Development and Validation of a Survey to Measure Perceived Team Communication Skills in Middle and High School STEM Out-of-School Time Programs

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Amy Lynn Grack Nelson

IN PARTIAL FULFILLMENT OF THE REQUIERMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Dr. Frances Lawrenz, Advisor

December 2017

Acknowledgements

This dissertation research would not have been possible without the assistance and support of many people. Words can't adequately express the depth of my gratitude to all of you. It certainly took a village to reach the end of my dissertation journey.

Thank you to the faculty and staff at the University of Minnesota who have supported me throughout my doctoral studies. Frances Lawrenz, thank you, thank you, thank you! I was so lucky to have you as an adviser. You have been such a great mentor and friend. I've learned so much from working with you. Thank you for believing in this research project and collaborating with me on this NSF grant. Thank you to my committee members, Jean King, Mark Davison, and Andy Zieffler, who also provided guidance and feedback throughout my research process. I'd also like to thank Ginny Michel for all of her help as a Graduate Research Assistant on the NSF grant that funded this research.

I am blessed to work in such an unbelievably supportive work environment at the Science Museum of Minnesota. Marjorie Bequette, you've been so supportive as my supervisor throughout my doctoral journey. Thank you for all of your feedback and encouragement from the development of this idea as a collaborative grant proposal between the Science Museum of Minnesota and the University of Minnesota all the way to completion of this research. I couldn't have carried out this research without the assistance of many of my colleagues in the Department of Evaluation and Research in Learning who worked with me on the NSF grant that funded this work. You helped carry out think-aloud interviews, reviewed versions of the survey, transcribed think-alouds, translated consent forms into Spanish and Hmong, helped coordinate the pilot test, entered survey data, helped run reports for STEM OST programs, provided feedback on the study instruments, and so much more! Thank you to Marjorie Bequette, Patrick Cox, Choua de Sala Her, Juan Dominguez, Megan Goeke, Chrissy Haikel, Gretchen Haupt, Zdanna King, Cynthia Matthias, Shannon McManimon, David Milavitz, Felicia Orozco, Rebecca Orozco, Maggie Sandford, Joe Schantz, Kalyssa Soucy, Amanda Svantesson DeGidio, Scott Van Cleave, Katherine Wood, and Della Xiong, I'd also like to thank my colleagues who took on additional work

responsibilities on our collaborative projects so that I could dedicate more time and space to finish writing my dissertation: Marjorie Bequette, Lauren Causey, Gretchen Haupt, Choua de Sala Her, Megan Goeke, Kaleen Tison Povis, and Zdanna King.

A huge thank you to all of the STEM OST programs across the country (and youth in these programs) that participated in this study. I'm so grateful for your interest, support, and involvement in this project.

Thank you to all of the organizations and individuals listed in Appendix C that spread the word about the study and helped with recruitment.

Thank you to the National Science Foundation's Promoting Research and Innovation in Methodologies for Evaluation (PRIME) program that provided funding for the research project Collaboration in the 21st Century (Award#1335883) that is the basis of this dissertation. Thank you to Eduardo Salas, Danial Bolt, and Beki Sato who were research advisors on this grant and provided valuable feedback on the construct, instrument, and research process.

A special acknowledgement to all of my friends and family that have been by my side throughout this journey. There are too many of you to name, but you all know who you are. Thank you for much needed study breaks, words of encouragement, and unending support and love. You kept me lifted up and moving forward over these last seven years.

And finally, a very special thank you to my husband, Travis. Thank you for escaping with me to the north shore every year so I could unplug and focus on my research, keeping me well fed when all I could do was focus on getting my dissertation done, and supporting me throughout this journey. I am so lucky to have you by my side!

Dedication

Mom and Dad

You taught me to always follow my dreams, believed in me no matter what path I decided to take, and instilled in me the value of hard work and dedication. I made it this far because of the amazing love and support you have shown me throughout my academic career from elementary to graduate school.

Travis

You are my rock. I could not have taken this journey and reached my destination without you.

Abstract

Twenty-first century skills are vital for preparing youth for careers in science, technology, engineering, and math (STEM) fields. STEM out-of-school time (OST) programs play an important role in helping youth develop these skills, particularly the teamwork skills necessary for the growing collaborative nature of STEM jobs. However, there is a lack of appropriate measures to evaluate this key programmatic outcome in STEM OST settings. This dissertation research addresses the lack of measures through the development of an instrument to assess team communication skills in middle and high school STEM OST programs.

