and support policies designed to promote social change. (p. 848)

They intimated the difficulty in this and the past reluctance of women to acknowledge sex discrimination and support affirmative action to end it, that is, their complicity in allowing inequality to persevere.

Tougas, et al. also indicated a linkage between racism and sexism, providing evidence that bias is often not one dimensional, that is, if individuals are biased towards

one sub-group, they likely have prejudices and intolerances towards others, as confirmed by Aosved, Long, and Voller (2009).

Limited exposure to E & T activities. Recruitment and retention are critical functions in broadening female participation in E & T roles, and that is where engaging young females in associated activities under the guidance of more experienced females in E & T can help or harm. Based on the research, activities that are engaging to middle school females coupled with the exposure to positive role models should have a positive impact on girls' interest in the fields of engineering and technology (Phelps, 2012; Weber, 2011; 2012), and, ultimately, *help* or increase the numbers of females entering or persisting in engineering and technology fields.

Socialization and belongingness. There is prodigious evidence that biology or genetics is not the problem (Ceci, Williams, & Barnett, 2009; Fine, 2010) – engineering design or technological creation aptitudes are seen in both sexes (Bem, 1975) and most technological expertise involving spatial skills merely requires practice or training (Feng, Spence, & Pratt, 2007; Sorby, 2009).

Why high-performing females choose to be doctors, lawyers, or engineers is not necessarily based on academic achievement, rather on more subtle cues, including socialization factors (Buday, Stake, & Peterson, 2012; Seymour, 1997), sense of belonging (Ceci, Williams, Sumner, & DeFraine, 2012), and having choices or career options (Eccles, 2011; Wang, Eccles, & Kenny, 2013). Enhanced sense of belonging does provide some positive mediation for interest level in occupational domains that involve math. Thoman, et al. found that "women feel pushed out of STEM when they feel a low

sense of belonging, [and] for women with specific self-esteem contingencies, competing experiences of belonging in non-STEM can pull interest away from STEM" (2013, p. 1).

This research, and interventions focused on improving sense of "belongingness" and social support in engineering and technology career fields, support the significance of the belongingness found in communities of practice on academic and work achievement (Draper, 2010) and female engagement in engineering (Chesler & Chesler, 2002).

Biological or genetic limitations. At the center of any discussion of biological or genetic sex differences are evolutionary psychology and reproduction. Evolutionary psychology gained momentum with Charles Darwin's advancement of the concept of *sexual selection,* which is how he categorized all behaviors and responses that were not able to be explained by the concept of natural selection or survival of the fittest (Eckes & Trautner, 2000). Mating dances, colorful mail plumage, or extensive antlers are examples of characteristics which likely evolved because they gave certain males advantages in reproduction, but otherwise appear to have no functional purpose. Darwin's research focus then shifted to emotions of all animals, including humans (1872), which efforts informed the work of psychologists including Freud, Skinner, and many others (Eckes & Trautner, 2000).

Spatial skills. Long before there were cars or video games, societal structures were based on occupational roles within those social structures, specifically females as agrarians, initially gatherers, and males as hunters (Silverman, Choi, & Peters, 1994). Certain skills, in particularly visual-spatial, either developed to support or served to predetermine specific roles (Joseph, 2000). The functions served by these roles may be

vestigial in today's society, however, evolutionary theories serve to explain lingering sex differences in terms of spatial abilities.

Females tended to be stationary and closer to home, learning to orient themselves using landmarks, while men ranged in search of animal food and mates (Ecuyer-Dab & Robert, 2004) and became more facile at dynamic orientation, today important in sports and most construction activities. Women generally are still poorer at mental rotation tasks and other visual spatial skills which involve dynamic situations, but performance improves with training (Feng, Spence, & Pratt, 2007; Moe, 2009; Moe & Pazzaglia, 2009; Neubauer, Bergner, & Schatz, 2010; Sorby, 2009), which can include sports which involve hand-eye coordination, video games, 3-D rendering, and other ways to practice the skills.

Reproductive functions. The human reproductive function, along with the different roles prescribed within the reproductive process, presents the most significant biological difference between males and females. From pre-natal hormones which may have some influence on spatial skills (Bull, Davidson, & Nordmann, 2010; Kempel, et al., 2005) to males competing for mating rights (Ecuyer-Dab & Robert, 2004), reproduction drives sex-differentiated behaviors. By random chance or evolutionary design, women drew the role of carrying offspring, thereby defining certain physical and psychosocial aspects of females that would facilitate bearing and nurturing offspring (Casey, 1996; Ecuyer-Dab & Robert, 2004). As in many species, however, bearing is different than raising and, while there is no current biological reason that males *should* be less nurturant towards offspring than females, both sexes are implicitly biased against males in nurturing roles (Park, Smith, & Correll, 2010).

Career choices and options. At the undergraduate and beyond level, opting in to an engineering and technology field from a non-E&T field may be more challenging, though not insurmountable, because of prerequisite coursework that is most often completed by the end of high school. Opting out of engineering and technology is more the norm for females (Good, Rattan, & Dweck, 2012; Williams & Ceci, 2012), as is transferring into engineering-related roles that capitalize on feminine communal or expressive skills, such as in technical sales, training, or management (Castro, Scandura, & Williams, 2004; Fletcher, 1999; Jagacinski, 1987; Williams & Ceci, 2012; Wong, Kettlewell, & Sproule, 1985). From secondary school graduation forward, mentoring is focused more on retaining women in the fields of engineering and technology rather than engaging them in the first place (Brand & Kasarda, 2014; Drury, Siy, & Cheryan, 2011; Geber & Roughneen, 2011; Poor & Brown, 2011).

Effort versus reward. Several recent studies examined achievement expectancy related to females in STEM careers and discovered that when women have career choices based on aptitudes, they often choose paths that minimize effort-to-return (Smith, Lewis, Hawthorne, & Hodges, 2013). Antecol and Cobb-Clark offer a model where people choose different careers either because they have "characteristics that lead them to be more productive or face less discrimination in some occupations than others (and hence enjoy higher returns) or because they have idiosyncratic preferences for certain occupations over others" (2013, p. 65)**.** They contend that psychosocial traits affect future productivity (through wages) as well as preferences for certain occupation-related working conditions.

Work preferences. In their study of young adults, Antecol and Cobb-Clark assert that it "seems more likely that gender segregation in discipline areas stems in part from the different preferences that men and women have for the job attributes associated with different occupations" (2013, p. 66). They did not discern a significantly higher level of masculine-typed females in traditionally male roles in their sample, although they did state that both "men and women who report that they are willing to work hard are employed in occupations which are more male-dominated, while the tendency to avoid problems is related to working in occupations with relatively more women" (p. 68).

Occupational congruence. Their results are "consistent with the psychological evidence that who is hired depends in part on a degree of congruence between the gender of the applicant and the sex type of the job" (Antecol & Cobb-Clark, 2013, p. 68). Specifically,

> entry into male-dominated fields of study and male-dominated occupations are both related to the extent to which individuals have "masculine" traits (i.e., are independent, assertive, not shy, not sensitive, and not emotional) and believe they are intelligent, while only entry into male-dominated occupations is related to the willingness to work hard, impulsivity, and the tendency to avoid problems. Moreover, the effect of psychosocial traits on field of study and occupational attainment (based on percent male) tends to differ by gender (e.g., women who believe they are intelligent are more likely to study/work in male-dominated fields of study). (p. 70)

E & T careers are rejected not because the return is not great enough, but because the effort is perceived as too costly by some (Eccles, 2011; Frome, Alfeld, Eccles, & Barber, 2006; Larose, et al., 2008; Smith, Lewis, Hawthorne, & Hodges, 2013).

Circumscription and compromise. According to Gottfredson (1981), acceptable gender roles are decided before middle school, then further narrowed based on social acceptability during middle school. During high school and into early post-secondary years, a refining of occupational aspirations takes place that incorporates an evolving self-identity based on abilities and affective traits such as self-confidence, self-esteem, and other psychosocially affected characteristics.

Like Piaget's stages of child development, Gottfredson sees the stages of occupational aspirations as one-direction only – decisions made early on are not expected to be overturned in later stages (1981, p. 554) – so it becomes important to introduce non-circumscriptive alternatives as early as possible. Students' occupational self-concepts tend to start stabilizing in high school through "circumscription and compromise" (p. 545), leading to feasible post-secondary academic and vocational options that suit their social class, by which time it may be too late to change direction.

Engineering identity development. Capobianco, French, & Diefes-Dux (2012) applied Capobianco's earlier qualitative research examining female undergraduates' engineering identities (2006) to develop an instrument to measure engineering identification in elementary-aged boys and girls. They discovered that engineering identity in younger students most strongly comprises two factors: academic abilities and engineering career alignment. As in her earlier research, this study emphasized that "identity is not fixed, absolute, or pre-given but rather a product of students' own lived

learning experiences with engineering-related tasks" (Capobianco, French, & Diefes-Dux, 2012, p. 709). In her earlier study, Capobianco suggests that improvement in female engagement in E & T fields might best be accomplished through "authentic, team-based, design-oriented tasks" (2006, p. 115), similar to those being examined here.

Summary

This study serves as a validation of data from a survey measuring the latent variable of middle school females' engagement in engineering and technology following a one day immersive engineering and technology event. It also provides a jumping-off point for additional discourse and research on the effects of mentored E & T activities on female engagement in male-dominated career fields such as engineering and technology. Survey data that demonstrate validity lend the measurement instrument credibility for future studies involving replications, designated E & T activities, role model characteristics, experimental designs involving targeted interventions, and others.

In a field whose workforce ranges between 85-90% male, one would expect that gender attitudes in the general workforce would reflect that makeup, characterized by what R.W. Connell refers to as "hegemonic masculinity" (2005, p. 11). Attitudes about men and women in the workplace would represent the attitudes of predominantly men, so as more women joined the workforce, it makes sense to think that attitudes would eventually change. Gendered attitudes and behaviors are those conscious and unconscious attributes that proscribe social or cultural constraints on the activities, roles, and behaviors of one or more individuals because of their sex.

In this sense, gender is a psychosocial concept rather than a biological categorization. Gendered attitudes are typically dyadic in nature, consistent with

stereotypes. Highly gendered individuals view masculinity and femininity as a dichotomous measure of biological maleness and femaleness; they often do not understand that gender is a psychosocial construct that serves societal purposes, and as such, can be (and is often) manipulated, e.g. gender-differentiated product marketing strategies or gender discrimination, the latter also known as sexism.

Engineering has not historically fit most societal norms for acceptable female occupations, consequently, attempts over the last 100 years to change that have been only minimally successful. Social structures build up over years, developing a thick layer of resistance to change that is mired in socioeconomic politics and requires significant inertia to overcome. Over the last 40 or so years, researchers have chipped away at gender role issues, effectively reversing once-entrenched beliefs regarding inherent biological cognitive deficits that rendered females incapable of performing as well as males in math and science, and leaving only spatial skills (in particular, mental rotation) as a cognitive difference, but one that can be further reduced through training and practice. Biological differences related to the ancestral origins of reproduction are elemental to the evolutionary psychology theory of gender development, however, as societies learn to effectively manage reproduction through technological innovations, arguments related to its ancestral motivations become less relevant.

Engaging females in fields and occupations that may not be congruent with their social and gender identity has been a Sisyphean task over the last 50 years, but has seen some incremental growth through interventions such as the mentored engineering and technology activities sponsored by the Society of Women Engineers and similar organizations such as ASEE’s K-12 Division, FIRST, and others. Any credible claim of

the efficacy of those mentored activities however, can only start with a credible and valid instrument with which to measure participants' feedback and needs additional research using that instrument.

CHAPTER III

Methodology

This study is a factor analysis of quasi-experimental secondary data collected from middle school females over a five year period after specific SWE-mentored engineering and technology events known as WOW. The purpose of the analysis is to validate the data collected via the instrument for continued reliable program evaluation and replication.

Instrument Validation

Instrument validation is not as much about the instrument as it is about how the data collected through an instrument can be statistically interpreted, that is, the types of inferences, conclusions, generalizations, and predictions that can be made to some level of confidence, typically 90-95% in most psychosocial studies, depending on the stakes. Data collected from an instrument that reliably demonstrates validity would likely show little variance in responses across groups that are supposed to be the same, and appropriate variance between groups that are supposed to be different. The data fits the construct (validates) that the instrument is supposed to be measuring.

Unified construct-based model of validity. Messick‘s early studies on validity (1989) ultimately led to his unified construct-based model of validity which emphasizes six key concepts when considering data validity: “content, substantive, structural, generalizability, external, and consequential aspects” (Dimitrov, 2012, p. 41). Messick provides guidance regarding appropriate validity evidence:

1. The content aspect of construct validity includes evidence of content relevance, representativeness, and technical quality (Lennon, 1956; Messick, 1989b);

2. The substantive aspect refers to theoretical rationales for the observed consistencies in test responses, including process models of task performance (Embretson, 1983), along with empirical evidence that the theoretical processes are actually engaged by respondents in the assessment tasks;
3. The structural aspect appraises the fidelity of the scoring structure to the structure of the construct domain at issue (Loevinger, 1957; Messick 1989b);
4. The generalizability aspect examines the extent to which score properties and interpretations generalize to and across population groups, settings, and tasks (Cook & Campbell, 1979; Shulman, 1970), including validity generalization of test criterion relationships (Hunter, Schmidt, & Jackson, 1982);
5. The external aspect includes convergent and discriminant evidence from multitrait-multimethod comparisons (Campbell & Fiske, 1959), as well as evidence of criterion relevance and applied utility (Cronbach & Gleser, 1965);
6. The consequential aspect appraises the value implications of score interpretation as a basis for action as well as the actual and potential consequences of test use, especially in regard to sources of invalidity related to issues of bias, fairness, and distributive justice (Messick, 1980, 1989b). (1995, p. 745)

Instrument development. Most of the literature on validity of data, including that collected from surveys, is from the perspective of creating a new survey or instrument. This typically involves deciding what is going to be measured, creating observable indicators that report a reliable and valid measure of the desired variable(s) and how they map to instrument objectives in a model. Once the model is confirmed by domain experts, the next step is collecting pilot data, then analyzing and statistically

interpreting the collected data to modify the instrument to remove sources of invalidity. This may include changing test administration method (paper to online or vice versa), adding new or better or reworded items, removing negative responses, and other changes expected to improve validity and/or reliability (Decoster, 2005; Devellis, 2012).

Evaluating extant instruments. One of the first steps in developing a new instrument is to decide what is being measured; with an existing survey, the data collected reflect what is being measured. They represent a measurement or empirical model of a certain set of constraints on the data based on a construct not yet validated. Through interviews with several authors of the early and current instruments, the reference survey (Appendix A) was deconstructed and provided the basis for a proposed structural model in which latent variables are being measured. They are latent because they represent attitudinal dispositions that cannot be directly measured.

WOW survey. Following an in depth theoretical and practical literature review on the topics of female engagement in E & T and instrument validation, and coupled with knowledge that the instrument has been operationalized for more than five years, certain assumptions were made regarding its content validity as specific to the domain of female engagement in E & T, especially given that it underwent a recent re-look in preparation for an annual event (October 2014).

Notwithstanding those assumptions, a key piece of validating the survey was to confirm that the items that it comprises measure what they are supposed to measure, reliably, so having a model to start from was important, and having one that the experts are in consensus on contributes to the face validity of the data collected.

Latent factors. In this study, the four event objectives provided the key factors to be measured and were somewhat mirrored in the four research questions. The items in each scale of observed responses in the survey were originally derived from the event objectives and together represent the strength of a latent factor that is not directly measureable. Based on the literature and the specified survey items, it was expected that items would resolve into several latent factors which would contribute to one higher-order factor having to do with predisposition or propensity for engagement in engineering and technology, a priming factor related to career entry.

Scales and items for complex concepts. Specifically, the topic of female engagement is complex, and according to Devellis, for complex concepts "multiple items may capture the essence of such a variable with a degree of precision that a single item could not attain" (2012, p. 12). Measurement instruments which are "collections of items combined into a composite score and intended to reveal levels of theoretical variables not readily observable by direct means are often referred to as scales" (p. 10).

Scales. The survey which produced the data being validated includes multiple scales or groupings of items. Multiple scales often provide a source of content validity for complex concepts, particularly in measuring psychosocial or affective constructs or variables such as self-efficacy or interest (Bandura, 2006). For this study, each scale presumably measures a different latent construct with all scales together representing the latent factor of the participant's likelihood of getting more engaged in engineering and technology in the future.

Scoring. Most of the items in the survey were measured using a Likert-type response based on five levels of response from *strongly disagree* to *strongly agree*. The

midpoint is *neither agree nor disagree*. One scale used 4-point ratings *improved, stayed the same, got worse,* or *I don't know,* while another rated importance using a 5-point scale, so all scoring was converted to a 5-point format for statistical analysis (Colman, Norris, & Preston, 1997).

Many argue that, irrelevant of the number of selection choices, data from Likert-type responses are not truly continuous in that they have a specified range of equal intervals. They deem it better evaluated as categorical or ordinal variables using non-parametric methods, however there is no easy or practical way to ferret out factors or scales from survey data without treating all of the Likert-type responses as continuous data using parametric methods.

This survey is both a Likert-type scale and uses the Likert response format in that it measures attitudes using data collected in a 5-point interval format with 1 being most negative and 5 being most positive. The relevant examination of the data collected via Likert response format on this deals first with normality and then with validity.

Normality. Theoretically, in order to perform a factor analysis using covariance structure analysis, multivariate data must be assumed somewhat normal, which is not often the case for Likert response data. Conventional statistical analysis experts suggest that the absolute value of skewness and kurtosis be less than one and as close to zero as possible. Fabrigar, Wegener, MacCallum, and Strahan (1999) suggest that, unless normality is severely violated (skew > 2; kurtosis >7) (p. 283), either maximum-likelihood (ML) or principal axis factor (PAF) extraction may be used for exploratory factor analysis.

Severe non-normality should limit factor analysis to the principal axis extraction method, but otherwise ML is preferred because it "allows for the computation of a wide range of indexes of the goodness of fit of the model . . . [and] permits statistical significance testing of factor loadings and correlations among factors, and the computation of confidence intervals" (p. 277). An assumption of multivariate normality in Likert response data may be relatively harmless, however, when between group behavior comparisons are not being evaluated or are limited to the item and not factor level and the sample is homogeneous.

Validity. According to Carifio and Perla (2007),

> five (5) basic and widely agreed-upon kinds of validity in the psychometric literature may be conceptually reduced to [two types of validity,] logical/semantic (content and face validity) and empirical (concurrent, predictive and construct). . . . with the empirical validities being confirmatory of the logical validities (i.e., concept/theory and observed facts/agreements). (p. 109)

Many statistical purists assign greater weight to empirical validity, arguing that Likert responses do not adequately support construct validity tests, while simultaneously discounting content or face validity. Carifio and Perla (2007) counter that Likert placed equal value on the logical/semantic or content validity of his attitude scale and the empirical (in this case construct) validity, and knew that the two must be intrinsically tied together in order to produce reliable and valid responses. They even went so far as to compare measurements taken on a continuum response with those taken using a Likert response format, with a very strong correlation ($r = .92$) between the two, which was further corroborated by others (p. 109).

When items load onto factors, the total effect of all of the items loading onto that factor could be represented as a summation or an average of all of the items' scores. Depending on whether any of the variables prove to be theoretically linked but statistically uncorrelated, Bobko and Schwartz (1984) offer a mathematical model for summating such oblique relationships to compare variables. Although not in the scope of this study, such continuous data may provide more variance in scoring than the individual scales when interpreting data. This aids with discriminant validity, which looks for differences where differences should be, perhaps between different extremes of ages within the sample.

Data Validation

This study used exploratory factor analysis (EFA) as a first level and preliminary confirmation for structure validity, followed by confirmatory factor analysis (CFA) using the pattern matrix and preliminary factor loadings from the EFA on secondary data available from SWE. Along with normality checks, evidence of validity tests were completed as per Table 2 to suggest areas of consistency and concern. Table 2 was adapted from Dimitrov's "Statistical Methods for Validation of Assessment Scale Data in Counseling and Related Fields" ((2012, pp. 48-49).

Sample size. There were N = 332 data records available for analysis following deletion of records missing data (< 10%). Depending on several different criteria, including ratios of sample size to variables or variables to indicators, overall sample size, type of test to be performed, and other conditions, sample size thresholds will fluctuate. For factor analysis specifically, Osborne and Costello offer that "larger samples are better than smaller samples (all other things being equal) because larger samples tend to

minimize the probability of errors, maximize the accuracy of population estimates, and increase the generalizability of the results" (2004, p. 2). Their review of the literature discusses three different approaches: subject to variable/factor ratio, subject to item ratio, and absolute minimum, and concludes that it depends, but larger generally is better.

Subject to variable/factor ratio. When counting variables for exploratory factor analysis, all relevant items were considered variables (p = 24). Since a construct was proposed wherein several items contribute to certain scales, confirmatory factor analysis presumed four factors. When generalization is critical, the N:p ratio could be as high as 30:1, but nominally 5 – 10:1 is acceptable.

For the EFA with 24 dependent variable scores, the smallest sample size should be 120, preferably 240. For the CFA with four expected endogenous or dependent variables: N would need to be 40 minimum, preferably 80-100. Using this approach to determine valid sample size could have significant effects on pattern comparison, Type I error rates, and obtaining the correct loading pattern, especially at lower Ns.

Type I errors detect effects that are not present, while type II errors fail to detect an effect that is present. According to Costello and Osborne, "in general, the stronger the data, the smaller the sample can be for an accurate analysis. 'Strong data' in factor analysis means uniformly high communalities without cross loadings, plus several variables loading strongly on each factor" (2005, p. 4).

Subject to item ratio. Osborne and Costello (2004) echo Gorsuch (1983) and Hatcher (1994) in specifying a subject to item ratio of at least 5:1 for EFA, although Nunnally's widely-cited rule of thumb (1978) has been 10:1. For 24 items, the required N would be between 120 and 240.

Table 2.

Proposed Analytical Process for Validating Data

Aspects of Validity	Importance	Evidence of	Test
Content	Content relevance Representativeness	Face validity Logical validity	Collect subjective relevance judgments of experts and/ or examinees.
	Technical quality	Subscale items correlation	Examine Pearson correlations between items that serve as indicators for the same construct (subscale).
Substantive	Measurement stability Measurement accuracy Accurate description of sample	Scale functioning	Examine scale functioning and ensure responses consistent with intended response characteristics.
	Inferences to target population	Cognitive modeling	Review instrument in context of existing theory and empirical research on cognitive processes related to the construct of female engagement. Ensure samples for test groups are randomly stratified.
Structural	Fidelity of scoring structure to structure of the construct domain	Correlational / measurement consistency between constructs and indicators (items)	Pearson correlations Factor analysis (both EFA and CFA)
Generalizability	Identifines boundaries of meaning of scores across tasks and contexts	Extent to which score properties and interpretations generalize to and across population groups, settings , and tasks	Create equivalent groups and test factorial invariance of constructs across groups (CFA), confirming (a) configural invariance: baseline score factor structure is same; (b) measurement invariance: score meaning and interpretation is same (c) structural invariance: construct scores variability (and correlations in case of two or more constructs) is same
		Reliability	Using Pearson's coefficient of reliability, examine degree to which data are accurate, consistent, and replicable when (a) different people conduct the measurement, (b) different instruments that purport to measure the same trait are used, and (c) there is incidental variation in measurement, such as due to an instrument modification during measurement period.
External	Expected between-group construct model differences support external validity. Responsiveness, referred to also as sensitivity, is the capacity of the assessment instrument to detect change.	Differential construct loading between groups known or expected	Examine the effects of age on the construct model, with expectation that there is a difference in factor loadings between oldest annd youngest on E & T self efficacy.

Absolute minimum. Osborne and Costello also cite Comfrey and Lee (1992) in suggesting raw *N*s as follows for factor analysis: 50 – very poor; 100 – poor; 200 – fair; 300 – good; 500 – very good; 1000 or more – excellent. Raw *N* has a significant influence on the average percent of Type II errors, with higher errors at low *N*s. Total *N* matters more when subject to item ratio is low but loses its unique impact once the ratio of *N*:p is accounted for.

Grouping. After accounting for significant mismatches in data samples from different sites and the changes in questions over time, only three of eight data samples were similar enough to the current day survey to be used; $N = 332$. Dividing this sample into homogeneous groups $n = 221$ for the exploratory factor analysis and $n = 111$ for the confirmatory was borderline for interpretability, but factorial invariance across groups provided evidence that could contribute to generalizability. Saving one sample for confirmatory factor analysis following the preliminary EFA enabled testing of the proposed model based on theory and then refined using empirical evidence. Any subsequent revisions would require another confirmatory sample.

Congruence. A simple measure of congruent validity for this survey was if factors extracted the same way across comparably distributed groups as for the larger sample population. Presumably, if the survey was designed correctly, the factors should align with the event objectives, all of which are obliquely related. Obliqueness implies some communality between factors because of the affective and psychosocial nature of engagement, versus orthogonal factors which posits binary (either-or, on-off) categorization.

In this study, the four event objectives provide the key factors to be measured and were reflected in the four research questions. The items in each scale of observed responses in the survey were presumed to be derived from the event objectives and would contribute to a composite priming score that represents the strength of a variable that is not directly measureable (latent). Based on the literature and the survey items, it was expected that items would resolve into several latent factors which might also contribute to one or more higher-order factors, together referred to as the engagement priming factor (EPF).

Survey construct and samples. Understanding the survey instrument's purpose and objectives in the context of the target audience, middle school girls, enabled data validation to begin.

***Event objectives*.** The specific objectives to be measured for signature events were that the events:

1. Facilitate a change in attitude about careers in engineering [and technology] by introducing participants to the creative, innovative, and forward thinking sides of engineering.
2. Engage participants in hands-on activities that build self-confidence and critical thinking skills around engineering [and technology] principles.
3. Foster mentoring relationships between participants and role models to encourage an expanded sense of community and a deeper understanding of what engineers [and technologists] do.
4. Connect priorities about what participants want in their future work and life to the work and life of an engineer [or technologist] by sharing personal stories

> and celebrating the accomplishments of women engineers [and technologists].(Society of Women Engineers, 2012, p. 1)

Survey objectives. The purpose of the survey is to report middle school girls' responses to a hands-on mentoring event so that their input can be used as a measure of how well the event met each of its objectives. Briefly, the survey is intended to measure 1) changes in attitudes about and interest in engineering, 2) understanding of what an engineer does from interactions with volunteer role models, 3) changes in confidence and self-efficacy due to hands-on activities, and 4) preferred future work type. Identifying the target audience as children assumed that certain survey accommodations would likely be needed in order to obtain valid and unbiased data from less mature cognitions: brevity, simple language, and low risk or reward.

Preliminary structural model. Based on cursory dataset analysis and appropriate literature review, Figure 1 (p. 12) presents a proposed structural model of the WOW event objectives defined as important to measure. Beyond informed groupings, the researcher did not pre-assign weightings to any particular factors, variables, or items. Only through confirmatory factor analysis (a form of structural equation modeling, or SEM) could a specified model be compared to what the data say is happening.

Data Collection

All data were purposively collected at eight WOW events via paper-and-pencil surveys between 2009 and 2013, entered into SWE's Survey Monkey account for further analysis following the events, and maintained in password-secured data storage files by the Society of Women Engineers to ensure participant privacy. Authorized users are

allowed to download data files locally and have a responsibility to remove information identifiable to any individual, such as email addresses.

Sample characteristics. The samples were characterized using the demographic indicators age, gender, and race / ethnicity (optional). Prior to data screening, the median age was 12, males represented a little over 3% of respondents, and the total sample reported themselves as approximately 38% white, 13% Hispanic, 18% African American, 18% Asian American or Pacific Islander, 1% Native American, and 8% mixed heritage, with 5% making no selection.

Survey sample size. The initial raw sample for secondary analysis of the survey data for the years 2009-2013 was $N = 999$, however, over the course of the five years of events, the survey was modified several times until stabilizing in 2011. Until detailed analysis began after IRB approval, this was not apparent. The current form has been in general use since 2011 for a revised sample $N = 332$, which is considered a large enough sample to perform either exploratory or confirmatory factor analysis on.

The demographics of the reduced-size sample, however, were different than those of the raw data: median age was 13 years, male representation doubled to 6%, and race/ethnicity was redistributed as follows: white 19%, Hispanic 23%, African-American 17%, Asian-American or Pacific Islander 27%, Native American 2%, mixed 12% and less than one percent opting out of identifying. Since the data were highly consistent in both samples, however, the demographic differences were deemed to be irrelevant to validity, although of note in delimitations.

Typically, EFAs are done early in the process of scale development with a smaller sample, then confirmed on a larger sample. This survey was already somewhat

institutionalized for the Society of Women Engineers events with a large data sample, so it was decided to see how well the existing survey results' factor analysis using Maximum Likelihood (ML) corresponded to the proposed model in SPSS, then correct by eliminating items from the model to provide a meaningful model (loading > .45) that reflects the highest parsimony without sacrificing internal consistency and discriminant validity of the constructs. The resultant pattern matrix and loading was then used to build a measurement model in AMOS for a confirmatory factor analysis.

In order to perform an EFA followed by a CFA, the data needed to be purposively divided into two sets based on the demographic characteristics of the larger sample. Since a rough rule of thumb for sample size needed for exploratory factor analysis is approximately 10 times the number of indicators (p) (Brown, 2006), the larger sample set was assigned for the EFA analysis (N = 221). Following the initial EFA/CFA analysis, the samples were reversed and the smaller sample analyzed for the EFA and the larger sample used for the CFA to examine whether items loaded onto factors differently. Finally, the full sample set (N = 332) was also examined for model fit.

Factor weights. MacCallum, Widaman, Zhang, & Hong suggest that defining sample size when using factor analysis is rather complex and that researchers can expect an interactive effect between unique factor weights and sample size such that,

> when unique factor weights are small (high communalities), the impact of this source of sampling error will be small regardless of sample size, and recovery of population factors should be good. However, as unique factor weights become larger (low communalities), the impact of this source of sampling error is more

strongly influenced by sample size, causing poorer recovery of population factors. (1999, p. 89)

Factor over-determination. They also posit that factor over-determination is the "degree to which each factor is clearly represented by a sufficient number of variables" (p. 89) as represented by the equation p/r, where p is the number of items or indicators and r is the number of factors. Highly overdetermined factors exist when the number of indicators is at least several times the number of factors. McCallum, et al. also posited "an interactive effect between sample size and degree of over-determination of the factors, such that when factors are highly overdetermined, sample size may have less impact on the quality of results" (p. 90).

Heterogeneity. Several of the events were held in locations that provided opportunities to oversample underrepresented ethnicities such as African-American (Tuskegee, AL and Baltimore-Washington), Asian-Pacific American (Hawaii and San Diego), and Hispanic (San Diego), so the survey has been tested on a heterogeneous population. Participants were invited from local public and private schools, using various media and networks, such as Girl Scouts of the USA, FIRST, Girls, Inc., and other girl-serving organizations. Attendance was voluntary, but the girls who attended were there non-randomly, so the data had a good chance of being non-normal.

Each location managed their own invitations and registration process, although all were required to collect a standardized parent consent form for minors, copies of which are maintained in the event locations. At the events, students were told they could choose to opt out of completing the survey at the end of the day. Most locations administered a

paper copy of the survey, then submitted them to a service for data entry, with a nearly 100% response rate.

Consent and assent. Middle school females are considered minors, therefore, their parents must consent to their participation in events that may result in potential publication such as research, videos, or photos. During the registration process, parents gave consent for their child to participate in the event, along with any data gathering that would ensue. Participants could opt out of any of the activities or final survey, but they were then ineligible for complimentary collateral such as t-shirts or goody bags unless they assented to completing the survey.

Depending on the registration process used, soft copies of parental consent forms were embedded in online registration records or kept in hard storage at the local venues, redacted for any identifiable data. Unless someone waived a release during a manual registration process, that is, no photos, videos or data collection from the child, the assumption was that most participants had parental consent by their presence at the event since a parent or guardian had to drop them off.

Data entry. Survey responses were shipped to SWE headquarter offices following events and manually entered into Survey Monkey as separate data files. There is always some inherent error in manually transposing data from one medium to another that could bias an otherwise robust analysis, so data was carefully examined for outlying anomalies or obvious data entry errors.

Data Analysis

Most of the analytical work involved the validation of the data collected with the survey instrument. A large portion of analytical effort consisted of data screening, that is,

deciding how to deal with missing data, evaluating the normality consistency, and adequacy of the data, and compiling two demographically matched samples for the separate factor analyses. Some of the scales comprised four levels of response, while most comprised five, so the former responses were rescored using a 5-point format (Colman, Norris, & Preston, 1997). Data were also examined for common method bias.

Factor analysis methodology. Once the data were deemed adequate for analysis based on normality, consistency, and sample size, a decision had to be made about which approach was the best to use for extracting factors: maximum-likelihood (ML) or principle-axis factor extraction (PAF). Each had its own benefits, but the ability to use ML was preferred because it has a wide range of goodness-of-fit indices which could be used during confirmatory factor analysis that PAF does not.

Maximum-Likelihood extraction. ML methods generate parameter estimates that are most likely to produce the observed correlation matrix, assuming the sample is from a normal distribution. The reproduced correlation matrix is compared with the actual correlation matrix through an iterative process that ideally concludes with convergence between the two matrices (Fleming M., et al., 2013, p. 1004). When comparing the reproduced correlation matrix to the actual correlation matrix, the fit is represented by the discrepancies between the two matrices, with the goal being non-significance.

Goodness-of-fit. Goodness-of-fit describes how closely the specified model fits the collected data but is dependent on several things, i.e., the number of parameters (items), factors, and sample size, so there is no one index that suits all cases. Many experts recommend ensuring that tests for goodness-of-fit include at least one fit index from each of the absolute, parsimonious, and comparative fit indices (Brown, 2006;

Brown & Cudeck, 1992). The goodness-of-fit statistics used in the confirmatory analysis were chi-square (absolute), comparative fit index (CFI, comparative), root mean square error of approximation (RMSEA) and PCLOSE (parsimony-adjusted), and Tucker-Lewis index (comparative).

Factor analysis process. Two-thirds of the sample (*N* = 221) were examined using EFA to first ferret out the "natural" factors and determine whether any of the 24 dependent items were not contributing to the overall model thereby threatening validity, and second, to test the proposed four factor model, A confirmatory factor analysis was conducted on the remaining one-third (n = 111) of the sample using the model developed during the EFA.

Factor retention. Defining the preliminary model was somewhat iterative. SPSS's factor analysis was used on the first data set to define a model, then a measurement model followed using AMOS to confirm the second set of data's goodness-of-fit to the previously extracted EFA construct. Which factors to keep in the final model was somewhat subjective, but largely based on Costello and Osborne's guidelines (2005, pp. 4-5): 1) moderate to high communalities (> .40), 2) moderately weak cross-loadings (< .40), 3) moderately strong factor loadings (> .40), and 4) each factor had at least two indicators that loaded strongly (> .5).

Goodness-of-fit correction. When the goodness-of-fit for the CFA was problematic, the various fit indices in AMOS guided remedies. If a variable or item was removed from the solution to improve fit, the EFA was rerun using the remaining variables, then fed back to AMOS for CFA. Dimitrov suggests that, in accordance with Jöreskog, EFA can be used

> in a CFA framework to obtain preliminary statistical information (e.g., on standard errors of factor loadings) that can be useful in a subsequent CFA (e.g., to select appropriate anchor items). In this case EFA is used to provide a technical jump-start for CFA, not to generate hypotheses about the factor structure of interest. (2012, p. 91)

Variables. Table 1 (p. 9) delineates the survey variables of concern for the factor analysis. Although they were included in the consistency and data adequacy estimations, the independent moderator variables of age, gender, and race/ethnicity were not used during the factor analysis.

Variable responses. All of the variables were polytomous, that is, they had more than two levels of response. Polytomous variables may be categorical or interval, continuous or discrete. In this study, two independent variables were categorical (race and gender), while the rest were considered continuous for the purposes of parametric testing such as factor analysis. During the analytical and interpretive phase of the study, it was expected that some variables would disappear or collapse into others based on the EFA results, which is why the final analysis was a confirmatory factor analysis using a clean data set.

Likert-type responses and factor analysis. According to Dimitrov, factor analysis using "Pearson correlations [as in SPSS] is appropriate for continuous variables but not [necessarily] for . . . . polytomous variables [such as Likert-type ratings], because . . . . EFA [or CFA] for polytomous data should be based on polychoric correlations" (2012, p. 90). Polychoric correlation estimates the statistical relationship between two theoretically normally distributed continuous latent variables from two observed ordinal variables,

such as ratings from a 5-point Likert-type scale of agreement, from *strongly disagree* to *strongly agree.* The fewer the number of response categories, the more any correlation between latent continuous variables will tend to be moderated, therefore, it is often recommended to use a computer program that uses polychoric correlation in factor analysis, such as MPlus or FACTOR (p. 90), especially when evaluating factor loadings for either EFA or CFA.

With 5- and 7-point response ratings, however, Carifio and Perla (2007) demonstrated that Likert ratings did not have to be treated as polytomous data and that AMOS' maximum-likelihood factor extraction (which does not use polychoric correlation) could effectively be used to determine structural correlations.

Validity tests. Beyond face (or content) validity, and goodness-of-fit which contributes to construct validity, many analyses spend little time on convergent or discriminant validity, or determining the effects of common method bias. Convergent validity demonstrates correlations where they should be, typically higher between the variables loading onto each factor and more moderately between theoretically-related factors in the same instrument. Discriminant validity measures how clearly the responses for each factor can be differentiated from the other factors' responses.

Tests for common method bias examine the shared variances to determine whether data responses are due to true differences or differences due to methods. For instruments such as questionnaires, method bias is most typically due to participant response style combined with format (response scale, response type, proximity of similar items, etc.) (Weijters, Geuens, & Schillewaert, 2010b). Additionally, Weijters, et al.

(2010a) demonstrated that a 5-point scale measuring attitudes via questionnaire exhibited less method bias than a 7-point scale, providing another validation point.