The instrument was developed and validity evidence was gathered through a rigorous four-phase process. Phase 1 focused on identifying and operationalizing the teamwork skill area to be measured by the instrument. The skill area of team communication was most common among STEM OST programs and was defined as information exchange, closed-loop communication, and listening. In Phase 2, the survey scenario and items were developed and then reviewed by experts in STEM OST, youth development evaluation, teamwork, and measurement. Phase 3 involved think-aloud interviews and a national pilot test. Revisions to the survey occurred throughout each phase, leading to the final phase: a national field test of the instrument with 959 youth from 40 STEM OST programs across the country. Through confirmatory factor analysis, a five-factor model of team communication skills was found to be a good fit. The model included two factors for closed-loop communication, two factors for information exchange, and a listening factor. Responses for each of the five factors were reliable with coefficient alphas ranging from $\alpha = .70$ to $\alpha = .79$. The final instrument is a 28-item scenario-based, self-report measure of middle and high school youths' perceptions of their team communication skills. The survey instrument and operationalization of team communication skills in STEM OST programs will be valuable for both the evaluation and STEM OST fields.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Twenty-first century skills are vital for preparing our nation's youth to become tomorrow's innovators, researchers, and leaders in the science, technology, engineering, and math (STEM) fields. STEM out-of-school time (OST) programs play an important role in helping youth develop the 21st century skills they need to prepare them for the workforce, particularly the teamwork skills necessary for the growing collaborative nature of work in STEM (National Research Council, 2015). However, there is a lack of appropriate measurement tools to evaluate this key programmatic outcome in STEM OST settings. Many evaluators of middle and high school STEM OST programs need to develop their own data collection instruments, modify existing instruments, or use ones that lack appropriate validity evidence for the population and/or context being studied. The National Science Foundation-funded research described in this dissertation addresses this need by developing and validating an instrument to measure the teamwork skills of team communication in middle and high school STEM OST programs.

Importance of 21st Century Skills

Twenty-first century skills constitute a wide range of skills including creativity, innovation, critical thinking, problem solving, communication, teamwork, leadership, information literacy, and many more (Binkley et al., 2012; National Research Council, 2011, 2012a; Partnership for 21st Century Learning, 2015). There has been a strong federal emphasis on the importance of these skills. In 2002, the Partnership for 21st Century Skills was formed with funds from the U.S. Department of Education to bring together "educators, administrators, parents, business and community leaders, and others to build a consensus on the definition of 21st century skills" (Partnership for 21st Century Skills, 2003, p. 2). From 2008 to 2009, the Institute of Museum and Library Services led a national 21st century skills project and as a result created a vision for how libraries and museums can become part of the national efforts around 21st century skills (Institute of Museum and Library Services, 2009). Starting in 2010, The National Academies of Science, Engineering, and Medicine convened three separate workshops and published

follow up reports focused on various topics related to 21st century skills: the intersection of science education and 21st century skills (National Research Council, 2010); the research on assessing 21st century skills (National Research Council, 2011); and defining 21st century skills, how to teach them, and how they relate to adult outcomes (National Research Council, 2012a). The national importance of 21st century skills has also been stressed at the policy level with the introduction of the 21st Century Readiness Act in 2011 and 2013 and the reestablishment of a Congressional 21st Century Skills Caucus in 2016, which then introduced the Critical Thinking, Collaboration, Communication, and Creativity for Careers Act (govtrack.us, 2017a, 2017b; Hand, 2016).

Importance of Teamwork Skills in STEM

The 21st century skill of teamwork is of particular importance for preparing today's youth to participate in the STEM workforce of the future. The collaborative nature of STEM is emphasized throughout various STEM education publications and standards including the Next Generation Science Standards, the Framework for K-12 Science Education, and the Principles of K-12 Engineering Education (National Research Council, 2009a, 2012b; NGSS Lead States, 2013). Another indicator of the emphasis on collaboration in STEM fields is the increase in the average number of authors on science and engineering articles over time, with the latest statistics stating 90% of all science and engineering publications have two or more coauthors (National Research Council, 2015). Seeing this trend, the National Science Foundation recently requested a consensus study to understand the benefits and challenges of collaborative scientific research or "team science" (National Research Council, 2015).