Summary

This was a study of an existing instrument with the goal of validating the data collected with it or providing statistically sound change recommendations to the owner based on a confirmatory factor analysis and other tests of validity and reliability. Secondary data from events (N = 332 responses) in three locations across the United States were used to test the proposed survey structural model.

Validating the data collected via the survey involved multiple steps, including establishing any correlations or communality between factors, evaluating scale consistency, and examining reliability, all before testing the proposed structure using exploratory, then confirmatory factor analysis, and followed by various tests including convergent and discriminant validity and common method bias analyses.

CHAPTER IV

Results

Characteristics of the Study Sample

The sample consisted of 332 middle school females who completed the survey following participation in one of three Wow! That's Engineering!? events between 2011 and 2013. This sample was subdivided for analysis into $N = 221$ for exploratory factor analysis (EFA) and $N = 111$ for confirmatory factor analysis (CFA). The samples differed somewhat by age and race, but general proportions were retained when purposively separating sample records.

The median age of the total sample was 13 years. Gender distribution was skewed intentionally towards females (97%). The sample's participants self-identified as 19% white, 23% Hispanic, 17% African American, 27% Asian American or Pacific Islander, not quite 2% Native American, and 12% mixed. Less than 1% did not provide the optional information. Each of the EFA and CFA sub-samples maintained the above distribution by age, gender and race/ethnicity.

Item Analysis

Descriptive statistics were used to characterize the responses of participants to survey questions, including mean rating, standard deviation (*SD*), variance, skewness, and kurtosis for each of the survey items shown in Table 3. Distributions are typically considered normal if skewness and kurtosis are less than +/- 1; normal distributions are a typical assumption for many parametric procedures, including factor analysis, especially for the maximum likelihood (ML) method of extraction. If the sample does not severely

Table 3

Item-wise Descriptive Statistics by Descending Mean

	Mean (SD)	Skewness	Kurtosis
Engineers are creative (EDOB)	4.52 (.66)	-1.633	4.274
How likely are you to participate in this event again? (REP)	4.51 (1.06)	-2.113	3.529
Would you recommend this event to your friends? (REC)	4.51 (.98)	-1.848	2.619
Work that is fun to do (FWKD)	4.48 (.87)	-2.153	5.154
Engineers do work that is hands-on (EDOC)	4.45 (.73)	-1.293	2.009
Engineers do work that allows them to help their community and/or society (EDOE)	4.42 (.71)	-1.324	2.639
Engineers are innovative (EDOA)	4.42 (,67)	-1.083	1.797
After this event, I know what an engineer does (KNOB)^{X}	4.36 (.79)	-1.554	3.233
Engineers work in many different kinds of career fields (EDOF)	4.33 (.77)	-.879	.199
My confidence in building and designing things…(ESEB)	4.29 (1.10)	-1.512	1.631
Work that is creative (FWKB)	4.27 (.94)	-1.527	2.446
Work that allows me to help my community and/or society (FWKE)	4.22 (.99)	-1.450	2.061
Work that is hands-on (FWKC)	4.22 (.90)	-1.340	2.171
My ability to think of many different possible ways to solve a problem…(ESED)	4.18 (1.07)	-1.075	.481
My ability to brainstorm solutions to problems (ESEC)	4.16 (1.13)	-1.202	.770
Engineers do work that is fun (EDOD)	4.11 (.86)	-.585	-.307
Work that could be in many different career fields (FWKF)	4.10 (.96)	-1.062	1.009
Work that is innovative (FWKA)	3.99 (.95)	-.991	1.090
My confidence in problem-solving (ESEA)	3.96 (1.16)	-.845	.040
I made some new friends today (ATT1)^{X}	3.95 (1.13)	-1.224	.889
I enjoyed the fact that this was a just girls event (ATT2)^{X}	3.84 (1.23)	-.848	-.276
After this event, I am interested in becoming an engineer (INTB)^{X}	3.63 (.96)	-.270	-.337
Before this event, I knew what and engineer did (KNOA)^{X}	3.56 (1.10)	-.506	-.496
Before this event, I was interested in becoming an engineer (INTA)^{X}	2.98 (1.16)	.153	-.717

Notes.
1. SEX, RACE, and AGE are independent (control) variables and were not used in factor analysis.
2. Variables marked with superscript ^{X} were removed in final model.

violate normality (skewness < 2, kurtosis < 7) and instrument validation tests do not

examine causal relationships between variables or factors, the maximum-likelihood method and its goodness-of-fit indices may apply, as in this study. Item-wise analyses were completed before factor analysis. Other than these descriptive statistics, no item-level tests were conducted since it was the overall survey construct under examination, not whether the program was effective or not.

Internal Consistency

To assess the reliability (internal consistency) of the data produced by the survey instrument, coefficient alpha, (Cronbach's α), was calculated for the set of variables in the model and for each latent factor scale before and after each revision of the factor structure. Table 4 provides details related to the internal consistency of datasets and scales within the data.

Data Adequacy and Factor Analysis

Bartlett's test of sphericity, along with the KMO (Kaiser-Meyer-Olkin) Measure of Sampling Adequacy confirmed the two group samples used in the CFA and EFA as suitable for factor analysis as shown in Table 5.

Model fit. To assess the goodness of fit of the measurement model to the structural model, the maximum likelihood (ML) extraction method was used for both EFA (SPSS 23) and CFA (AMOS 23), applying an oblique rotation (Promax) for factor loadings since the factors are theoretically related. Associated goodness-of-fit indices were re-evaluated every time the model changed during the confirmatory stage. Table 4 also reports the different goodness-of-fit indices used for different cases as identified, noting that at least one measure each from the absolute, parsimony-adjusted, and comparative (or incremental) groups of indices were used in accordance with Brown and Cudeck (1992) and others (Perry, Nicholls, Clough, & Crust, 2015) to ensure the highest construct validity. Note that the only goodness-of-fit index used for EFA is chi-square.

Table 4

Internal Consistency Before, During, and After Factor Analysis

Case	Cronbach's $\alpha > .70$	
	Items	Observed
Full dataset pre-factor analysis (N = 332)	27	.81
EFA dataset (N = 221)	24[a]	.83
CFA dataset (N = 111)	18[a]	.85
Final Future Work Preferences scale (FWK)	6	.90
Final Knowledge of What engineers Do scale (EDO)	6	.89
Final E&T self-efficacy scale (ESE)	4	.83
Psychosocial satisfaction scale (NPS)	2	.70

[a] Independent variables AGE, SEX, RACE removed

Proposed model testing. The first model for exploratory factor analysis using the full set of 24 dependent variables was unconstrained other than to suppress coefficients smaller than .30, resulting in a seven-factor model contributing 63% of the model variance, but with factor 7 having low loading indicators (< .30), factor 6 having only one indicator, and factors 4 and 5 having only two indicators. For the initial EFA, χ^2 (129, N = 221) = 158.296, p = .041.

At the other end, the single factor model produced an unacceptable fit, χ^2 (252, N = 221) = 965.206, $p < .001$, as did the two (χ^2 (229, N = 221) = 642.994, $p < .001$), three (χ^2 (207, N = 221) = 460.319, $p < .001$), and five (χ^2 (166, N = 221) = 258.680, $p < .001$)

factor models. The four-factor model using all 24 dependent variables also had an unacceptable fit (χ^2 (186, N = 221) = 334.299, $p < .001$), so low loading variables were removed (< .4) and the different models re-tested. With the low-loading variables KNOA, KNOB, INTA, INTB, ATT1, and ATT2 removed, the four-factor model resulted in a reasonable fit with (χ^2 (87, N = 221) = 114.887, $p > .01$). This model was used for the CFA since it best approximated the theoretical model, although it also suggested some construct concerns with several of the variables that would have to be addressed with SWE in the future.

Confirmatory model testing. Using the model resulting from the EFA with 18 dependent factors, the four factor model produced the correlations and loadings shown in Figure 2 when using the original EFA sample (*N* = 221) and co-varying error variables e18 and e20. Lubke & Muthan (2004) and Brown (2006) recommend that, when subdividing samples for validation purposes, each sample be evaluated against the best of all samples, so a CFA was also run on the sample set aside specifically or the CFA (*N* = 111) and the full sample (*N* = 332). Of the three sets of data, the full sample best fit the four factor confirmatory model with 18 indicators and Table 5 provides those test of fit results.

The primary reason that the survey owners gave when questioned about the revisions made to the survey over the years was that reducing the number of questions also reduced the perceived burden on the participant, prompting the researcher to further analyze two six-item scales: FWK and EDO. Eliminating the two lowest loading and less reliable items from each resulted in the best fit using the full sample (*N* = 332). Of the

three sets of data, the full sample best fit the four factor confirmatory model with both 14 and 18 items and Table 5 provides those test of fit results.

Table 5

Data Adequacy and Goodness of Fit of EFA and CFA Models

Fit Index	KMO[b] Measure of Sampling Adequacy	Bartlett's test of sphericity	χ^2	χ^2/ df (chi-square/df)	CFI	RMSEA	PCLOSE	ρ^2 (Tucker-Lewis index)
Threshold/cutoff	> .70	$p < .05$	$p > .001$	< 5	> .90	< 0.08[c]	> .05	> .90
Model Dataset *Index type* [a]	-	-	A	A	C	P	P	C
EFA (22 variables, N = 221)	0.79	.000	334.30, p = .000	1.788	-	-	-	-
EFA (18 variables, N = 221)	0.83	.000	114.87, p = .024	1.321	-	-	-	-
CFA (18 variables, EFA sample, N = 221)	-	-	166.97, p = .012	1.304	.97	0.04	.92	.96
CFA (18 variables, CFA sample, N = 111)	-	-	187.09, p = .001	1.462	.94	0.07	.12	.93
CFA (18 variables, full sample, N = 332)	-	-	164.87, p = .016	1.288	.98	0.03	.99	.97
CFA (14 variables, CFA sample, N = 332)	-	-	74.55, p = .333	1.065	.95	0.07	.18	.94

[a] Fit index type: A – absolute, P – parsimony-adjusted, C – comparative.[b] Test named after developers Kaiser, Meyer, and Olkin. [c] From Brown & Cudeck, 1992, p. 239

Figure 2 presents the factor loadings for the 18 variable, four factor model using the EFA sample, Figure 3 represents the CFA sample, Figure 4 represents the full sample and Figure 5 represents the four factor, 14 item model using the full data set (N = 332). Of import is that all loadings are greater than .50 and factor inter-correlations are less than .40, indicating good discriminant validity between factors while still exhibiting sufficiently correlated latent constructs.

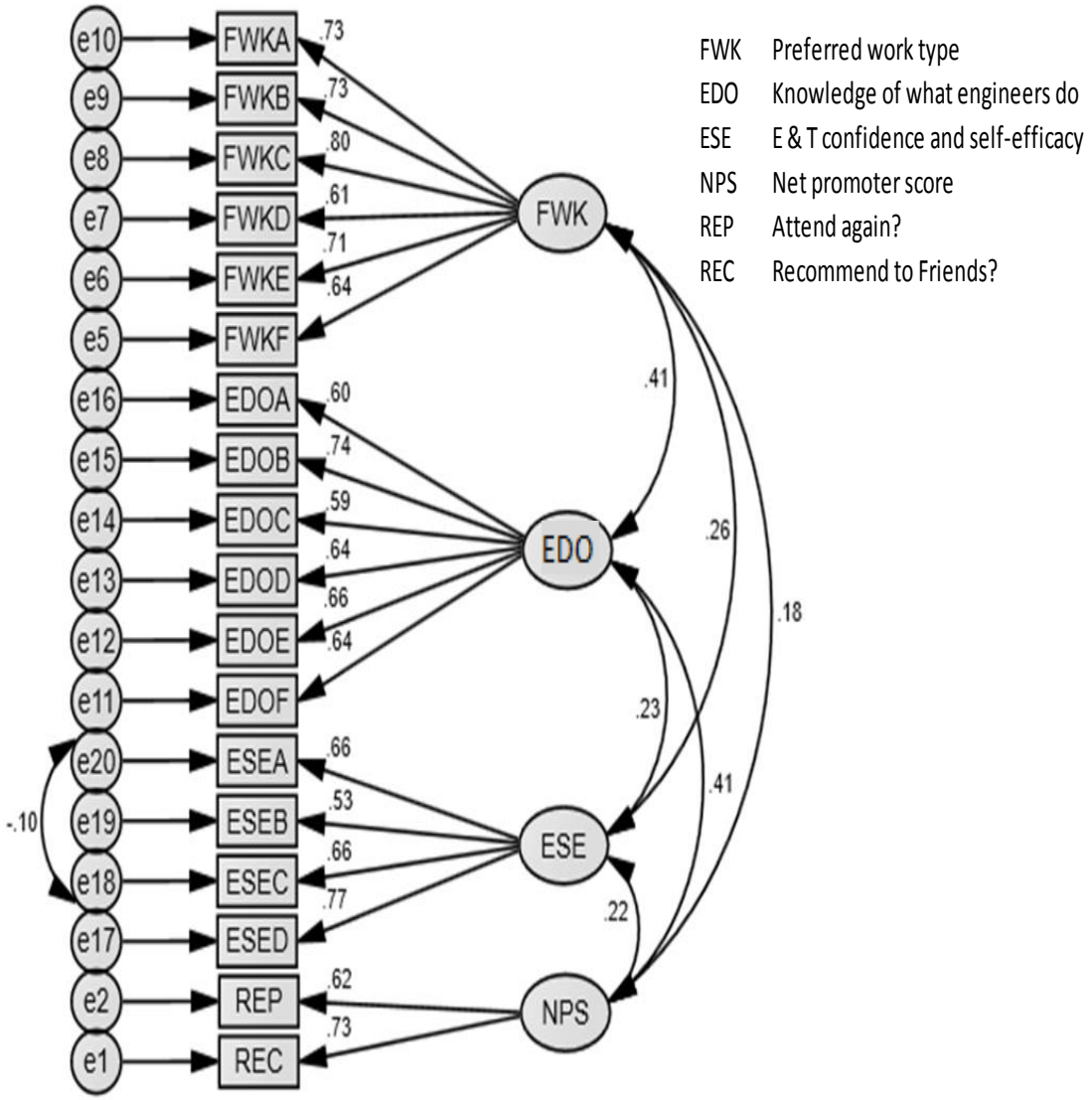

Figure 2. Path diagram representing four factor model of engagement priming using 18-item EFA data sample (*N* = 221).

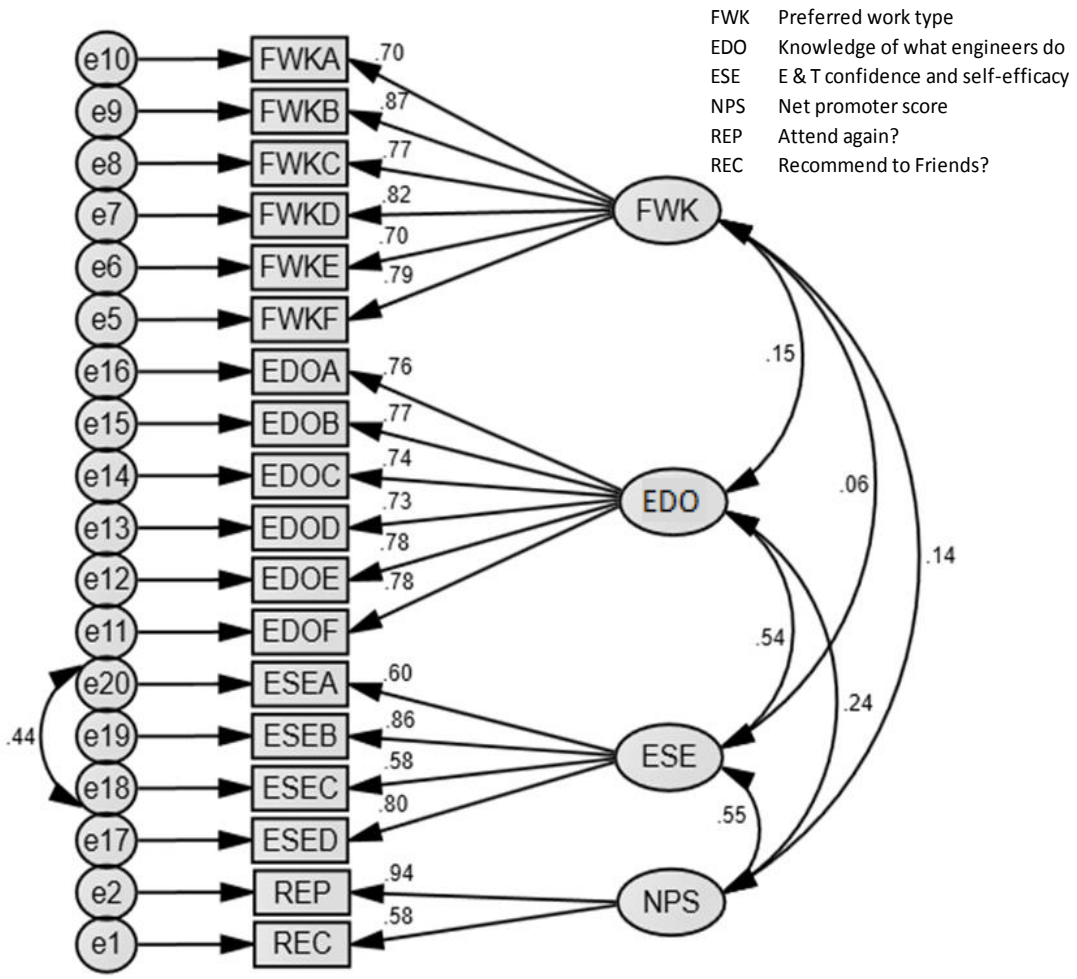

Figure 3. Path diagram representing four factor model of engagement priming using 18-item CFA data sample (N = 111).

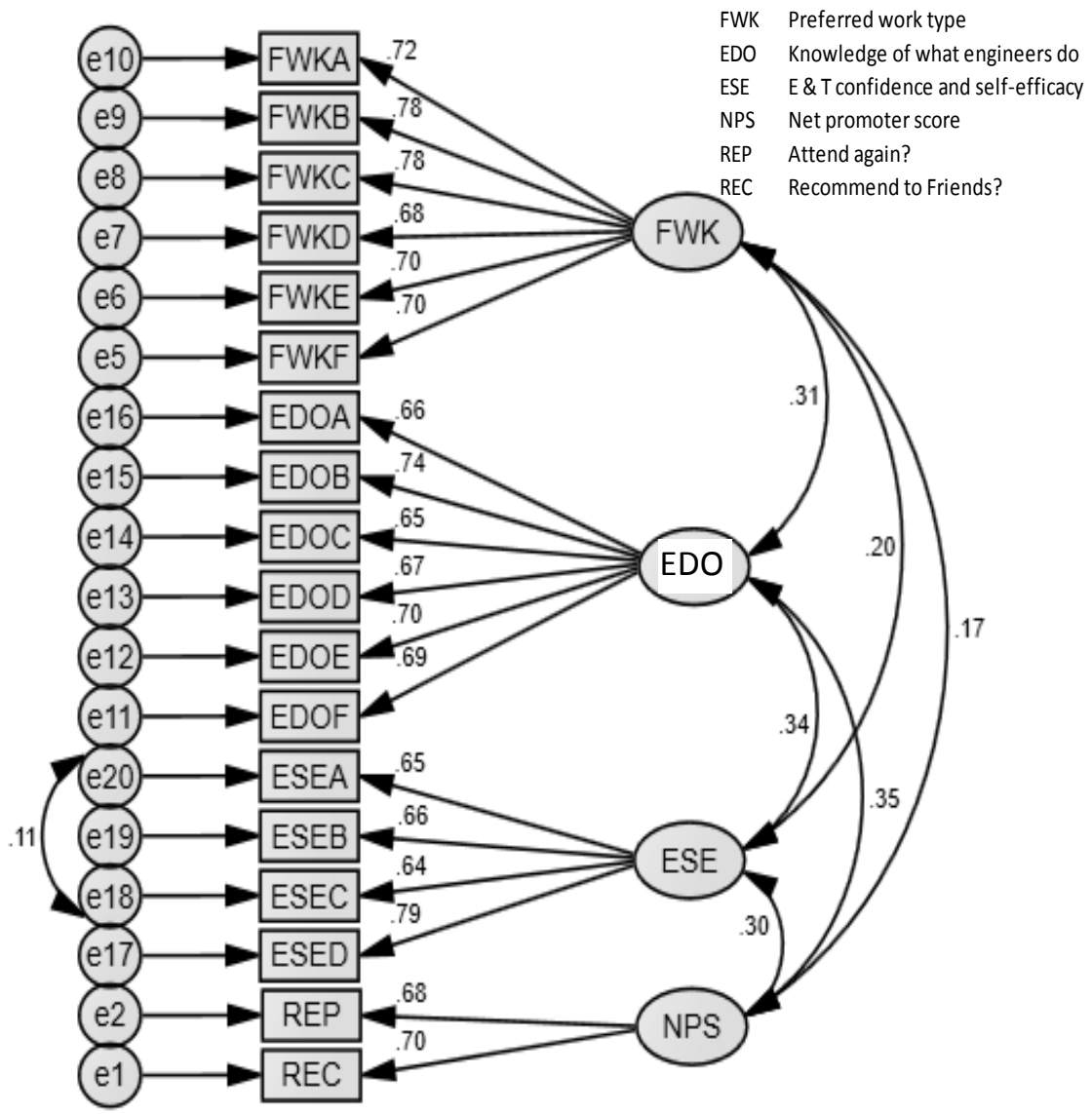

Figure 4. Path diagram representing four factor model of engagement priming using 18-item full data sample (N = 332).

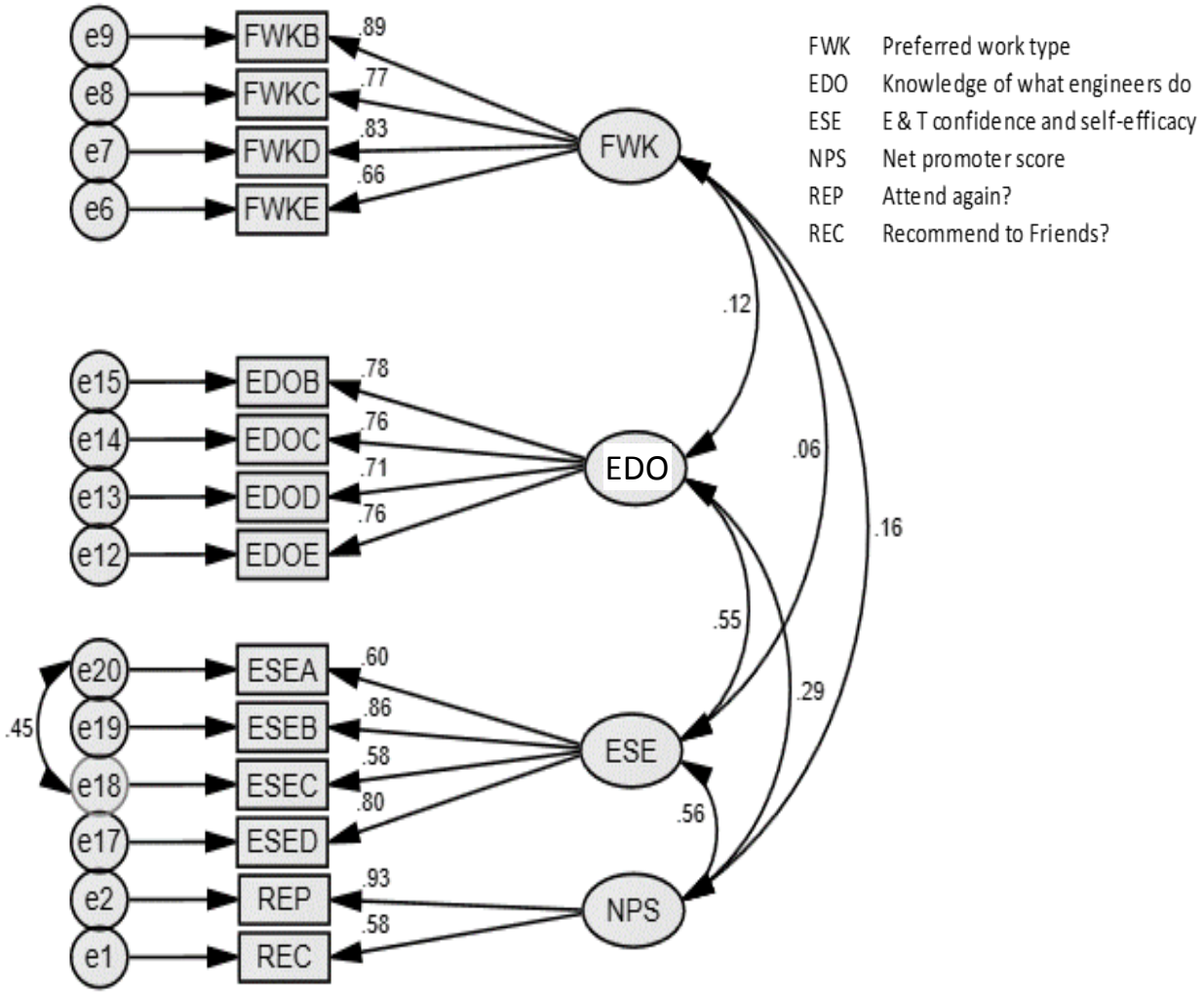

Figure 5. Path diagram representing four factor model of engagement priming using 14-item full data sample (N = 332).

Discriminant and Convergent Validity

Data may fit well to a model and also have high reliability, yet have validity concerns if the factors are either too highly correlated (non-discriminant) or variables are not sufficiently correlated or convergent within factors.

Discriminant validity. According to Farrell and Rudd (2009), discriminant validity is the degree to which

> latent variables discriminate from other latent variables. . . . [and] means that a latent variable is able to account for more variance in the observed variables associated with it than a) measurement error or similar external, unmeasured influences; or b) other constructs within the conceptual framework. (p. 2)

Shared variance is the amount of variance that one variable is able to explain in another, represented by the square of the correlation (or factor loading) between the two variables. Farrell and Rudd (2009) suggest that average variance extracted (AVE) is similar to shared variance in that "the AVE estimate is the average amount of variation that a latent construct is able to explain in the observed variables to which it is theoretically related" (p. 3), while shared variance is "the amount of variance in observed variables relating to another construct that a latent construct is able to explain" (p. 3), where a construct can be either a factor or a variable.

If factor weighting is much less than 0.7, then over 50% of the variable's variance is attributable to either error or other constructs in the model. For discriminant validity, the maximum shared variance (MSV) and average shared variance (ASV) should both be less than the average variance extracted (MSV, ASV < AVE) and inter-factor correlations should be less than the square root of the AVE.

Convergent validity. Satisfactory evidence of convergent validity requires variables within each factor to be highly correlated. Factor loadings represent these correlations. Determining whether loadings are sufficient is dependent on the dataset. Gaskin's StatWiki table (2012) outlines the thresholds for sufficient/significant factor

loadings for convergent validity based on sample size, with smaller sample size requiring higher loadings. He cautions though that "regardless of sample size, it is best to have loadings greater than 0.500 and averaging out to greater than 0.700 for each factor" (Convergent validity, para. 3)

Gaskin also provides an Excel tool to examine convergent and discriminant validity within his Stats Tool Package (Validity master, 2012) that utilizes the correlation and standardized regression estimates from AMOS to tabulate average variance extracted (AVE), maximum shared variance (MSV) and average shared variance (ASV), along with producing the factor correlations table needed for discriminant validity. Recall that the MSV and ASV should both be less than the AVE and inter-factor correlations should be less than the square root of the AVE. The full data sample (N = 332) had both satisfactory discriminant and convergent validity between the factors for the 14-item model as shown in Table 6.

Table 6

Reliability and Convergent / Discriminant Validity of Latent Factors in Full Sample

	α	AVE	MSV	ASV	NPS	FWK	EDO	ESE
RQ1 - NPS	0.740	0.600	0.314	0.142	0.775	--	--	--
RQ2 - FWK	0.870	0.628	0.027	0.015	0.164	0.792	--	--
RQ3 - EDO	0.840	0.568	0.297	0.132	0.290	0.123	0.754	--
RQ4 - ESE	0.807	0.519	0.314	0.205	0.560	0.059	0.545	0.720

Note: AVE = average variance extracted, MSV = maximum shared variance, ASV = average shared variance, NPS = likelihood of recommending, FWK = future work preference, EDO = Knowledge of what engineers do. ESE = E&T confidence and self-efficacy.

Common Method Bias

Although there is heterogeneity present in the sample populations, including different sites, different event content, and different sociocultural contexts, there was enough similarity due to the survey itself that it was appropriate to confirm whether common method bias (CMB) had a significant impact on participant responses. Surveys are notorious for threatened validity due to common method bias (CMB), typically having to do with the way the instrument is formatted or administered.

As Podsakoff, Mackenzie, Podsakoff and Lee aver:

> regardless of whether one considers various rater response styles, item characteristics, and aspects of the measurement context to be "method" factors, they are all sources of systematic measurement error that threaten the validity of a study's findings. Indeed, if they are ignored they can threaten construct validity, distort the dimensional structure of psychological domains, and obscure relationships between constructs/traits (Messick 1991). (2012, p. 542)

Harmann's single factor test was used to assess the degree of CMB, whereby an EFA was modeled using one factor to represent how much of the variance a single factor could explain. If a single factor explained a moderate to large amount of the variance in the dataset, the likelihood of common method bias influencing responses was high. In a four factor model, less than 25% would indicate low common method bias, as in this study (22%). The variety of scale formats used in the survey also helped control CMB due to over-familiarization with a specific scale format, however further work is recommended to explore controlling the effects of participants' response styles, which is more relevant with selection or rating type responses.

Research Questions

Each of the four research questions asks how well the survey produces data responses that fit the theoretical model suggested by the literature (Figure 1, p. 12) and the objectives of the event being evaluated. A summary of the factor analysis is found in Table 6. Of note is that the cross-construct correlations with the other factors were all lower than the square root of the average variance estimated for each factor representing the objective / research question, which is evidence of discriminant validity.

RQ1. *How valid and reliable are the data measuring change in girls' attitudes about engineering and technology?*

This factor (NPS, acronym for net promoter score, which is a measure of customer satisfaction) had the fewest variables (REC and REP) and had the lowest reliability (r = .74), although still deemed reliable by statistical standards. Removal of the low loading variable REC may aid reliability slightly but would detract further from validity by breaking a cardinal rule in survey development of avoiding single item scales (Decoster, 2005; Devellis, 2012). Most of the variables that were intended to load to this factor in the proposed model (such as whether they made friends and their interest before and after the event) were removed from the model due to low loadings on not just this factor, but any.

This suggests model misspecification, confusion on the part of the survey designer or respondent, or simple irrelevance of the items removed. Additionally, having only two items loading to this factor introduces validity concerns. This factor was deemed marginally acceptable in the model and the recommendation is to rework the items feeding it for increased clarity and validity.

RQ2. *How valid and reliable are the data measuring cognitive alignment of future work priorities with engineering and technology career possibilities?*

The data measuring this factor (FWK) produced the most reliable scale (r = .87) with the least amount of shared variance and the highest average loading .79. The four remaining items in this scale, coupled with EDO, may provide the best assessment of how well the respondents connect activities they prefer with career possibilities as engineers and technologists.

RQ3. *How valid and reliable are the data measuring sense of community encouragement and a deeper understanding of what engineers and technologists do*?

One of the issues that often surfaces when translating objectives to research questions is that the original objective may be imperfect. This objective is compound so it may be difficult to interpret whether it is met when evaluating the program (not in the scope of this study): do the data reflect a sense of community encouragement or knowledge of what engineers do, or both?

This ambiguity can threaten content validity, although on the surface, the data produced by this scale (EDO) are consistent (r = .84), with an average loading .76 for the items. It is suggested that this objective be more clearly defined and conjunctive concepts removed to strengthen construct and convergent validity.

RQ4. *How valid and reliable are the data measuring girls' engineering and technology self-efficacy?*

The engineering and technology self-efficacy scale (ESE) produced relatively strong data in four variables with a residual correlation between the first item which measures confidence in problem solving and the third item which measures confidence in

brainstorming solutions to problems. With high internal consistency (r = .81), changing the wording of one or the other item to provide better clarity and differentiation between items may eliminate the high shared variance.

Summary

This study served to validate data from a survey measuring the latent variables that contribute to middle school females' engagement in engineering and technology following a one day immersive engineering and technology event. It also provides a jumping-off point for additional discourse and research on the effects of mentored E & T activities on female engagement in male-dominated career fields such as engineering and technology.

Survey data that demonstrate validity lend the measurement instrument credibility for future studies involving replications, designated E & T activities, role model characteristics, experimental designs involving targeted interventions, and others. The results from this study suggest that, based on the final model, the survey will produce suitably valid and reliable data for use in both program evaluation and research on the effects of mentored activity events on middle school female engagement in engineering and technology.

CHAPTER V

Discussion

The results of this study support the use of the subject survey for measuring middle school females' attitudes about and propensity for engagement in engineering and technology as possible career fields, with certain recommended changes based on validity literature and practice. While the data did not support the specific proposed latent factor path diagram from Figure 1 (page 12), they were highly consistent across samples and groups within samples, revealed several expected correlations, and suggested areas that could threaten validity for further research.

Data Consistency and Validity

When examining data consistency and validity, the underlying concern is whether the instrument used to collect the data was correctly specified to produce results that can be reliably used to predict future behaviors or establish cause and effect. Does the instrument produce responses that do not severely violate normality and are homogeneous across group characteristics?

One of the challenges with working with an extant questionnaire that produces data that has never been validated is that the "custodians" of the survey may not understand the impact that arbitrarily removing or inserting items or questions has on the validity of the instrument, particularly as it relates to making assertions or predictions based on the data. Consequently, over time, the questionnaire may look very different, as in the case of the post-event SWE survey for middle school girls which was first used in 2008 and continues to be used today.

Data consistency. The data sample (N = 332) represented three different sociocultural settings, so it was important to know whether the data were homogeneous across sites, in addition to knowing whether each data sample used in the two factor analyses were consistent with each other. Measures of association, or eta-squared (η^2) represent between group correlations. According to Cohen, the values of ".01, .06, and .14 [should] be used to indicate small, medium, or large associations between variables, respectively" (1988, pp. 280-287). Eta-squared is a comparable measure to Cohen's *d*, and by those guidelines, of the 18 variables across the three sites, seven showed low correlations or effect sizes (mostly in future work preferences FWK) between groups and the remaining showed medium correlations. No correlations were large which would have been problematic and indicative of poor discriminant validity.

The samples used for the exploratory (N = 221) and confirmatory (N = 111) factor analyses were purposively stratified so that each subset reflected similar age, gender, and race distribution as in the larger sample, and that a proportionate number of responses from each site were included in each to ensure homogeneity. The internal consistency of each sample and group was high across scales, although there were some significant item-wise discrepancies between groups. Brown and Cudeck (1992) caution researchers to evaluate such significance conservatively, that is, do not discard a variable simply because there are significant differences between groups, especially if the rest of the scale fits well.

Threats to validity. Many of the survey items have changed over time somewhat randomly, either to intentionally correlate with revised program objectives or due to space constraints. The survey version used for this analysis (late 2011-2013) was the

most stable and relevant due to its continued usage, so little analysis was committed to the older versions.

Sample size. Unfortunately, this also reduced the suitable records from the original data set from 999 to 332, which, while still respectable in sample size by many research standards, was borderline acceptable for factor analysis, especially when dividing further into one sample for exploratory and one for confirmatory analysis. A greater quantity of suitable samples would have been better.

Construct validity. The proposed model was a construct that was informed by the survey itself along with the literature review.

Item issues. During development of the model, the literature review did suggest that several of the items used in the survey to measure interest or psychosocial dispositions might be problematic in that they either did not appear to be specifically related to the objective or required respondents to select a response a priori, as if scoring before the event, immediately followed by the post-event assessment item, for each of interest (INTA and INTB) and knowledge (KNOA and KNOB).

Using a contrived scenario to solicit responses indicating a change from A to B is not credible for temporal results. Instead, asking respondents to rate an item's stem that reads *my interest in engineering and technology increased from this event* or *I know more about what engineers do after this event* may provide a more reliable and valid indicator of change due to the event. The survey designers are also cautioned that such an indication of change is instantial, not longitudinal, and if longer term predictions are of interest, responses from the same participants over time are required.

Low communalities. Several of the survey items had low means and communalities and did not strongly load on any of the factors or scales, including ATT1 (made new friends), ATT2 (prefer girls only), INTA (interest, before event), INTB (interest, after event), KNOA (knowledge of what engineers do, before event), and KNOB (knowledge of what engineers do, after event). Low communalities imply irrelevance. These variables were removed from the model when analysis loadings were set to > .40 with four factors, however no new items replaced them in loading onto the attitudinal scale. Scale developers generally prefer a minimum of three indicators or items, placing the factor NPS at risk with only two.

New items should focus on psychosocial attributes that reflect an indication of satisfaction, belonging, or interest and are correlated to the other two items. Since there are no items in the studied survey that address the influence of role models or mentors, and female role models are asserted to be of significance to younger female engagement in E & T, it is important to get participants' feedback on the relative value of the mentor or role model in their learning and engagement. Such an item might ask the participant to rate how much they agree that *the role model(s) I worked with were inspirational* or how important it was that *role models shared their stories.*

Revised survey. Notwithstanding that many items were revised over the period in question, additionally, a newer version of the event survey was introduced in late 2014 for the Invent it. Build it. event. On the surface, it appears to have at least one potentially significant threat to construct validity in the factor FWK that represents *cognitive alignment of future work priorities with engineering and technology career possibilities.*

In the latest version, the FWK scale was reduced from six variables down to one strictly for brevity, which is not a particularly sound psychometric practice and will likely have ripple effects on the overall factor structure (Gliem & Gliem, 2003). One remedy for that is to add relevant items back in, which would contribute positively towards overall consistency. Since several low loading items in this scale and others were recommended for removal to improve the four factor model fit, the resultant scale should be briefer than the original survey it was extracted from.