STEM OST Programs and Teamwork Skills

STEM OST programs are important environments for youth from diverse backgrounds to develop the teamwork skills they need to enter and prosper in the STEM workforce. According to the 2014 America After 3PM survey, over 10 million youth participate in afterschool programs and, of these, approximately seven million participate in programs that offer STEM activities (Afterschool Alliance, 2015). A defining feature of STEM OST programs is that they frequently offer collaborative learning

experiences (National Research Council, 2009b; Schwarz & Stolow, 2006). In fact, some programs are completely team-based and refer to the program itself as a team (FIRST LEGO League, 2017; Science Museum of Minnesota, 2017). In addition to the collaborative experience, many STEM OST programs place emphasis on teamwork skills as a key programmatic outcome (Grack Nelson, 2013). This emphasis on the importance of teamwork outcomes is echoed in a number of publications outlining common outcomes across STEM OST programs and OST programs in general (Afterschool Alliance, 2013; Partnership for After School Education, 2010; The Forum for Youth Investment & National Collaboration for Youth Research Group, 2012; Wilson-Ahlstrom, DuBois, Ji, & Hillaker, 2014; Wilson-Ahlstrom, Yohalem, DuBois, & Ji, 2011).

Defining the Construct of Teamwork Skills

To understand the construct of teamwork skills, it is important to understand what teams are and what makes them effective. A team is "two or more individuals with different roles and responsibilities who interact socially and interdependently within an organizational system to perform tasks and accomplish common goals" (National Research Council, 2015, p. 2). Certain factors are necessary for a team to work effectively, including a certain level of proficiency in various skills. Much has been written over the last few decades in organizational psychology about what these skills entail. A number of researchers have synthesized this theoretical and empirical literature in an effort to outline and define the skill requirements for effective teamwork in adult teams (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995; Salas, Rosen, Burke, & Goodwin, 2009; Stevens & Campion, 1994). Teamwork and collaboration skills are also operationalized within the context of 21st century skills in the education literature. Much of the work that has been done around defining the skills in education is not directly based on the theoretical underpinnings of effective teams, but based on syntheses of various definitions and lists of 21st century skills (Binkley et al., 2012; National Research Council, 2012a; Organization for Economic Cooperation and Development, 2005; Partnership for 21st Century Learning, 2015). Overall, the organizational psychology literature describes a wider range of skills necessary for teamwork in comparison to the 21st

century skills literature (see Table 1 in Chapter 2 for a detailed comparison). The differences between the two literatures suggest that the theoretically- and empirically-based organizational psychology literature may be of value in strengthening the ways teamwork skills are operationalized within the context of 21st century skills. (See Chapter 2 for a more detailed discussion of these two literatures and how they each define teamwork skills).

Measuring Teamwork Skills

A review of the organizational psychology, 21st century skill, and STEM OST literatures provides insight into how evaluators and researchers have started to measure teamwork skills in both youth and adults. A variety of data collection methods have been used, including self-report surveys, situational judgment tests, ratings from other people, and observations. Each of these methods has benefits and limitations for assessing teamwork skills. Self-report surveys tend to be easy to administer, low in cost, and allow people to reflect on all of their teamwork experiences when responding (Lai & Viering, 2012; Zhuang, MacCann, Wang, Liu, & Roberts, 2008). A limitation is that people may not realize the extent of their skill and thus rate themselves inaccurately, or the opposite could occur and they rate themselves in a way that they perceive to be the socially desirable response (Bedwell, Fiore, & Salas, 2011; National Research Council, 2011). Situational judgment tests are detailed teamwork scenarios where individuals are asked to indicate what they would be likely to do in the described situation (Lai & Viering, 2012; National Research Council, 2011). Since they are based on real-world experiences and people are asked to apply their skills in an actual situation, they can be more authentic measures (Lai & Viering, 2012). However, similar to concerns with self-report surveys, people might be able to figure out the socially desirable response in a situation (Lai & Viering, 2012; National Research Council, 2011). Observations and ratings from other people, such as peers or teachers, are based on seeing someone exhibit their skills in an actual teamwork experience. Ratings and observations can have fewer problems with faking and bias than self-report measures (Connelly & Ones, 2010; Lai & Viering, 2012). A limitation is that in order to get a holistic view of an individual's teamwork skills, they need to be observed on multiple occasions,

which can be time intensive and costly. There are also issues around how observers might interpret what they see and consistency in ratings across observers (Bedwell et al., 2011; Kyllonen, 2012).