Compound items. Several items in the engineering and technology self-efficacy scale ESE identify more than one measurement task in an item or variable stem such as *my confidence in building* <u>*and*</u> *designing things* introduces ambiguity, which threatens reliability and validity. How should the participant respond if they feel more confident about building but not so confident about designing? Depending on their response style, different people respond differently to ambiguous or ill-defined items (misspecifications) in an attitude measurement scale. While this study did not extend to a detailed examination of the impact of response style on data validity, such a study would contribute to a more accurate representation of intra- and inter-factor loadings

Double-barreled items. Fortunately, this survey has few double-barreled questions. Putting an ‘and’ between two ideas in a survey question is often done to consolidate space, but it makes answering the question more challenging because it will likely be confusing for the respondent: how do they respond if they agree with one ‘barrel’ of the question, but not the other? Item ESEB *my confidence in designing and building things* is the only compound item in the survey being examined, but the newest version has an additional item that will need to be revised to ensure data validity: *I*

worked with a mentor who was helpful <u>*and*</u> *easy to talk to*. Additionally, some girls may not understand the word *mentor*, so using the term *volunteer*, *role model* or even *person* may be clearer.

Response style. Specifically, Weijters, et al. (2010a; 2010b) examined participant response styles to determine whether or how much impact they had on rating values and subsequent data validity. Podsakoff, et al. agreed that response style did play a role in biasing the data, but offered mitigation for, or minimally, tests of response style-induced bias typical of self-rated questionnaires as related to:

> (*a*) acquiescence response style (ARS)—calculate the extent of agreement with both positively and negatively worded items in each set (before negatively worded items have been reverse-scored), (*b*) dis-acquiescence response style (DRS)—calculate the extent of disagreement with both positively and negatively worded items in each set (before negatively worded items have been reverse-scored), (*c*) extreme response style (ERS)—calculate the proportion of items in each set on which the respondent endorses the most extreme (positive or negative) scale categories, and (*d*) midpoint response style (MRS)—calculate the proportion of items in each set on which the respondent endorses the middle scale category. (2012, p. 558)

Several items utilize the *agree/disagree* format which often exhibit stronger effects of ARS and DRS, but whether that is relevant or not is for future study. Weijters, et al. also noted that there is a lower level of mid-response style bias from items using a 5-point response format than from a 7-point format (2010a).

Proximity of measures. Podsakoff, et al. also suggest that "researchers should try to position measures of the same-construct at least six items apart, separated by measures of other constructs using the same or different formats, or by means of dedicated buffer items" (2012, p. 550). While it may not be practical, it may assist in enhancing the convergent or discriminant validity of data collected with the instrument.

Method bias. For evaluating and controlling bias related to method, they further offer that

> researchers should consider the extent to which their questions fail to define ambiguous or unfamiliar terms, refer to vague concepts without providing clear examples, have complicated syntax, or are double-barreled. In addition, Krosnick (1991) notes that item ambiguity is greater if only the end points of a response scale are labeled (rather than every point). (p. 561)

Ambiguous definitions. Several of the items in this study are examples of one or more of the cautions from Podsakoff, et al., which, if heeded carefully, should result in improved validity. Examples include the ambiguity of measures ESEA and ESEC, the first asking respondent to rate improvement in problem solving skills after the event, and the second asking them to rate improvement in their ability to brainstorm solutions to problems, two different but functionally related concepts that provide measures of engineering and technology self-efficacy. Finding ways to express individual concepts more distinctly is key to improved validity.

Complicated syntax. Another example is the complicated syntax for item FWKA: *Work that is innovative (where you can come up with new ideas and inventions).* How is that different from *work that is creative?* 12 and 13 year old children do not have the

cognitive sophistication to detect that level of nuance, so are very likely to score one the same as the other from a response style perspective. If their meaning is intended to be different, more clarity is needed to differentiate the two items for ratings; if not, one item should likely be removed.

Researcher bias. Although not directly related to construct validity, the researcher's intimate knowledge of and voluntary participation in the leadership of the Society of Women Engineers (and events such as the one the measurement model of this study is based on), potentially bias any inferences drawn from the secondary data evaluated. As no causal relationships are being tested or inferred in this study, however, the effects of such bias can be effectively moderated by the current leadership of SWE.

Proposed Factor Model

The proposed model was based on literature-informed analysis of the extant survey, coupled with the objectives of the event the survey was supposed to measure to assess whether the program met its objectives. Over the years, much discussion has surfaced from time to time regarding the Society's ability to measure the impact their outreach events have on young females and how that might translate to improved engagement, which is the first step to cultivating interest and commitment. Since 2008, SWE has been collecting data based on that post event survey, but until now, the data was never validated in support of broader usage within and outside of SWE events.

Event objectives as factors. When the researcher proposed a four factor model to mirror the four objectives being measured for the event, there was some concern with the original instrument developers' mild confusion between learning and survey objectives

which translated into several weak loadings or misleading acceptability of a seemingly sufficient over-identified model.

Objective 1: Facilitate change (NPS). The first event learning objective states *facilitate a change in attitude about careers in engineering and technology*, yet the only attitudinal items which appear to have both a pre- and post- aspect to them have to do simply with degree of interest in engineering. The only other before and after items ask about understanding of what engineers do, which although not attitudinal, do link most strongly with the concept of change due to their pre- and post- nature.

The post-event survey, which is completed with pencil and paper before girls leave the event, asks the girls to rate how much they agree with several statements on a five point scale. The first item states *Before this event, I knew what an engineer did*, while the next item replaces *Before* with *After*, *knew* with *know*, and *do* with *does*. The next two items use the same mechanism for interest in becoming an engineer. From a construct validity standpoint, such an approach is flawed for at least two reasons.

1. All four were rather weak variables in the scheme of factor analysis, with communalities of less than .300 and factor loadings less than .400. Since the difference between the two readings was really of concern, the two variables together are essentially a single indicator, which, according to Brown (2006) does not meet the "minimum of three indicators per latent variable" (p. 72) for just- or over-identification.
2. The validity of self-reported assessments of attitudes is continually being challenged as being biased, so, when coupled with asking pre-pubescent females to objectively remember how they felt about engineering six hours

earlier and, in the next item, how they feel now, responses are not likely to be normal or reliable. It is likely the responses would be more consistent if the survey asked how much they agreed with *my interest in engineering and technology increased today* to assess the change between Before and After.

This change would not address the fact that there are still too few variables for the latent variable NPS, so it is recommended that, if one of the survey objectives continues to be to measure a change in attitudes, several more items/indicators/statements get added to contribute to a valid and reliable change in attitudes about engineering and technology factor, otherwise the factor as defined may not be of much credible value.

Objective 2: Build E & T self-efficacy (ESE). This survey objective is to measure girls' self-confidence and critical thinking skills having to do with engineering and technology principles following guided hands-on activities. Although the survey does not make use of the before-and-after mechanism used for objective 1, it does use a different form of Likert-scale in asking whether confidence levels *improved, got worse, stayed the same* or *I don't know*, which required interpretation as to what constituted endpoints (*improved, got worse*) and how to score the in-between on first a 4-point scale, then convert to 5-point. Certainty (*stayed the same*) was ranked higher than uncertainty (*I don't know*).

This objective was pretty straight forward except for some high shared variance (r^2 = .45) between the indicators *my confidence in problem-solving* and *my ability to brainstorm solutions to problems*. Girls this age may not be cognitively able to discriminate between self-confidence and ability and may interpret them as the same concept. Additionally, this factor did have one compound variable in it *my confidence in*

building AND designing things that may pinpoint confidence issues better if broken into one item for building and one for designing. In literature and practice, while females are beginning to develop improved confidence levels with respect to design, confidence in physical manipulation of materials to produce a prototype of a design has not increased correspondingly yet (Phelps, 2012), and linking them together may lead to flawed results.

Bandura (2006) cautions that asking children to self-assess ability is an imperfect process and that objective performance measures may be a better way to assess skills, either with a quick test of the skills in question or an assessment from a third party.

Objective 3. Develop knowledge of what engineers do (EDO). This was another compound objective in that the ultimate goal was to use role models to *encourage an expanded sense of community AND a deeper understanding of what engineers and technologists do*. Since only the 2014 version of the survey used on another signature event even touches on the value of role models to encourage sense of community (one item), and there were four strongly loading items here that examined knowledge of what engineers do, the factor was redefined to reflect the delimitation.

Unfortunately, that meant that part of this objective, to measure whether the events' role models were able to *foster relationships to build a sense of community*, was not met as there are no items which examine the mentors/role models effect in this version (there is one item in the latest survey version). Perhaps their effect can be implied from the girls' responses to other questions, but there is no direct linkage in the survey data themselves.

It is recommended that either a separate scale and factor be created to specifically measure girls' attitudes towards role models, with at least three items, or that three items

get added to the first objective NPS. A good start for what to include would be examining the construct from the role model survey administered at the 2014 Invent it. Build it. event to make sure we are asking girls to assess mentors for the same things that SWE has determined to be important for outreach role models to exhibit.

Objective 4: Connect future priorities with engineering and technology (FWK). This objective exhibits some of the same issues as the previous in that underlying the objective is the implied role of the role models in facilitating a connection between the girls' interests and preferences and career possibilities in engineering and technology. Additionally, there exists a connection between *by sharing personal stories AND celebrating the accomplishments of women engineers and technologists*. Even if there were items in the survey being examined that measure this objective (there are none in this version), how does the evaluator tell when that is accomplished or whether it occurred because of the *personal stories* or *celebrating accomplishments*? As in the previous objective, a strong recommendation is to create a mentor assessment scale that incorporates these two criteria, in addition to the *fostering relationships to build a sense of community* from objective 3.

The factor for this objective then became alignment between engineering and technology careers/work and personal values, which was expected to include loadings from items reflecting psychosocial attitudes such as whether the participant would participate again or recommend the event to their friends, whether they made new friends, and whether they enjoyed that is was a girls-only event, in addition to six indicators of work preferences. In the EFA, most of the psychosocial factors ended up

being removed from the model because of weak communalities and low correlations with each other and the latent factors.

Consequently, the remaining psychosocial factors (*Would you recommend this to friends?* and *Would you participate in this event again?*) loaded on a new factor the researcher labeled NPS (for net promoter score) which replaced the former Change in Attitudes factor, but which was also imperfectly identified with only two variables loading on it, and one of them rather weakly (< .500). Since this factor is an amalgam of their own attitudes (psychological) and their perception of the attitudes of others (sociological), it is recommended that at least two more items/indicators be developed which measure psychosocial constructs such as *My family and friends would support my career in engineering or technology* or *A career in engineering or technology would suit me personally*.

Recommendations

On the whole, the instrument appears to satisfactorily measure most of SWE's event objectives. Although there are several validity concerns in the examined instrument, most could be expected to improve through changes in the assessment process used by the Society and/or the instrument itself.

Face or content validity. There are no performance-related data collected through this instrument (such as test scores or project portfolios) that might provide an objective assessment of the participants' level of knowledge of what engineering is or what engineers do. The scale that asks participants to subjectively rate their agreement with statements about the field of engineering is limited by the "truth" that exists in the participants' minds based on their psychosocial state at the time of answering, a truth that

may be both duly and unduly influenced by what they ate for breakfast, what they saw on the way to school, or whether they are only kids, among a multitude of other possible factors such as economics, religion, race, gender, age, etcetera.

Bandura says it is to be expected that children's attitudes and self-efficacy will change over time (2006) but they are most influenced (and likely to be influenced significantly by family) in the early years, especially with respect to gender norms (Bandura & Bussey, 1999). Measures of propensities to achieve (what Bandura calls self-efficacy) within children are challenging to validate; they could like something today and hate it tomorrow. However, if there is little at stake, they may be less inclined to prevaricate when completing their assessment. If there is a lot at stake (test, scholarship, award, etc.), like many humans, they will often tell the assessor what they think he/she wants to hear, as early as pre-school (Bronson & Merryman, 2009). How valid the data from children is depends on what is being assessed and how clear the instructions.

In fairness to the instrument's authors, however, their charter was not to provide valid data for researchers, but rather to measure the general effectiveness of a specific outreach program, which resultant data is likely sufficiently valid to accomplish that. Whether the results would be considered valid to other researchers generally comes down to the numbers: how many items, factors, participants, raters, methods, etcetera? What is the reliability, validity, normality, covariance of the data?

When examining concepts such as engagement and self-efficacy within the framework of engineering and technology fields and through the lens of gender, being able to make reliable predictions about girls' future decisions about career possibilities based on a single response after a single event is ambitious and fraught with the potential

for flaws. Such data can however provide a baseline which can be followed through their career-formative years.

Research on female engagement in engineering and technology speaks strongly to the sense of belonging that drives both self-efficacy and engagement. In many cases of successful exemplars it was a sense of belonging to engineering or technology, but not necessarily to the environment in which it was performed, designed, or created that drove them to find their own ways of being successful and often involved a relationship with the Society of Women Engineers or other social organizations that shared their mission.

Just as females who persisted in E & T careers primed their own interest and engagement through multiple opportunities with other women over their careers, longitudinal data offers the potential to map girls' attitudinal changes over time. Therefore, it could be valuable to establish a way to maintain continuity with event participants over time, particularly if inferences or high stakes decisions could potentially be made from the resulting data.

Recommendation 1. To address the secondary purpose of the survey (measuring impact), re-assess participants at one through eight years post-event to determine residual effects of the event on participants' level of engagement. Control for intervening outreach events and determine whether number of events is a moderator of engagement.

To simplify future data analysis, either a CFA of a revised survey or interpretive analysis of program data, there are several mechanistic steps that can be taken to minimize invalidity.

Recommendation 2. Ensure the response format for all scales within the instrument are consistent. Note that 'the same' is not necessary, but researchers Colman, Norris, & Preston (1997) suggest that response formats should ALL

- be 5- or 7- point optimally , and
- labeled gradations should be used such as *strongly agree* to *strongly disagree*, with avoidance of *yes*, *no*, or *I don't know* responses, which are more filters than directive and contribute little to analytical interpretations or conclusions, rather broadly slotting the data into groups.

Recommendation 3. Ensure subscales have a minimum of three items. This comes from good practice, especially when measuring psychosocial concepts. More items than five or six are not necessary if the concept is less complex, but single-item scales are to psychosocial and educational research as dividing by zero is to mathematics, a non-starter from a validity perspective. Additionally, ensure that items are not redundant. High communalities (above .85) are sufficient cause to question whether the items are discriminant enough from each other to offer independent-yet-related responses.

Recommended model. The final confirmed model comprises four factors as did the proposed model. One factor (NPS) is conceptually somewhat different than the originally proposed, likely due to misspecification of the proposed model, which produced weak loadings of several variables (< .400), begging the question of whether they were needed to measure the objectives. These included all of the psychosocial measures: ATT1 *making new friends*, ATT2 *just girls event*, INTA/B *improved interest in E & T*, and *improved knowledge of E & T* KNOA/B.

The modified factor has been labeled NPS for *net promoter score*, which in this sense is a measure of psychosocial predisposition to the career possibilities offered by technology and engineering. Adding two items for NPS that measure the psychosocial relevance of family and peer approval of career decisions would likely boost the salience of w*ould you recommend this event to your friends?* (r = .58) and w*ould you participate in this event again*? (r = .92), as would asking the respondent to rate agreement with *my interest in E & T increased after this event* rather than whether their interest went up, down, or stayed the same.

Recommendation 4. Add a minimum of two items to factor NPS that reflect a measure of level of psychosocial acceptance of E & T career decisions or interest, such as *my family and friends would support my career in engineering or technology* or *a career in engineering or technology would suit me personally*.

Building a survey that produces valid data regarding middle school females' propensity for pursuing a career in engineering and technology is but one step in enacting effective interventions that make a difference in female engagement. Meanwhile, there are other barriers to female engagement that are not directly related to middle school females' decisions, but which have been demonstrated to impact psychosocial attitudes.

Eliminating Psychosocial Barriers

As Becker and Swim present in their 2011 study on benevolent sexism, although woman have come a long way from consideration as property or chattel, residual sexist attitudes about the roles of males and females are deeply embedded and in survival mode, that is, tenacious, as society creeps inexorably towards gender equality. As in the arduous

global sociopolitical journey towards race equality, there are often two steps forward and one step back.

Becker and Swim discovered that the benevolent and modern sexist attitudes of 38 male and 82 female undergraduates were influenced by daily journal reflection on sexist occurrences, with a significant correlation for women only; in addition to awareness of sexism, men required the additional stimulus of empathy induced by describing women's feelings as acts of sexism occurred, in order for their benevolent and modern sexism attitudes to change.

Many researchers and practitioners alike attest to the value of first, increasing awareness of prejudicial issues, then invoking empathy in order to reduce it (Aosved, Long, & Voller, 2009; Becker & Swim, 2012; Campbell, Schellenberg, & Senn, 1997; Cheryan, Davies, Plaut, & Steele, 2009; Chesler & Chesler, 2002; Jackman, 1994; Jones, Doveston, & Rose, 2009; Jost & Kay, 2005). Others recommend introducing cognitive support structures such as mentors and physical or online communities of practice that may provide psychosocial resonance and help offset some of the negative affect associated with discrimination (Allen & Eby, 2004; Amelink, 2008; Buday, Stake, & Peterson, 2012; Cheryan, Plaut, Handron, & Hudson, 2013; Chesler & Chesler, 2002; Holmes, Redmond, & Thomas, 2012; Jones, Doveston, & Rose, 2009; London, 2011; Pisimisi & Ioannides, 2005).

Stereotype inoculation model (SIM). Dasgupta's stereotype inoculation model (SIM) (2011) is cited by many as key to providing the relevant support that females need to combat stereotypes and its resultant sexist manifestations. Her model comprises

> four interrelated processes [which] are predicted to drive stereotype inoculation of the self-concept when people encounter ingroup experts and peers: enhanced sense of belonging, self-efficacy, increased challenge, and reduced threat. . . . [and] proposes that diversity cues in the environment may involve successful ingroup experts who are advanced in their career or one's peers in an achievement context . . . . by virtue of personal contact with these individuals or mediated exposure from afar. (2011, p. 239)

She and her colleagues assert that transitional periods such as that between K-12 and undergraduate or between college and work are when in-group experts are most influential (Stout, Dasgupta, Hunsinger, & McManus, 2011). She also recommends that organizations craft or re-craft

> policies that promote interpersonal contact and . . . . mentoring relationships with ingroup experts . . . . to have the strongest effects on newcomers' sense of belonging, self-efficacy, and engagement in the domain because contact personalizes ingroup experts and makes it easier for beginners to identify with them as similar others and not view them as unattainable superstars. (Dasgupta, 2011, p. 242)

Cultural awareness. Bandura and Bussey (1999) cautioned that, while modern communication devices such as TV and internet provide a much needed broadcasting vehicle, they also perpetuate social structures which support gendered dichotomies that assume all males have masculine traits and all females have feminine traits. As discovered by the National Academy of Engineering in their seminal *Changing the Conversation: Messages for Improving Public Understanding of Engineering* (2008), in

spite of all of the technological advances that have occurred in society in recent years, most Americans were challenged to articulate what engineers did.

Most did not relate it to helping or making a difference, two concepts that resonated more with girls than boys, likely significantly based on historical gender stereotyping that females are more people-oriented than males. Media representation of engineers, and especially subscription to stereotypes, has a negative impact on female recruitment and retention in male-dominant fields such as engineering (Bystydzienski & Brown, 2012; Cheryan, Siy, Vichayapai, Drury, & Kim, 2011; Cheryan, Davies, Plaut, & Steele, 2009; Cheryan, Plaut, Handron, & Hudson, 2013; Drury, Siy, & Cheryan, 2011; Eccles, 2011; Frome, Alfeld, Eccles, & Barber, 2006; Larose, et al., 2008; Patterson, Campbell, Busch-Vishniac, & Guillaume, 2011).