In the literature review in Chapter 2, a variety of instruments that measure teamwork skills for youth and adults are reviewed. However, these instruments have limitations that make them inadequate for STEM OST program evaluation. One limitation is validity evidence collected for the instruments. Some of the instruments have only been tested with adults and are related to adult situations such as college courses or workplace settings, so they do not have validity evidence for use with middle and high school youth (Aguado, Rico, Sánchez-Manzanares, & Salas, 2014; Baker, Horvath, Campion, Offermann, & Salas, 2004; Chiu, 2014; Loughry, Ohland, & Moore, 2007; Stevens & Campion, 1999). Some instruments developed for youth do not have the necessary validity evidence for a wide range of STEM OST programs. One instrument has validity evidence for formal education settings, but not informal (Zhuang et al., 2008). Another has validity evidence for camps, but camps are only one of the many types of STEM OST programs (Sibthorp, Bialschki, Morgan, & Browne, 2013). In one case it is unclear what validation, if any, an instrument went through as there is lack of reported validity evidence (MHA Labs, n.d.). There is a concern about the quality of the items for a few instruments, and it is unclear if response process validity evidence was gathered for the items (Hansen & Larson, 2002, 2005). Another limitation beyond appropriate validity evidence is the ability to easily access some of the instruments either because there is a fee associated with their use or an instrument is not easily found online (American Camp Association, 2013; MHA Labs, n.d.).

Research Questions

This research recognizes the important role STEM OST programs play in developing teamwork skills and the lack of measurement tools with the validity evidence necessary to use for the evaluation of STEM OST programs. This research helps to fill this gap through the development and validation of an instrument to measure perceived teamwork skills in middle and high school STEM OST programs. The instrument development was guided by the following research questions:

1. What skill area of the broad range of teamwork skills best aligns with middle and high school youth outcomes in STEM OST programs and the evaluation needs of these programs?
2. Is the construct of team communication skills unidimensional or multidimensional?[1]
3. To what extent does the developed instrument gather reliable data from youth in STEM OST programs?
4. To what extent is there adequate validity evidence for the developed instrument?

These research questions were answered through a rigorous four-phase instrument development and validation process based on standards from the field of educational measurement (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). The process helped to ensure that the survey developed through this research gathered reliable data and had validity evidence for use with the diverse middle and high school youth populations participating in STEM OST programs.

[1] This research question emerged after answering Research Question 1 (as described in Chapter 3) by identifying team communication skills as the construct area of focus for the instrument.

Chapter 2: Literature Review

Teamwork skills are known as one of the 21st century skills, a group of skills that are important for preparing today's youth for the world of work. As this literature review details, there is a lack of instruments to measure teamwork skills. This is particularly true for STEM OST programs where evaluators are looking for no-cost tools with adequate validity evidence for these settings. Presently, evaluators of STEM OST programs have to develop their own instruments, modify existing instruments, or use ones that lack validity evidence for the population and/or context being studied. This dissertation research addressed this need through the development and validation of instruments to measure teamwork skills in middle and high school STEM OST programs. To inform the development of the instruments, the literature review was guided by two main questions:

1. How are teamwork skills operationalized in the literature?
2. To what extent do the instruments currently available for measuring teamwork skills have validity evidence for use with middle and high school youth in STEM OST contexts?

The chapter begins with an examination of the history and importance of 21st century skills and more specifically why teamwork skills are vital for future careers, particularly in STEM. The chapter will then discuss the various ways teamwork skills have been defined in the literature. Finally, the chapter will examine a variety of instruments that have been used to measure teamwork skills and discuss the benefits and limitations of these methods in relation to their potential use in evaluating middle and high school STEM OST programs.

History and Importance of the 21st Century Skills

Twenty-first century skills constitute a wide range of skills vital for preparing the nation's youth to enter the 21st century STEM workforce. These skills are also sometimes referred to as soft skills, workforce skills, applied skills, non-cognitive skills, and life skills (Casner-Lotto & Barrington, 2006; Silva, 2008; The Secretary's Commission on Achieving Necessary Skills, 1999). Twenty-first century

skills include (Binkley et al., 2012; National Research Council, 2011, 2012a; Partnership for 21st Century Learning, 2015):

- Creativity and innovation.
- Critical thinking and problem solving.
- Communication.
- Teamwork and collaboration.
- Flexibility and adaptability.
- Self-direction, self-regulation, and self-management.
- Leadership and responsibility.
- Social and cross-cultural skills.
- Information literacy.
- Information and communication technology literacy.
- Productivity and accountability.
- Media literacy.

The emphasis placed on these skills has increased over the last 20 years. Workforce skills gained national attention in 1990, when the Secretary's Commission on Achieving Necessary Skills (SCANS) was charged with identifying the skills necessary for workplace readiness of the nation's young people (The Secretary's Commission on Achieving Necessary Skills, 1999). The Commission's resulting skill list included both workplace competencies (resource, information, interpersonal, systems, and technology competencies) and foundation skills (basic skills, thinking skills, and personal qualities). Many of these competencies and skills overlap with those that are now considered 21st century skills.