As Daniel Goleman, author of *Emotional Intelligence* (1995), captured in his 1984 NY Times interview with researcher on bias Anthony Greenwald, humans are egocentric or self-serving, something Greenwald terms “beneffectance” (p. 1), which combines *beneficence* (doing good) with *effectance*, a psychological term for competence. In essence, beneffectance represents the human tendency to take personal credit for good things happening, while looking outside oneself to explain personal failures. Men are generally better at achieving beneffectance than women, who tend to blame themselves for something going wrong in someone else’s life. Consider how frequently after someone jostles you or bumps into you that you hear women apologize as compared to men (Schumann & Ross, 2010).

As Steele points out in *Whistling Vivaldi* (2010), being confronted with the possibility of prejudice, many humans either deny or reject evidence of their own

complicity, often rationalizing their behaviors and attitudes so that they make some convoluted sense of it, known as system justification (Jost & Banaji, 1994; Jost, et al., 2009; Kilianski & Rudman, 1998; Betz & Sekaquaptewa, 2012). Often the choices or responses to self-enacted and observed behaviors around them are automatic and involuntary until the perceiver becomes aware that he or she has control over them (Bandura, 1986). With awareness comes control, and finally mastery of one's bias.

The results of Becker and Swim's study of ways to reduce sexism's effects suggested that "women endorse sexist beliefs because they lack recognition of subtle forms of sexism, discount sexist incidents, and do not notice the aggregate amount of sexism in their daily lives" (2012, p. 239), which is especially evident in male-dominant occupations where sexism is more rampant than in sex-typed occupations. Once made explicitly aware of the subtle sexist cues that proliferate in their environments, women tend to reject the seemingly benevolent behaviors and the people who perpetrate them, although men must be first directed to role-play female emotions to sexist behaviors before their benevolent sexism attitudes lessen.

In *A Class Divided* (Peters, 1985), teacher Jane Elliot demonstrated the power of prejudice and stereotypes with her classroom experiment in which boys and girls were treated differently dependent on their eye color. In the experiment, her explicit eye color discrimination of first one group, then the other, enabled all of the children in the class to personally experience the isolation, negative emotions, and lowered self-esteem that goes with prejudice from the in-group towards the out-group, inducing a psychosocial awareness which impact and influence was still felt by many nearly fifteen years later.

Surveys such as the one studied here have the potential to begin laying the groundwork for predicting long term psychosocial change related to female engagement in engineering and technology careers. Demonstrating the validity of the data produced via the one in this study was the first step on that journey.

Future Research

Future research comprises two primary topical areas: further research related to the current instrument and / or data set and research related to the application of the instrument outside the WOW framework.

Current instrument. As demonstrated in this study, validating the data collected via an instrument is an ongoing process and requires re-confirmation every time the instrument changes significantly. What is significant? Anything that could introduce a threat to validity, such as under-identifying a factor (too few items loading onto it). Additional confirmatory analysis should be completed every time the instrument's authors remove, add or change the wording of questions.

Given that the final recommended model with 14 items is somewhat different than the starting point with 24 items, one study of interest might examine how the raw data (N = 999) fits a partial factor model. This could provide an additional validation input for the model. Another study might evaluate the presence of higher order latent factors, that is, do two or three of the four factors actually load onto another heretofore unidentified factor that loads onto the E&T priming factor (EPF)?

Current data set. With validated data, the efficacy of the WOW program can be credibly evaluated and correlations between factors and items examined for contextual significance. Such results could be used to further revise the theoretical or measurement

model. Additional examining differences between groups such as event site and race or ethnicity could provide insights that inform instrument generalizability.

Non-WOW instrument usage. The current instrument was the basis for the Invent it. Build it. (IIBI) immersion event survey, although the most recent version (2014) has some marked differences which will likely reflect on validity. Evaluating the significance of those differences is suggested prior to further usage, especially if program results are published, using 2014 data (N = 544).

Longitudinal studies provide an opportunity to measure test-retest reliability along with potentially enabling researchers to infer predictive validity of one or more factors. Administering the survey to event participants immediately following an event, then again at fixed intervals over time (say, every two or four years up to however many years it takes for participant to enter final career) provides a set of data points that may better inform the relationship between the factors and engagement.

Summary

This study examined the theoretical and measurement models represented in a survey for middle school girls developed by the Society of Women Engineers to measure program and organizational effectiveness. Analytical tests included evaluating the reliability and validity of a purposive sample of responses (N = 332) collected between 2011 and 2013 via the instrument.

Ultimately, the goal of the survey is to collect data that demonstrates a salient connection between the activities that occurred at events it is administered at and the development of a positive attitude in the respondent about engineering and technology as possible future career fields, in short, a priming factor that is comprised of several

different latent factors that tie to event and program objectives. Deconstructing and assessing whether the survey questions precisely and accurately measured those key objectives was the functional goal of the study (defining what is).

The survey data evaluated in this study have successfully validated a four factor construct, albeit not exactly as proposed with respect to item loading, however three out of four factors have a sufficient number of items loading, while the fourth factor, the psychosocial one (NPS) would likely see an improvement in scale consistency (reliability) and item salience with additional items that better reflect the first program objective of positive change in interest or attitude.

The future holds several possibilities including re-constructing and revalidating the survey for improved validity (especially in the area of psychosocial indicators that may better predict long term engagement, including the effects of role models); completing a program assessment using the full dataset with high confidence in the results; or performing a similar validation exercise on other outreach assessment instruments to ensure they are measuring what they are supposed to measure. The possibilities are endless and the Society can now be confident that they are reliably measuring most of their objectives with the existing survey.

References

Bureau of Labor Statistics, U.S. Department of Labor. (2014). *DOL > WB > Data & Statistics > Latest Annual Data* . Retrieved from United States Department of Labor: http://www.bls.gov/opub/ooq/2011/winter/art02.pdf

Allen, T. D., & Eby, L. T. (2003). Relationship effectiveness for mentors: Factors associated with learning and quality. *Journal of Management, 29,* 469-486. doi:10.1016/S0149-2063_03_00021-7

Allen, T. D., & Eby, L. T. (2004). Factors related to mentor reports of mentoring functions provided: Gender and relational characteristics. *Sex Roles, 50*(1/2), 129-139.

Allendoerfer, C., Wilson, D., Bates, R., Crawford, J., Jones, D., Floyd-Smith, T., . . . Veilleux, N. (2012). Strategic pathways for success: The influence of outside community on academic engagement. *Journal of Engineering Education*, 512-538.

Allwin, D., & Krosnick, J. (1991). The reliability of survey attitude measurement: The influence of question and respondent attributes. *Sociological Methods & Research, 20,* 139-181. doi:10.1177/0049124191020001005

Amelink, C. (2008). *Overview: Mentoring and women in engineering.* Retrieved from http://www.AWEonline.org.: http://www.engr.psu.edu/awe/misc/ ARPs/ARP_Mentoring_overview120408.pdf

American Bar Association. (2013, March). First year and total J.D. enrollment by gender: 1947-2010. Retrieved from Knowledge Center - Catalyst: http://www.catalyst.org/ knowledge/women-law-us

American Bar Association Commission on Women in the Profession. (2013). A current glance at women in the law. Chicago, IL: American Bar Association.

American Psychological Association. (2014, January). 2013: APA member profiles. Retrieved from APA Center for Workforce Studies: http://www.apa.org/workforce/publications/13-member/index.aspxAntecol, H., & Cobb-Clark, D. A. (2013). Do psychosocial traits help explain gender segregation in young people's occupations? *Labour Economics, 21*, pp. 59-73. doi:10.1016/j.labeco.2012.12.005

Antill, J. K., Cunningham, J. D., & Cotton, S. (2003). Gender-role attitudes in middle childhood: In what ways do parents influence their children? *Australian Journal of Psychology*, 148-153.

Aosved, A. C., Long, P. J., & Voller, E. K. (2009). Measuring sexism, racism, sexual prejudice, ageism, classism, and religious intolerance: The intolerant schema measure. *Journal of Applied Social Psychology, 39*(10), 2321–2354.

Asgari, S., Dasgupta, N., & Stout, J. G. (2012). When do counterstereotypic ingroup members inspire versus deflate? The effect of successful professional women on young women's leadership self-concept. *Personality and Social Psychology Bulletin, 38*(3), 370–383. doi:10.1177/0146167211431968

Askari, S. F., Liss, M., Erchull, M. J., Staebell, S. E., & Axelson, S. J. (2010). Men want equality, but women don't expect it: Young adults' expectations for participation in household and child care chores. *Psychology of Women Quarterly, 34*, 243–252.

AWE. (2008). *Longitudinal Assessment of Engineering Self-Efficacy (LAESE) - High School Version.* Retrieved from Assessing Women and Men in Engineering: https://www.engr.psu.edu/awe/secured/director/diversity/efficacy.aspx

Bakir, A., & Palan, K. M. (2010). How are children's attitudes toward ads and brands affected by gender-related content in advertising? *Journal of Advertising*, 35–48.

Bandura, A. (1971). *Social learning theory.* New York: General Learning Press.

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

Bandura, A. (Ed.). (1995). *Self-efficacy in changing societies.* New York: Cambridge University Press.

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

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

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares, & T. Urdan (Eds.), *Self-Efficacy Beliefs of Adolescents* (pp. 307–337). Greenwich, CT: Information Age Publishing. Retrieved from http://www.uky.edu/~eushe2/BanduraPubs/ BanduraGuide2006.pdf

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science, 1*(2), 164-180.

Bandura, A., & Bussey, K. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review, 106*(4), 676-713.

Bandura, A., & Bussey, K. (2004). On broadening the cognitive, motivational, and sociostructural scope of theorizing about gender development and functioning: Comment on Martin, Ruble, and Szkrybalo (2002). *Psychological Bulletin, 130*(5), 691–701. doi:10.1037/0033-2909.130.5.691

Barnard, S., Hassan, T., Bagilhole, B., & Dainty, A. (2012). They're not girly girls: An exploration of quantitative and qualitative data on engineering and gender in higher education. *European Journal of Engineering Education, 37*(2), 193-204.

Becker, J. C., & Swim, J. K. (2011). Seeing the unseen: Attention to daily encounters with sexism as way to reduce sexist beliefs. *Psychology of Women Quarterly, 35*(2), 227-242.

Becker, J. C., & Swim, J. K. (2012). Reducing endorsement of benevolent and modern sexist beliefs. *Social Psychology, 43*(3), 127-137. doi:10.1027/1864-9335/a000091

Bell, A. (1970). Role modelship and interaction in adolescence and young adulthood. *Developmental Psychology, 2*(1), 123-128.

Beltz, A. M., Swanson, J. L., & Berenbaum, S. A. (2011). Gendered occupational interests: Prenatal androgen effects on psychological orientation to things versus people. *Hormones and Behavior, 60*, 313-317.

Bem, C., & Lenney, E. (1976). Sex typing and the avoidance of cross-sex behavior. *Journal o] Personality and Social Psychology, 33*(1), 48-54.

Bem, S. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 155-162.

Bem, S. (1975). Sex role adaptability: One consequence of psychological androgyny. *Journal of Personality and Social Psychology, 31*(4), 634-643.

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88*(4), 354-364.

Bem, S. L. (1981). The BSRI and gender schema theory: A reply to Spence and Helmreich. *Psychological Review, 88*(4), 369-371.

Benjamin, C. (2014, 1 25). Dear LEGO company. *The Society Pages: Sociological Images*. Retrieved from http://thesocietypages.org/socimages/2014/01/31/this-month-in-socimages-january-2014/

Betz, D., & Sekaquaptewa, D. (2012). My fair physicist? Feminine math and science role models demotivate young girls. *Social Psychological and Personality Science*, 1-9.

Bjorklund, D. F., & Brown, R. D. (1998). Physical play and cognitive development: Integrating activity, cognition, and education. *Child Development, 69*(3), 604-606.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246 – 263.

Blickenstaff, J. (2005, October). Women and science careers: leaky pipeline or gender filter? *Gender and Education, 17*(4), 369-386.

Bobko, P., & Schwartz, J. P. (1984). A metric for integrating theoretically related but statistically uncorrelated constructs. *Journal of Personality Assessment, 48*(1), 11-16.

Bogue, B., Cady, E. T., & Shanahan, B. (2013). Professional societies making engineering outreach work: Good input results in good output. *Leadership and Management in Engineering*, 11-26.

Bona, L., Kelly, A., & Jung, M. (2010). Exploring factors contributing to women's nontraditional career aspirations. *Psi Chi Journal of Undergraduate Research, 15*(3), 123-129.

Brand, B., & Kasarda, M. (2014). The Influence of social Interactions on female students' retention in two engineering programs. *Insights on Learning Disabilities, 11*(1), 33-45.

Bronson, P., & Merryman, A. (2009). *Nurtureshock: New thinking about children.* New York: Twelve.

Brown, M., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230-258.

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

Buck, G. A., Clark, V., Leslie-Pelecky, D., Lu, Y., & Cerda-Lizarraga, P. (2008). Examining the cognitive processes used by adolescent girls and women scientists in identifying Science Role Models: A Feminist Approach. , 92(4), 688-707. doi:. *Science Education, 92*(4). doi:10.1002/sce.20257

Buday, S., Stake, J., & Peterson, Z. (2012). Gender and the choice of a science career: The impact of social support and possible selves. *Sex Roles, 66*, 197-209.

Bull, R., Davidson, W. A., & Nordmann, E. (2010). Prenatal testosterone, visual-spatial memory, and numerical skills in young children. *Learning and Individual Differences, 20*, 246–250.

Byrne, B. (2008). Testing for time-invariant and time-varying predictors of self-perceived ability in math, language arts, and science: A look at the gender factor. In H. M. Watt, & J. S.

Eccles (Eds.), *Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences.* (pp. 145-169). Washington, DC: American Psychological Association.

Bystydzienski, J. M., & Brown, A. (2012, Winter). “I just want to help people”: Young women’s gendered engagement with engineering. *Feminist Formations, 24*(3), pp. 1-21. doi:10.1353/ff.2012.0027

Campbell, B., Schellenberg, E. G., & Senn, C. Y. (1997). Evaluating measures of contemporary sexism. *Psychology of Women Quarterly, 21*, 89-102.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-mutlimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

Campbell, S., & Collaer, M. (2009). Stereotype threat and gender differences in performance on a novel visuospatial task. *Psychology of Women Quarterly, 33*, 437-444.

Capobianco, B. M. (2006). Undergraduate women engineering their professional identities. *Jounral of Women and Minorities in Science and Engineering, 12*(2), 1-24.

Capobianco, B. M., French, B. F., & Diefes-Dux, H. A. (2012). Engineering identity development among pre-adolescent learners. *Journal of Engineering Education*, 698-716.

Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences, 3*(3), 106-116.

Casey, M. (1996). Understanding individual differences in spatial ability within females: A nature/nurture interactionist framework. *DEVELOPMENTAL REVIEW, 16*, 241-260.

Castro, S. L., Scandura, T. A., & Williams, E. A. (2004). Validity of Scandura and Ragins' (1993) multidimensional mentoring measure: An evaluation and refinement. *Management Faculty Articles and Papers* (pp. 1-39). San Antonio, TX: University of Miami Scholarly Repository. Retrieved from Unversity of Miami Scholarly Repository:

http://scholarlyrepository.miami.edu/cgi/viewcontent.cgi?article=1006&context=management_articles

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest, 15*(3), 75-141. doi:10.1177/1529100614541236

Ceci, S. J., Williams, W. M., Sumner, R. A., & DeFraine, W. C. (2012). Do subtle cues about belongingness constrain women's career choices? *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory, 22*(4), 255-258. doi:10.1080/1047840X.2011.619112

Ceci, S., Williams, W., & Barnett, S. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin, 135*(2), 218-261.

Cheryan, S. (2012). Understanding the paradox in math-related fields: Why some gender gaps remain while others do not. *Sex Roles, 66*, 184–190. doi:10.1007/s11199-011-0060-z

Cheryan, S., Davies, P. G., Plaut, V. C., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology, 97*(6), 1045–1060.

Cheryan, S., Plaut, V. C., Handron, C., & Hudson, L. (2013). The stereotypical computer scientist: Gendered media representations as a barrier to inclusion for women. *Sex Roles, 69*, 58–71. doi:10.1007/s11199-013-0296-x

Cheryan, S., Siy, J. O., Vichayapai, M., Drury, B. J., & Kim, S. (2011). Do female and male role models who embody STEM stereotypes hinder women's anticipated success in STEM? *Social Psychological and Personality Science*, 1-9.

Chesler, N. C., & Chesler, M. A. (2002). Gender-informed mentoring strategies for women engineering scholars: On establishing a caring community. *Journal of Engineering Education*, 49-55.

Childress, V. (2003, Summer). *Girls in Technology, A Supplement to Exploring Technology and Fuindamentals of Technology Curriculum Guides.*

Cohen, J. (1988). *Statistical power analysis of the behavioral sciences.* New York: Academic Press.

Colman, A. M., Norris, C. E., & Preston, C. C. (1997). Comparing rating scales of different lengths: Equivalence of scores from 5-point and 7-point scales. *Psychological Reports, 80*, 355-362.

Committee on Maximizing the Potential of Women in Academic Science and Engineering. (2007). *Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering.* Washington, DC: National Academy of Engineering Press. Retrieved from http://www.nap.edu/catalog/11741.html

Constantinople, A. (1973). Masculinity-femininity: An exception to a famous dictum? *Psychological Bulletin, 80*(5), 389-407.

Corbett, C., & Hill, C. (2012, October 24). *Graduating to a pay gap.* Retrieved from American Association of University Women: http://www.aauw.org/graduatetoapaygap/

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10*(7). Retrieved from http://pareonline.net/getvn.asp?v=10&n=7

Crisp, G. (2009). Conceptualization and initial validation of the college student mentoring scale (CSMS). *Journal of College Student Development, 50*(2), 177-194. doi:DOI: 10.1353/csd.0.0061

Crisp, G., & Cruz, I. (2010). Confirmatory factor analysis of a measure of "mentoring" among undergraduate students attending a Hispanic serving institution. *Journal of Hispanic Higher Education, 9*, 232-244. doi:10.1177/1538192710371982

Crocker, J. (2011). Safety in numbers: Shifting from egosystem to ecosystem. *Psychological Inquiry, 22*, 259-264.

Dasgupta, N. (2011). Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory, 22*(4), 231-246. doi:10.1080/ 1047840X.2011.607313

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology, 40*, 642-658.

Decoster, J. (2005). Scale construction notes. Tuscaloosa, AL. Retrieved from http://www.stat-help.com/notes.html

Derks, B., Ellemers, N., Laar, C. v., & Groot, K. d. (2011). Do sexist organizational cultures create the queen bee? *British Journal of Social Psychology, 50*, pp. 519–535. doi:10.1348/014466610X525280

Devellis, R. (2012). *Scale development: Theory and application.* Los Angeles, CA: Sage Publications.

Dimitrov, D. M. (2012). *Statistical methods for validation of sssessment scale data in counseling and related fields* . Alexandria VA: American Counseling Association.

Doms, M. (2011, August 3). *STEM: Where are the women?* Retrieved from Economics and Statistics Administration: http://www.esa.doc.gov/Blog/2011/08/03/stem-where-are-women

Draper, D. (2010). *The instructional effects of knowledge based communities of practice learning environments on student achievement and knowledge convergence.* Retrieved from The Penn State University: https://etda.libraries.psu.edu/paper/10629/6692

Dreher, G. F., & Ash, R. A. (1990). A comparative study of mentoring among men and women in managerial, professional, and technical positions. *Journal of Applied Psychology, 75*(5), 539-546.

Drury, B. J., Siy, J. O., & Cheryan, S. (2011). When do female role models benefit women? The importance of differentiating recruitment From retention in STEM. *Psychological Inquiry*, 265–269.

Dunleavy, V. O., & Millette, D. M. (2007). Measuring mentor roles and protégé initiation strategies: Validating instruments in academic mentoring . (pp. 1-25). Chicago: National Communication Association, Instructional Development Division.

Dweck, C. S. (2006). Is math a gift? Beliefs that put females at risk. In S. J. Ceci, & W. M. Williams (Eds.), *Why aren't more women in science?: Top researchers debate the evidence* (pp. 47-57). Washington DC: American Psychological Association.

Dweck, C. S. (2008). *Mindsets and math/science achievement.* Prepared for Carnegie-IAS Commission on Mathematics and Science Education, New York, NY.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256-273.

Dworkin, T. M., Maurer, V., & Schipani, C. A. (2012). Career mentoring for women: New horizons/Expanded methods. *Business Horizons, 55*, pp. 363—372. doi:10.1016/ j.bushor.2012.03.001

Eagly, A. H. (1987). *Sex-differences in social behaviors: A social-role interpretation.* Hillsdale, NJ: Lawrence Erlbaum.

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology, 46*, 735-754.

Eagly, A. H., & Wood, W. (1999, June). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist, 54*(6), pp. 408-423.

Eagly, A., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. E. Trautner (Ed.), *The developmental social psychology of gender* (pp. 123-174).

Eby, L. T., Allen, T. D., Evans, S. C., Ng, T., & DuBois, D. L. (2008). Does mentoring matter? A multidisciplinary meta-analysis comparing mentored and non-mentored individuals. *Journal of Vocational Behavior, 72*, 254-267. doi:10.1016/j.jvb.2007.04.005

Eccles, J. (1994). Understanding women's educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *Psychology of Women Quarterly, 18* , pp. 585-609.

Eccles, J. (2006). Where are all the women? Gender differences in participation in physical science and engineering. In S. Ceci, W. M. Williams, S. Ceci, & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 199-210). Washington, DC: American Psychological Association.

Eccles, J. (2011). Understanding women's achievement choices: Looking back and looking forward. *Psychology of Women Quarterly, 35*, pp. 510-516. doi:10.1177/ 0361684311414829

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly, 11*, 135-172.

Eckes, T., & Trautner, H. M. (Eds.). (2000). *The developmental social psychology of gender.* New York: Psychology Press - Taylor & Francis Group.

Ecuyer-Dab, I., & Robert, M. (2004). Have sex differences in spatial ability evolved from male competition for mating and female concern for survival? *Cognition, 91*, 251-257.

Eisenhart, M., Bystydzienski, J., & Bruning, M. (2010). *Female recruits explore engineering (FREE) project.*

Evans, O., & Steptoe, A. (2002). The contribution of gender-role orientation, work factors and home stressors to psychological well-being and sickness absence in male- and female-dominated occupational groups. *Social Science & Medicine, 54*(4), 481–492.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299.

Faludi, S. (1991). *Backlash: The undeclared war against American women* . New York, NY: Crown Publishing Group.

Farrell, A. M., & Rudd, J. M. (2009). Factor analysis and discriminant validity: A brief review of some practical issues. *ANZMAC 2009 conference proceedings*, (pp. 1-9). Birmingham, UK.

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science, 18*(10), 850-855.

Fernández-Ballesteros, R., Díez-Nicolás, J., Caprara, G. V., Barbaranelli, C., & Bandura, A. (2002). Determinants and structural relation of personal efficacy to collective efficacy. *Applied Psychology: An International Review, 51*(1), 107-125.

Fetterolf, J. C., & Eagly, A. H. (2011). Do young women expect gender equality in their future lives? An answer from a possible selves experiment. *Sex Roles, 65*, 83-93. doi:10.1007/ s11199-011-9981-9

Fine, C. (2010). *Delusions of gender: How our minds, society, and neurosexism create difference.* New York, NY: W.W. Norton and Company.

Fleming, M., House, S., Hanson, V. S., Yu, L., Garbutt, J., McGee, R., . . . Rubio, D. M. (2013). The mentoring competency assessment: Validation of a new instrument to evaluate skills of research mentors. *Academic Medicine, 88*(7), 1002-1008. doi:10.1097/ACM. 0b013e318295e298

Fletcher, J. K. (1999). *Disappearing acts: Gender, power, and relational practices at work.* Cambridge, MA: Massachusetts Institute of Technology.

Flouri, E., & Hawkes, D. (2008). Ambitious mothers – successful daughters: Mothers' early expectations for children's education and children's earnings and sense of control in adult life. *British Journal of Educational Psychology*, 411-433.

Foels, R., & Pappas, C. J. (2004, June). Learning and unlearning the myths we are taught: Gender and social dominance orientation. *Sex Roles, 50*(11/12).

Fowler, J. L., & O'Gorman, J. G. (2005). Mentoring functions: A contemporary view of the perceptions of mentees and mentors. *British Journal of Management, 16*, 51–57 . doi:10.1111/j.1467-8551.2005.00439.x

Fried, T., & MacCleave, A. (2009). Influence of role models and mentors on female graduate students' choice of science as a career. *The Alberta Journal of Educational Research, 55*(4), 482-496.

Frome, P. M., Alfeld, C. J., Eccles, J. S., & Barber, B. L. (2006). Why don't they want a male-dominated job? An investigation of young women who changed their occupational aspirations. *Educational Research and Evaluation, 12* (4), 359 – 372.

Gaskin, J. (2012). *Convergent validity*. Retrieved from Gaskination's StatWiki: http://statwiki. kolobkreations.com

Gaskin, J. (2012). *Validity master.* Retrieved from Stats Tools Package: http://statwiki. kolobkreations.com

Geber, H., & Roughneen, C. (2011). Tinkering, tailoring, and transforming: Retention of scientific excellence of women researchers through WiSER mentoring. *International Journal of Evidence Based Coaching and Mentoring, 9*(1), 59-75. Retrieved from http://www.business.brookes.ac.uk/research/areas/coachingandmentoring/

Gibson, D. E. (2003). Developing the professional self-concept: Role model construals in early, middle, and late career stages. *Organization Science, 14*(5), pp. 591-610.

Girl Scout Research Institute. (2012). *Generation STEM: What girls say about science, technology, engineering, and math.* New York, NY: Girl Scout Research Institute.

Gitta H. Lubke ;Bengt O. Muthén (2004) Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons, , 11:4, 514-534, 10.1207/. (n.d.). *Structural Equation Modeling: A Multidisciplinary Journal*.

Glick, P., & Fiske, S. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*(3), 491-512.

Glick, P., & Fiske, S. (2001, February). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist, 56*, pp. 109-118. doi:10.1037//O003-066X.56.2.1O9

Glick, P., & Fiske, S. T. (2011). Ambivalent sexism revisited. *Psychology of Women Quarterly, 35*, 530-535. doi:10.1177/0361684311414832

Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. *2003 Midwest Research to Practice Conference in Adult, Continuing, and Community Education* (pp. 82-88). The Ohio State University.

Goldin, C., & Rouse, C. (2000, September). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *The American Economic Review, 90*(4), pp. 715-741. Retrieved from http://www.jstor.org/stable/117305

Goleman, D. (1984, June 12). A bias puts self at center of everything. *The New York Times*. Retrieved from http://www.nytimes.com/1984/06/12/science/a-bias-puts-self-at-center-of-everything.html

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology, 102*(4), 700–717. doi:10.1037/a0026659

Gottfredson, L. (1981). Circumscription and compromise: A developmental theory of occupational aspirations. *Journal of Counseling Psychology Monograph, 28*(6), 545=579.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4-27.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6*(4), 430-450. doi:10.1037//1082-989X.6.4.430

Gurung, R. A. (2009). Sex-role orientation. In H. T. Reis, & S. Sprecher (Eds.), *Encyclopedia of Human Relationships* (pp. 1450-1453). Thousand Oaks CA: SAGE Publications, Inc. doi:http://dx.doi.org/10.4135/9781412958479.n475

Hackett, G., & Betz, N. E. (1981). A self-efficacy approach to the career development of women. *Journal of Vocational Behavior, 18*, 326-339.

Halim, M. L., Ruble, D. N., & Amodio, D. M. (2011). From pink frilly dresses to 'one of the boys': A social-cognitive analysis of gender identity development and gender bias. *Social and Personality Psychology Compass*, 933–949. doi:10.1111/j.1751-9004.2011.00399.x

Heilbronner, N. (2009). Nurturing talented girls in the science classroom. *Gifted Child Today, 32*(1), 46-54.

Hill, C., Corbett, C., & St. Rose, A. (2010). *Why so few? Women in science, technology, engineering and mathematics.* Washington, DC: American Association of University Women.

Holmes, S., Redmond, A., & Thomas, J. (2012). Girls helping girls: Assessing the influence of college student mentors in an afterschool engineering program. *Mentoring & Tutoring: Partnership in Learning, 20*(1), 137-150.

Holth, L., & Mellström, U. (2011). Revisiting engineering, masculinity and technology studies: Old structures with new openings. *International Journal of Gender, Science, and Technology, 3*(2), 313-329. Retrieved from http://genderandset.open.ac.uk

Hudson, P., Skamp, K., & Brooks, L. (2005). Development of an instrument: Mentoring for effective primary science teaching. In M. Windschitl (Ed.), *Science Teacher Education* (pp. 657-674). Wiley Periodicals, Inc.

Hug, S., Jurow, A. S., & Chi, W. C. (2011). Evolving identities: Undergraduate women pursuing the engineering professoriate. *ASEE 2011 Conference Proceedings* (pp. 1-15). American Society for Engineering Education.

Huguet, P., & Régner, I. (2009). Counter-stereotypic beliefs in math do not protect school girls from stereotype threat. *Journal of Experimental Social Psychology, 45*, 1024-1027.

Jackman, M. R. (1994). *The velvet glove: Paternalism and conflict in gender, class, and race relations.* Berkeley: University of California. Retrieved from http://publishing.cdlib.org/ucpressebooks/view?docId=ft958009k3&chunk.id=d0e1322&toc.depth=1&toc.id=0&brand=ucpress&query=velvet+glove

Jagacinski, C. (1987). Androgyny in a male-dominated field: The relationship of sex-typed traits to performance and satisfaction in engineering. *Sex Roles, 17*(9-10), 529-547.

Jagacinski, C. M. (1987, March). Engineering careers: Women in a male dominated field. *Psychology of Women Quarterly, 11*(1), pp. 97-111.

Jöreskog, K. G. (1999). How large can a standarized coefficient be? Chicago: Scientific Software International. Retrieved from www.ssicentral.com/lisrel/column2.htm

Joseph, R. (2000). The evolution of sex differences in language, sexuality, and visual-spatial skills. *Archives of Sexual Behavior, 29*(1), 35-66.

Jost, J. T., & Banaji, M. (1994). The role of stereotyping in system justification and the production of false consciousness. *British Journal of Social Psychology, 22*, 1-27.

Jost, J. T., & Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: Consequences for specific and diffuse forms of system justification. *Journal of Personality and Social Psychology, 88*(3), 498-509. doi:10.1037/0022-3514.88.3.498

Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., & Hardin, C. D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior, 29*, 39–69.

Kammeyer-Mueller, J. D., & Judge, T. A. (2008). A quantitative review of mentoring research: Test of a model. *Journal of Vocational Behavior, 72*, 269–283. doi:10.1016/ j.jvb.2007.09.006

Kekelis, L. S., Ancheta, R. W., & Countryman, J. (2005, Summer). Role models make a diifference: a recipe for success. *AWIS Magazine*, pp. 17-24.

Kempel, P., Gohlke, B., Klempau, J., Zinsberger, P., Reuter, M., & Hennig, J. (2005). Second-to-fourth digit length, testosterone and spatial ability. *Intelligence, 33*, 215–230.

Kiecolt, K. J., & Nathan, L. E. (1985). *Secondary analysis of survey data.* Newbury Park, CA: SAGE Publications, Inc.

Kilianski, S. E., & Rudman, L. A. (1998). Wanting it both ways: Do women approve of benevolent sexism? *Sex Roles, 39*(5/6), 333-352.

Kimmel, L. G., Miller, J. D., & Eccles, J. (2012). Do the paths to STEMM differ by gender? *Peabody Jounral of Education, 87*(1), 92-113. doi:10.1080/0161956X.2012.642276

Kohlberg, L. (1973). Stage and sequence: The cognitive development approach to socialization. In D. Goslin, *Handbook of socialization theory and research* (pp. 347-480). Chicago, IL: Rand McNally College Publishing Company.

Kram, K. (1985). *Mentoring at work.* Boston, MA: Scott, Foresman.

Kram, K., & Ragins, B. R. (2008). *The handbook of mentoring at work: Theory, research, and practice.* Thousand Oaks, CA: SAGE Publications, Inc.

Lane, K. A., Goh, J. X., & Driver-Linn, E. (2011). Implicit science stereotypes mediate the relationship between gender and academic participation. *Sex Roles, 66*(3-4), 1-15. doi:10.1007/s11199-011-0036-z

Larose, S., Ratelle, C. F., Guay, F., Senécal, C., Harvey, M., & Drouin, E. (2008). A sociomotivational analysis of gender effects on persistence in science and technology: A 5-year longitudinal study. In H. M. Watt., & J. S. Eccles (Eds.), *Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences* (pp. 171-192). Washington, DC: American Psychological Association. doi:10.1037/11706-006

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge, UK: Cambridge University Press.

Leaper, C., Farkas, T., & Spears Brown, C. (2012). Adolescent girls' experiences and gender-related beliefs in relation to their motivation in math/science and english. *Journal of Youth Adolescence, 41,* 268–282. doi:10.1007/s10964-011-9693-z

Lee, T. L., Fiske, S. T., & Glick, P. (2010). Next gen ambivalent sexism: Converging correlates, causality in context, and converse causality, an introduction to the special issue. *Sex Roles, 62,* 395-404. doi:10.1007/s11199-010-9747-9

Leedy, P. D., & Ormrod, J. E. (2012). *Practical research: Planning and design.* Upper Saddle River NJ: Pearson Education, Inc.

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 Development, 45,* 79-122.

Li, C., & Kerpelman, J. (2007, January). Parental influences on young women's certainty about their career aspirations. *Sex Roles,* pp. 105-115.

Likert, R. (1932). A technique for the measurement of attitudes. (R. Woodworth, Ed.) *Archives of Psychology, 22*(140), pp. 1-55.

Little, A. J., & León de la Barra, B. A. (2009). Attracting girls to science, engineering and technology: an Australian perspective. *European Journal of Engineering Education,* 439-445.

Logel, C., Walton, G. M., Spencer, S. J., Iserman, E. C., & von Hippel, W. (2009). Interacting with sexist men triggers social identity threat among female engineers. *Journal of Personality and Social Psychology, 96*(6), 1089–1103. doi:10.1037/a0015703

London, B. R. (2011). The influences of perceived identity compatibility and social support on women in nontraditional fields during the college transition. *Basic & Applied Social Psychology, 33*(4), 304-321. doi:10.1080/01973533.2011.614166

London, B., Rosenthal, L., & Gonzalez, A. (2011). Assessing the role of gender rejection sensitivity, identity, and support on the academic engagement of women in nontraditional fields using experience sampling methods. *Journal Of Social Issues, 67*(3), 510-530. doi:10.1111/j.1540-4560.2011.01712.x

Lord, S. M., Camacho, M. M., Layton, R. A., Long, R. A., Ohland, M. W., & Wasburn, M. H. (2009). Who's persisting in engineering? A comparative analysis of female and male asian, black, hispanic, native american, and white students. *Journal of Women and Minorities in Science and Engineering*, 167–190.

Lubke, G. H., & Muthén, B. O. (2004). Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modeling, 11*(4), 514–534. doi:10.1207/s15328007sem1104_2

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84-99.

Maccoby, E. E., & Jacklin, C. N. (1974). Chapter 10: Summary and commentary. In E. E. Maccoby, & C. N. Jacklin, *The psychology of sex differences* (pp. 349-374). Stanford, CA: Stanford University Press.

Marikar, S. (2014, December 24). GoldieBlox tries to prove itself — This time without the Beastie Boys. *New York Times*. Retrieved from http://www.nytimes.com/2014/12/25/fashion/goldieblox-tries-to-prove-itself-this-time-without-the-beastie-boys.html?ref=fashion&_r=2

McCarthy, R. (2009, October). Beyond Smash and Crash:. *The Technology Teacher*, pp. 16-21.

McIntosh, P. (1988). *White privilege and male privilege: A personal account of coming to see correspondences through work in women's studies.* Wellsley: Wellesley College Center for Research on Women.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415-444. Retrieved from http://www.jstor.org/stable/2678628?origin=JSTOR-pdf

Meadows, L. A., & Sekaquaptewa, D. (2013). The influence of gender stereotypes on role adoption in student teams. *120th ASEE Annual Conference & Exposition.* Atlanta: American Society for Engineering Education.

Medeiros, D. J. (2011). The influence of female social models in corporate STEM initiatives on girls' math and science attitudes. *Dissertation*. Ann Arbor, MI: ProQuest LLC.

Merriam-Webster. (2014). *Mentor.* Retrieved from Merriam-Webster Online Dictionary: http://www.merriam-webster.com/dictionary/mentoring

Merriam-Webster. (2014). *Psychosocial.* Retrieved from Merriam-Webster Online Dictionary: http://www.merriam-webster.com/dictionary/psychosocial

Mertz, N. (2004). What's a mentor anyway? *Educational Administration Quarterly*, 541-560. doi:10.1177/0013161X04267110

Messick, S. (1989). Validity. In R. L. Linn, & R. Linn (Ed.), *Educational measurement* (pp. 13–103). New York, NY: Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons′ responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741– 749.

Milgram, D. (2011, November). How to recruit women and girls to the science, technology, engineering, and math (STEM) classrooms. *Technology and Engineering Teacher*. (ITEEA, Ed.) Reston, VA.

Mills, J., Ayre, M., & Gill, J. (2010). *Gender inclusive engineering education.* New York: Taylor & Francis Group.

Moe, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences, 19*, 21-27.

Moe, A., & Pazzaglia, F. (2009). Beyond genetics in mental rotation test performance - the power of effort attribution. *Learning and Individual Differences, 20*, 464-468.