The federal emphasis of these skills was revisited at the start of the 21st century. In 2002, the Partnership for 21st Century Skills was formed with funds from the U.S. Department of Education (Partnership for 21st Century Learning, 2017). Then U.S. Secretary of Education Rod Paige described the

Partnership as bringing together "educators, administrators, parents, business and community leaders, and others to build a consensus on the definition of 21st century skills" (Partnership for 21st Century Skills, 2003, p. 2). The Partnership developed a 21st Century Skills Framework that defined the 21st century skills and served as the foundation for the Partnership's national conversations around curriculum and instruction, professional development, and assessment of the skills.

In 2009, the National Academies began convening experts to address various topics around 21st century skills. The first workshop, in 2009, focused on the intersection of science education and 21st century skills (National Research Council, 2010). In 2011, with funding from the National Science Foundation and the National Institutes of Health, a workshop was held to discuss the research on assessing 21st century skills (National Research Council, 2011). In 2012, through partial funding from the National Science Foundation, a committee was gathered to further define the 21st century skills, how to teach them, and how they relate to adult outcomes (National Research Council, 2012a). The report resulting from this meeting, *Education for Life and Work: Developing Transferrable Knowledge and Skills in the 21st Century*, stressed the need for continued work and support around research of 21st century skills, stating that "foundations and federal agencies should support research aimed at establishing agreed-upon definitions of 21st century competencies and ways to measure and assess them" (National Research Council, 2012c, p. 4). In 2015, the National Academies of Sciences, Engineering, and Medicine held a symposium titled *Assessing Hard-to-Measure Cognitive, Interpersonal, and Interpersonal Competencies*. This symposium was a follow up to the *Education for Life and Work* report and provided an opportunity for staff from National Science Foundation funded projects addressing these competencies to come to together to talk about how to best measure them. The project described in this dissertation was one of the National Science Foundation-funded projects invited to present at the symposium.

The national importance of 21st century skills was also stressed at the policy level in 2011 when the 21st Century Readiness Act was introduced as an amendment to the Elementary and Secondary Education Act of 1965. The purpose of the bill was:

> To provide, develop, and support 21st century readiness initiatives that assist students in acquiring the skills necessary to think critically and solve problems, be an effective communicator, collaborate with others, and learn to create and innovate. ("21st Century Readiness Act," 2011)

The bill called for the addition of a 21st century skills focus to school reform, professional development, 21st Century Learning Center activities, and the creation of assessments to measure these skills. The bill was referred to the House Education and Work Force Committee, but died in committee (govtrack.us, 2014). The bill was reintroduced to the 113th Congress in 2013, but was not enacted (govtrack.us, 2017a). In 2016, Congressmen Loebsack and Costello reestablished the Congressional 21st Century Skills Caucus that had been originally formed during the 112nd Congress in 2011. The intent of the Caucus was to "serve as a bipartisan forum for advancing the discussion on how to effectively promote 21st century skills in our nation's educational system and better prepare our students with the right knowledge and skills to compete in a global economy" (Hand, 2016). In 2016, the Caucus introduced the Critical Thinking, Collaboration, Communication, and Creativity for Careers Act as an amendment to the Carl D. Perkins Career and Technical Education Act of 2006, but it was not enacted (govtrack.us, 2017b). One of the intentions of the Act was to:

> Improve the critical thinking, communications, collaboration, and creativity skills of students participating in career and technical education programs, including by (i) integrating such skills into coursework through project-based learning; (ii) building the capacity of educators to teach such skills; and (iii) providing ongoing support to help students achieve such skills. ("Critical Thinking, Collaboration, Communication, and Creativity for Careers Act," 2016)

The international STEM community has also endorsed the importance of 21st century skills. In 2009, three computing technology companies, Intel, Cisco, and Microsoft, sponsored the international research project: Assessment and Teaching of 21st Century Skills (Assessment and Teaching of 21st Century Skills, 2012). Participating in this project were education officials from six countries, including

the U.S. Secretary of Education. In 2010, the project commissioned a series of white papers from international experts to address measurement issues around 21st century skills (Griffin, Care, & McGaw, 2012).

Although much of the work around 21st century skills has focused on formal education, the informal education field also plays a significant role in helping individuals develop these skills. From 2008 – 2009, the Institute of Museum and Library Services led the Museums, Libraries, and 21st Century Skills project. A major outcome of the project was a report that outlined a vision for how libraries and museums could become part of the national efforts around 21st century skills, defined a set of 21st century skills relevant to museums and libraries,[2] and included a self-assessment to help museums and libraries identify strategies for developing the 21st century skills of their audiences (Institute of Museum and Library Services, 2009).

The importance of 21st century skills has also been stressed throughout a wide range of OST publications. The journal *New Directions for Youth Development* devoted an entire issue to the case for 21st century learning (Schwarz & Kay, 2006). Many of the 21st century skills also show up in national publications of common outcomes and indicators for OST youth programs, including STEM-focused programs (Afterschool Alliance, 2013; Every Hour Counts, 2014; Partnership for After School Education, 2010; The Forum for Youth Investment & National Collaboration for Youth Research Group, 2012; Wilson-Ahlstrom et al., 2014; Wilson-Ahlstrom et al., 2011).

Importance of Teamwork Skills in STEM

The 21st century skills of teamwork are particularly important for the workforce. In 2006, a national survey asked 431 employers across the country to rate the importance of 20 skills for workplace readiness (Casner-Lotto & Barrington, 2006). Their list included applied skills (which included many of the 21st century skills) and basic knowledge (such as math and science). Three-quarters of the employers

[2] The Institute of Museum and Library Services' list of 21st century skills was adapted from the Partnership for 21st Century Skills' framework (Institute of Museum and Library Services, 2009).

surveyed rated teamwork skills as a "very important" skill for high school graduates, even more so than the basic knowledge. The importance they expressed toward teamwork skills was similar to the importance that has been placed on these skills in relation to STEM workforce readiness.

The collaborative nature of STEM is emphasized throughout various STEM education publications and standards. The Next Generation Science Standards discuss teamwork when addressing the nature of science. Part of understanding the scientific enterprise is recognizing that science is a human endeavor often carried out by teams of individuals, which is illustrated in one of the learning outcomes for grades 3-5, "Most scientists and engineers work in teams" (NGSS Lead States, 2013, p. 434). *A Framework for K-12 Science Education* discusses how the fields of science, engineering, and technology can be interdependent and individuals in these fields often must work together to solve problems (National Research Council, 2012b). This is particularly true for engineering, where working in multidisciplinary teams to solve engineering problems is only becoming more common (National Research Council, 2004). The Principles of K-12 Engineering Education further stress this need and state that engineering education programs should include teaching engineering "habits of mind," which include collaboration and some of the other 21st century skills (National Research Council, 2009a).

Another indicator of the growing emphasis on collaboration in the STEM fields is the increase in the average number of authors on science and engineering articles over time. From 1955 to 2000, the average number of authors increased from 1.9 to 3.5 (Wuchty, Jones, & Uzzi, 2007). This was true across 171 science and engineering subfields, where 99% have had an increase in collaborative authorship. The trend in collaborative authorship of science and engineering articles continued, with two-thirds of articles in 2010 including more than one author and the average number of authors increasing from 3.2 in 1990 to 5.6 in 2010 (National Science Board, 2012). Further evidence of this continuing trend is the National Research Council publication *Enhancing the Effectiveness of Team Science*, which states that 90% of all science and engineering publications have two or more coauthors (National Research Council, 2015). The publication was the result of a consensus study requested by the National Science Foundation in response

to the increasing collaborative nature of research, referred to as "team science," the benefits team science offers, and the challenges it can pose. The purpose of the study was to "recommend opportunities to enhances the effectiveness of collaborative research in science teams, research centers, and institutes" (National Research Council, 2015, p. 1).

STEM OST Programs and Teamwork Skills

STEM OST programs are important environments for youth to develop the teamwork skills they need to enter and prosper in the STEM workforce. Youth spend only 18.5% of their waking hours in school, so OST programs that happen before school, afterschool, on weekends, and in the summer offer important learning opportunities (National Research Council, 2009b). According to the 2014 America After 3PM survey, over 10 million youth participate in afterschool programs and of these approximately 7 million participate in programs that offer STEM activities (Afterschool Alliance, 2015).

STEM OST programs offer opportunities to build on youths' interests and offer skill building activities that they may have limited access to during school. A defining feature of STEM OST programs is that they frequently offer collaborative learning experiences (National Research Council, 2009b; Schwarz & Stolow, 2006). In fact, some programs are completely team-based and refer to the program itself as a team. For example, the ThoughtSTEM program emphasizes the importance of their team structure in teaching youth how to code: "At ThoughtSTEM, you learn teamwork because no big piece of software was built by just one person" (ThoughtSTEM, 2014). STEM OST programs in the Science Museum of Minnesota's Kitty Andersen Youth Science Center are called "teams" or "crews" (Science Museum of Minnesota, 2017). On their website, a core value of the national FIRST LEGO League program is "we are a team" (FIRST LEGO League, 2017).