Moss-Racusin, C. A., & Rudman, L. A. (2010). Disruptions in women's self-promotion: The backlash avoidance model. *Psychology of Women Quarterly, 34*, pp. 186–202.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *PNAS, 109*(41), 16474–16479. doi:10.1073/pnas.1211286109

Mussen, P. (1973). Early sex-role development. In D. Goslin, *Handbook of socialization and research* (pp. 707-732). Chicago, IL: Rand McNally College Publishing Company.

National Academy of Engineering. (2008). *Changing the conversation: Messages for improving public understanding of engineering.* Washington DC: National Academies Press.

Nauta, M. M., & Kokaly, M. L. (2001). Assessing role model influences on students' academic and vocational decisions. *Journal of Career Assessment, 9*, 81-99. doi:10.1177/106907270100900106

Neubauer, A., Bergner, S., & Schatz, M. (2010). Two- vs. three-dimensional presentation of mental rotation tasks: Sex differences and effects of training on performance and brain activation. *Intelligence, 38*, 529-539.

Neuman, W. (2003). *Social research methods: Qualitative and quantitative approaches* (5 ed.). Boston: MA: Pearson:.

Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal, 48*(5), 1125–1156. doi:10.3102/0002831211410683

Opare, P. B. (2012, December). Dissertation: Factors that female higher education faculty in select science, technology, engineering, and mathematics (STEM) fields perceive as being influential to their success and persistence in their chosen professions. Norfolk, VA. Retrieved from ProQuest.

Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research & Evaluation, 9*(11), 1-12. Retrieved from http://PAREonline.net/getvn.asp?v=9&n=11

Park, B., Smith, J. A., & Correll, J. (2010). The persistence of implicit behavioral associations for moms and dads. *Journal of Experimental Social Psychology, 46*, 809–815.

Parks-Stamm, E. J., Heilman, M. E., & Hearns, K. A. (2008). Motivated to penalize: Women's strategic rejection of successful women. *Personality and Social Psychology Bulletin, 34*, 237–247. doi:10.1177/0146167207310027

Patterson, E. A., Campbell, P. B., Busch-Vishniac, I., & Guillaume, D. W. (2011). The effect of context on student engagement in engineering. *European Journal of Engineering Education, 36*(3), 211-224. doi:10.1080/03043797.2011.575218

Paulsen, C. A. (2014). *Invent It. Build It. 2013: evaluation report.* Concord, MA: Concord Evaluation Group, LLC.

Perry, J. L., Nicholls, A. R., Clough, P. J., & Crust, L. (2015). Assessing model fit: Caveats and recommendations for confirmatory factor analysis and exploratory structural equation modeling. *Measurement in Physical Education & Exercise Science, 19*, 12–21. doi:10.1080/1091367X.2014.952370

Phelps, M. (2012). Effects of hands-on activities on middle school females' spatial skills and interest in technology-based careers. *ASEE 2012 Conference Proceedings* (pp. AC2012-3268). American Society for Engineering Education. Retrieved from http://www.asee.org/public/conferences/8/papers/3268/view

Plante, I., de la Sablonnière, R., Aronson, J. M., & Théorêt, M. (2013). Gender stereotype endorsement and achievement-related outcomes: The role of competence beliefs and task values. *Contemporary Educational Psychology, 38*, 225-235.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 539-569. doi:10.1146/annurev-psych-120710-100452

Pratto, F., Sidanius, J., & Levin, S. (2006). Social dominance theory and the dynamics of intergroup relations: Taking stock and looking forward. *European Review of Social Psychology, 17*, 271-320.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology, 67*(4), 741-763.

Pratto, F., Stallworth, L. M., Sidanius, J., & Siers, B. (1997). The gender gap in occupational role attainment: A social dominance approach. *Journal of Personality and Social Psychology, 72*(1), 37-53.

Pronin, E., Steele, C. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: Women and mathematics. *Journal of Experimental Social Psychology, 40*, 152–168.

Quimby, J. L., & DeSantis, A. M. (2006, June). The influence of role models on women's career choices. *The Career Development Quarterly, 54*, pp. 297-306.

Ragins, B. R., & Winkel, D. E. (2011). Gender, emotion and power in work relationships. *Human Resource Management Review, 21*, 377–393. doi:10.1016/j.hrmr.2011.05.001

Rudman, L. A., & Phelan, J. E. (2010). The effect of priming gender roles on women's implicit gender beliefs and career aspirations. *Social Psychology, 41*(3), 192–202. doi:10.1027/1864-9335/a000027

Scandura, T. A., & Ragins, B. R. (1993). The effects of sex and gender role orientation on mentorship in male-dominated occupations. *Journal of Vocational Behavior, 43*, 251-265.

Schlenker, B. R. (1978). Attitudes as actions: Social identity theory and consumer research. *Advances in Consumer Research, 5*(1), 352-359.

Schmitt, M. T., & Wirth, J. H. (2009). Evidence that gender differences in social dominance orientation result from gendered self-stereotyping and group-interested responses to patriarchy. *Psychology of Women Quarterly, 33*, 429–436.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. The Journal of Educational Research, 99(6), 323-338.

Schumann, K., & Ross, M. (2010). Why women apologize more than men: Gender differences in thresholds for perceiving offensive behavior. Psychological Science, 21(11), 1649-1655. doi:10.1177/0956797610384150 Schunk, D. (1981). Modeling and attributional feedback effects on children's achievement: A self-efficacy analysis. *Journal of Educational Psychology, 74*, 93–105.

Sekaquaptewa, D. (2011). Discounting their own success: A case for the role of implicit stereotypic attribution bias in women's STEM outcomes. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory, 22* (4), 291-295. doi:10.1080/1047840X.2011.624979

Seymour, E. (1997). The role of socialization in shaping the career-related choices of undergraduate women in science, mathematics, and engineering majors. *Annals New York Academy of Sciences* (pp. 118-126). New York Academy of Sciences.

Shapiro, J., & Williams, A. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles, 66*(3/4), 175-183.

Shields, S. A., & Dicicco, E. C. (2011). The social psychology of sex and gender: From gender differences to doing gender. *Psychology of Women Quarterly*, 35(3) 491-499. doi:10.1177/0361684311414823

Sidanius, J., & Pratto, F. (1999). Social dominance theory: A new synthesis. In J. Sidanius, & F. Pratto, *Social dominance: An integoup theory of social hierarchy and oppression* (pp. 31-58). Cambridge, UK: Cambridge University Press.

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression.* New York: Cambridge University Press.

Sidanius, J., Devereux, E., & Pratto, F. (2001). A comparison of symbolic racism theory and social dominance theory as explanations for racial policy attitudes. *The Journal of Social Psychology, 132*(3), 377-395.

Sidanius, J., Pratto, F., & Bobo, L. (1994). Social dominance orientation and the political psychology of gender: A case of invariance? *Journal of Personality and Social Psychology, 67*(6), 998-1011.

Silverman, I., Choi, J., & Peters, M. (1994). The hunter-gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in recall of object arrays. *Ethology and Sociobiology, 15*(2), 95-105.

Singh, R., Fouad, N. A., Fitzpatrick, M. E., Liu, J. P., Cappaert, K. J., & Figuereido, C. (2013). Stemming the tide: Predicting women engineers' intentions to leave. *Journal of Vocational Behavior, 83*, 281–294. doi:10.1016/j.jvb.2013.05.007

Smith, J. L., Lewis, K. L., Hawthorne, L., & Hodges, S. D. (2013). When trying hard Isn't natural: Women's belonging with and motivation for male-dominated STEM fields as a function of effort expenditure concerns. *Personality and Social Psychology Bulletin, 39*, 131-143. doi:10.1177/0146167212468332

Society of Women Engineers. (2012, 9 26). How to plan and host Wow! Chicago, IL.

Sorby, S. A. (2009). Educational research in developing 3-D spatial skills for engineering students. *International Journal of Science Education, 31*(3), 459-480.

Spence, J. (1984). Gender identity and its implications for the concepts of masculinity and femininity. *Nebraska Symposium on Motivation* (pp. 61-92). Lincoln, NE: University of Nebraska.

Spence, J. T., & Buckner, C. E. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes: What do they signify? *Psychology of Women Quarterly, 24*, p. Psychology of Women Quarterly.

Spence, J., & Helmreich, R. (1978). *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents.* Austin, TX: University of Texas Press.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4–28. doi:10.1006/jesp.1998.1373

Springer, K. W., & Hauser, R. M. (2006). An assessment of the construct validity of Ryff's scales of psychological well-being: method, mode and measurement effects. *Social Science Research, 35*(4), 1080–1102.

Staines, G., Tavris, C., & Jayaratne, T. E. (1974). The queen bee syndrome. *Psychology Today, 7*(8), 51-58.

Stake, J. E., Zand, D., & Smalley, R. (1996). The relation of instrumentality and expressiveness to self-concept and adjustment: A social context perspective. *Journal of Social and Clinical Psychology, 15*(2), 167-190.

Stallard, K., Ehrenreich, B., & Sklar, H. (1983). Poverty in the American dream: Women and children first. Boston, MA: South End Press.

Steele, C. (2010). *Whistling Vivaldi: And other clues to how stereotypes affect us.* New York: W.W. Norton & Company.

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, 255-270.

SWE Corporate Partnership Council. (2014, October). *SWE: Magazine of the Society of Women Engineers*, 124. Chicago, IL, USA: Society of Women Engineers.

Swim, Aikin, Hall, & Hunter. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology, 68*, 199-214.

Swim, J. K., & Cohen, L. L. (1997). Overt, covert, and subtle sexism: A comparison between the attitudes toward women and modern sexism scales. *Psychology of Women Quarterly, 21*, 103-118.

The GE Foundation. (2002). *Upping the numbers: Using research-based decision making to increase diversity in the quantitative disciplines.* Newton, MA: Education Development Center.

Thoman, D. B., Arizaga, J. A., Smith, J. L., Story, T. S., & Soncuya, G. (2013). The grass Is greener in non-science, technology, engineering, and math classes: Examining the role of competing belonging to undergraduate women's vulnerability to being pulled away from science. *Psychology of Women Quarterly, 0*(0), 1-13. doi:10.1177/0361684313499899

Tougas, F., Brown, R., Beaton, A. M., & Joly, S. (1995). Neosexism: Plus ça change, plus c'est pareil. *Personality and Social Psychology Bulletin, 21*, 842-849. doi:10.1177/0146167295218007

Trautner, H. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., & Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in childhood: Developmental or differential? *Infant and Child Development, 14*, 365–381. doi:10.1002/icd.399

Tse, C., Logel, C., & Spencer, S. J. (2011). Stereotype inoculation: An elegant and practical model. *Psychological Inquiry, 22*, 296–298. doi:10.1080/1047840X.2011.623266

Vescio, T. K., Gervais, S. J., Snyder, M., & Hoover, A. (2005). Power and the creation of patronizing environments: The stereotype-based behaviors of the powerful and their effects on female performance in masculine domains. *Journal of Personality and Social Psychology, 88*(4), 658-672. doi:10.1037/0022-3514.88.4.658

Vest, S. N., & Niccolai, M. J. (2003). Increasing the number of women in the technological sciences through role model intervention. *International Association for Computer Information Systems* (pp. 734-740). Las Vegas NV: International Association for Computer Information Systems.

Virtanen, S., & Ikonen, P. (2009). Best practice: "Trick-track" to encourage students, particularly girls, to learn technology in an inventive manner. *ECER 2009.* Vienna, Austria: University of Vienna.

Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30 year examination. *Intelligence, 38*, 412-423.

Wang, M.-T., Eccles, J. S., & Kenny, S. (2013). Not lack of ability but more choice : Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science, 24*, 770-775. doi:10.1177/0956797612458937

Watt, H. M. (2008). What motivates females and males to pursue sex-stereotyped careers? In J. S. Eccles, & H. Watt (Eds.), *Gender and occupational outcomes. Longitudinal assessments of individual, social, and cultural influences* (pp. 87-113). Washington DC: American Psychological Association.

Weber, K. (2011). Role models and informal STEM-related activities positively impact female interest in STEM. *Technology and Engineering Teacher, 71*(3), pp. 18-21.

Weber, K. (2012). Gender differences in interest, perceived personal capacity, and participation in STEM-related activities. *Journal of Technology Education, 24*(1), 18-33.

Weichselbaumer, D. (2004). Is it sex or personality? The impact of sex stereotypes on discrimination in applicant selection. *Eastern Economic Journal, 30*(2), 159–186.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010a). The effect of rating scale format on response styles: The number of response categories and response category labels. *Intern. J. of Research in Marketing, 27*(3), 236-247.

Weijters, B., Geuens, M., & Schillewaert, N. (2010b). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement. 34:105–21, 34*, 105-21. doi:10.1177/0146621609338593

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity.* Cambridge, UK: Cambridge University Press.

Wenger, E. (2002). *Cultivating communities of practice.* Boston, MA: Harvard Business School Publishing.

White, M. J., & White, G. B. (2006). Implicit and explicit occupational gender stereotypes. *Sex Roles, 55*, :259–266. doi:10.1007/s11199-006-9078-z

Wikipedia. (2014). *Latent variable*. Retrieved from Wikipedia, the Free Encyclopedia: http://en.wikipedia.org/wiki/Latent_variable

Williams, W., & Ceci, S. (2012). When scientists choose motherhood. *American Scientist, 100*(2). doi:10.1511/2012.95.138

Wong, P. T., Kettlewell, G., & Sproule, C. F. (1985). On the importance of being masculine: Sex role, attribution, and women's career achievement. *Sex Roles, 12*(7/8), 757-769.

Wood, A. M. (2011). Designing a new scale/questionnaire: Optimal psychometric practice. Manchester, UK.

Woodcock, A. (2012). Implicit ingroup biases and performance under threat: Do successful counterstereotypic exemplars help or harm? *Dissertation*. UMI Dissertation Publishing.

Yoon, S. Y., Evans, M. G., & Strobel, J. (2014). Validation of the teaching engineering self-efficacy scale for K-12 teachers: A structural equation modeling approach. *Journal of Engineering Education, 103*(3), 463-485. doi:10.1002/jee.20049

Young, D. M., Rudman, L. A., Buettner, H. M., & McLean, M. C. (2013). The influence of female role models on women's implicit science cognitions. *Psychology of Women Quarterly*, *37*(3), 283-292. doi:10.1177/0361684313482109

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25*, 82–91.

Appendix A

SWE Survey

Student Feedback Form

Thank you for your participation.

Please take a few minutes to fill out this survey. Your comments and ideas will help make this event better.

If you have questions, please ask a volunteer.

1. With what races or ethnicities do you most identify? (Choose all that apply)

    a. White or European American
    b. Hispanic, Latino, or Spanish
    c. Black or African-American
    d. Asian American
    e. Native Hawaiian or Pacific Islander
    f. Native American or Alaskan Native
    g. Other: _____

2. How old are you? _____

3. The following statements describe work you might do in the future. Please check the box to show your answer.

How important is it to you to do work that…

	Very Unimportant	Unimportant	Not Sure	Important	Very Important
…is innovative (where you can come up with new ideas and inventions)					
…is creative					
…is hands-on					
…is fun to do					
… allows you to help your community and/or society					
…could be in many different kinds of career fields					

Student Feedback Form

4. Please tell us how much you agree with the following statements. Please check the box to show your answer.

	Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
Engineers are innovative (they come up with new ideas and inventions)					
Engineers are creative					
Engineers do work that is hands-on					
Engineers do work that is fun					
Engineers do work that allows them to help their community and/or society					
Engineers work in many different kinds of career fields					

5. Please tell us how much you agree with the following statements about today's event. Please check the box to show your answer.

	Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
Before this event, I was interested in becoming an engineer.					
After this event, I am interested in becoming an engineer.					
Before this event, I knew what an engineer did.					
After this event, I know what an engineer did.					
I made some new friends today.					
I enjoyed the fact that this was an event just for girls.					

Student Feedback Form

	Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
I worked with a mentor who was helpful and easy to talk to.					

6. Think back on your day, look at the following items, and check the box that best completes the sentence.

	Improved	Stayed the Same	Got Worse	I Don't Know
My confidence in problem-solving...				
My confidence in building and designing things...				
My ability to brainstorm solutions to problems...				
My ability to think of many different possible ways to solve a problem...				
My ability to use the design process (brainstorm, design, build, test, redesign)...				

7. Are you more interested in being involved with the Girl Scouts after attending this event?

    a. Yes
    b. No
    c. Maybe
    d. I am already a Girl Scout

8. Would you recommend that other kids participate in events like this?

    a. Yes
    b. Maybe
    c. No (please explain):

9. What did you like most about the event today?

Student Feedback Form

10. If you were in charge, how would you change this event?

Thanks very much for your help! Please hand in your completed survey.

VITA

Mary Bonk Isaac
San Diego, California

Education

M.A.T.	North Carolina A&T University	2011 Technology Education
B.S.	Union College, NY	1982 Mechanical Engineering

Certifications

Technology Education Secondary Instruction License (North Carolina)
Technology Education Secondary and Adult Credential (California)

Professional Experience

2007-Present	HEDGE Co, *Principal Consultant*
2005-2007	GE Energy, *Customer Quality Manager and Account Executive*
2001-2005	GE Energy, *Region General Manager, Sales*
1998-2001	GE Power Systems, *Risk Leader*
1994-1999	GE Power Systems, *Commercial Director – India and Subcontinent*
1992–1994	GE Power Systems, *Programs Manager – Parts*
1990–1992	GE Power Systems, *Account Manager – Power Generation*
1989-1990	GE Power Systems, *Marketing Specialist – Power Generation*
1985-1989	GE Power Systems, *Turbine Service Engineer – Service Shops*
1982-1985	GE Power Systems, *Nuclear Field Engineer*
1979-1982	Union College, *Machining Assistant*
1976-1979	GE Power Systems, *Apprentice Machinist*

Service

Society of Women Engineers (SWE)

- *Program Development Grants Committee Chair 2015-2017*
- *Director of Advocacy 2013-2015*
- *Outreach Committee Chair 2010-2013; PDG liaison 2013-15*
- *Fellow 2011; Fellow award reviewer 2012-current*
- *VP Members at Large (MALs) 2011-2012, MAL rep 2013, 2014, 2015*

International Technology and Engineering Educators Association (ITEEA)

- 21st Century Leadership Academy – Class of 2012
- Marketing subgroup Council of Technology and Engineering Teacher Educators

American Society of Engineering Educators (ASEE)

- *At Large Director for Women in Engineering Division 2010-11; Webmaster 2011-2013*
- *At Large Director K-12 Division 2012-14*
- *Annual conference paper reviewer (K-12 and Women in Engineering divisions) 2012-2015*

Member Board of Directors, Girls Inc. of Wilmington (NC) 2008, 2009

Publications

Phelps, M. (2012). Effects of hands-on activities on middle school females' spatial skills and interest intechnology-based careers. *ASEE 2012 Conference Proceedings* (pp. AC2012-3268). American Society for Engineering Education. Retrieved from http://www.asee.org/ public/conferences/8/papers/3268/view

Phelps, M. B. (2013, November). What goes up, must come down... *Technology and Engineering Teacher*, pp. 26-30. https://dl.dropboxusercontent.com/u/18345362/wgumcd.pdf

Isaac, M. B. (2015, January). Reinvention: From customer executive to doctoral student. SWE: *Magazine of the Society of Women Engineers*. p. 56.