In addition to the collaborative experience, many STEM OST programs place emphasis on teamwork skills as key programmatic outcomes. The Afterschool Alliance (2013), a national organization with over 25,000 afterschool program partners, developed an outcomes framework for STEM OST programs with input from afterschool providers and funders. One of the common outcomes in their

framework was that "youth develop a capacity to productively engage in STEM learning activities" (Afterschool Alliance, 2013, p. 6). An indicator of this was that youth exhibit the "ability to exercise STEM-relevant life and career skills" with the sub-indicator "demonstration of ability to work in teams to conduct STEM investigations" (Afterschool Alliance, 2013, p. 6). While writing the grant proposal that funded this dissertation's research, a survey was administered to find out the extent to which STEM OST programs had goals or outcomes related to teamwork skills. The pre-proposal survey results echoed the Afterschool Alliance's findings that teamwork skills were common STEM OST outcomes. Of the 82 STEM OST professionals who responded to the survey, over two-thirds (68%) had teamwork goals or outcomes as part of their program (Grack Nelson, 2013). In the larger OST field, the Partnership for After School Education pulled together a Youth Outcomes Committee to develop an inventory of common youth outcomes across OST programs in general (Partnership for After School Education, 2010). Among these outcomes was the relationship skill of "increased ability to work with others to accomplish goals" (Partnership for After School Education, 2010, p. 10). A 2010 survey of members of the National Collaboration for Youth, a coalition of more than 50 of the nation's leading youth organizations, also resulted in a framework of common outcomes and indicators across all types of youth programs (The Forum for Youth Investment & National Collaboration for Youth Research Group, 2012). The ability to work effectively in groups was among the workforce readiness outcomes identified in the framework.

Defining the Construct of Teamwork Skills

To understand the construct of teamwork skills, it is important to first define what is meant by the term "team" and how it might differ from other kinds of groups. According to the National Research Council, a team is "two or more individuals with different roles and responsibilities who interact socially and interdependently within an organizational system to perform tasks and accomplish common goals" (2015, p. 2). Johnson and Johnson (2013) stress that a team can be a group, but a group of people is not necessarily a team.

A team consists of two or more individuals who are (a) aware of their positive interdependence as

> they strive to achieve mutual goals, (b) interact while they do so, (c) are aware of who is and is not a member of the team, (d) have specific roles or functions to perform, and (e) have a limited life span of membership. (Johnson & Johnson, 2013, p. 500)

Some key features of teams include "member interdependency, a common goal, dynamic exchange of information, coordination of task activities, and some structuring of member roles" (Prichard, Bizo, & Stratford, 2006, p. 120).

Certain factors are necessary for a team to work effectively. Salas, Sims, and Burke (2005) answered the question "What is teamwork?" by reviewing 138 theoretical and empirical models of teamwork. From their analysis, they identified the "big five" components most essential for effective teamwork. These core factors included: (a) team leadership, (b) mutual performance monitoring, (c) backup behavior, (d) adaptability, and (e) team orientation. They also identified three coordinating mechanisms that need to be in place to support these core factors: shared mental models among team members, closed-loop communication, and mutual trust. Salas et al. (2005) argued that these components and coordinating mechanisms must be present for a team to effectively complete a task.

The components that are essential for effective teamwork are closely connected to the teamwork skills team members need to possess. How teamwork skills are defined varies depending on the literature, as evidenced from the following review of the organizational psychology and 21st century skills literatures.

Organizational psychology literature. Much has been written in organizational psychology over the last few decades about teamwork skills and how these skills relate to team effectiveness. Stevens and Campion (1994) conducted a literature review to understand the knowledge, skills, and ability (KSA) requirements for teamwork. Their model was based on two broad categories of KSAs –interpersonal KSAs and self-management KSAs. Under interpersonal KSAs they stressed the need for individuals to be skilled in conflict resolution, collaborative problem solving, and communication. Self-management KSAs included goal setting and performance management, and planning and task coordination. Table 1 includes

descriptions of the knowledge, skills, and abilities under each of the subcategories of interpersonal and self-management KSAs.

Table 1

Stevens and Campion's (1994) Knowledge, Skill, and Ability Requirements for Teamwork (p. 505)

Subcategory of KSAs	Knowledge, skill, and ability (KSA) to…
Conflict resolution	• Recognize and encourage desirable, but discourage undesirable, team conflict. • Recognize the type and source of conflict confronting the team and to implement an appropriate conflict resolution strategy. • Employ an integrative (win-win) negotiation strategy rather than the traditional (win-lose) strategy.
Collaborative problem solving	• Identify situations requiring participative group problem solving and to utilize the proper degree and type of participation. • Recognize the obstacles to collaborative group problem solving and implement appropriate corrective actions.
Communication	• Understand communication networks, and to utilize decentralized networks to enhance communication where possible. • Communicate openly and supportively, that is, to send messages which are: (a) behavior- or event-oriented, (b) congruent, (c) validating, (d) conjunctive, and (e) owned. • Listen nonevaluatively and to appropriately use active listening techniques. • Maximize consonance between nonverbal and verbal messages, and to recognize and interpret the nonverbal messages of others. • Engage in ritual greetings and small talk and recognition of their importance.
Goal setting and performance management	• Help establish specific, challenging, and accepted team goals. • Monitor, evaluate, and provide feedback on both overall team performance and individual team performance.
Planning and task coordination	• Coordinate and synchronize activities, information, and task interdependencies between team members. • Help establish task and role expectations of individual team members, and to ensure proper balancing of workload in the team.

Around the same time as Stevens and Campion, Cannon-Bowers et al. (1995) also carried out a review of the empirical and theoretical literature on teamwork and came up with a list of 130 skills necessary for effective teamwork. From this list emerged eight major dimensions of teamwork skills and associated subskills (see Table 2). There was overlap between many of Cannon-Bowers et al. (1995) teamwork dimensions and subskills and Stevens and Campion (1994) subcategories of teamwork skills.

Table 2

Cannon-Bowers et al.'s (1995) Dimensions and Subskills of Teamwork Skills (pp. 344 - 346)

Dimension	Subskills
Adaptability	Flexibility, compensatory behavior, dynamic reallocation of functions.
Shared situational analysis	Situational awareness, Shared problem-model development.
Performance monitoring and feedback	Intramember feedback, mutual performance monitoring.
Leadership and team management	Task structuring, mission analysis, motivation of others.
Interpersonal relations	Conflict resolution, cooperation, assertiveness, morale building, boundary spanning.
Coordination	Task organization, task interaction, timing and activity pacing.
Communication	Information exchange, consulting with others.
Decision making	Problem assessment, problem solving, planning, metacognitive behavior, implementation,

Over a decade later, Cannon-Bowers et al.'s dimensions of teamwork skills were updated based on new literature and a growing body of research on teams (Salas et al., 2009). As illustrated in Table 3, the updated list of dimensions expanded on and offered more nuanced dimensions of teamwork skills with no accompanying subskills.

Table 3

Salas et al.'s (2009) Dimensions of Teamwork Skills (pp. 55 - 61)

Dimension	Definition
Mutual performance monitoring	Keeping track of the work of fellow team members while carrying out your own.
Adaptability	Adjusting strategies based on information and conditions.
Backup/supportive behavior	Anticipating team members' needs and adjusting workloads accordingly.
Implicit coordination strategies	Synchronizing team members' actions based on unspoken assumptions.
Shared/distributed leadership	Shifting leadership in order to take advantage of team members' strengths.
Mission analysis	Being able to interpret the team's objectives.
Problem detection	Sensing when there is or could be a problem.
Conflict resolution/management	Establishing conditions to address conflict both before and when it happens.
Motivation of others	Motivating others to reach the team's objectives.
Intrateam feedback	Providing feedback to team members.
Task-related assertiveness	Sharing ideas and opinions with team members.
Planning	Generating a plan of action to reach team goals.
Coordination	Sequencing and timing team member tasks.
Team leadership	Developing and coordinating team member activities and motivating members to reach goals.
Problem solving	Identifying a problem and potential solutions.
Closed-loop communication/ information exchange	Communicating effectively between the sender and receiver of messages within the team.

Twenty-first century skills literature. Teamwork skills are also operationalized within the 21st century skills literature. Many of the discussions within the 21st century skills literature around defining teamwork skills are not directly related to the theoretical or empirical underpinnings of effective teams, but instead based on syntheses of various definitions and lists of 21st century skills.

Starting in 1997, the Organization for Economic Cooperation and Development (OECD) led the international Definition and Selection of Competencies project to develop a set of key competencies for the 21st century (Organization for Economic Cooperation and Development, 2005). One of the competencies was the "ability to cooperate." The OECD (2005) emphasized that in order to cooperate individuals need to be able to do the following: "(a) present ideas and listen to those of others, (b) understand the dynamics of debate and follow an agenda, (c) construct tactical or sustainable alliances, (d) negotiate, and (e) make decisions that allow for different shades of opinion" (pp. 12-13).

The Partnership for 21st Century Learning (2015) defined teamwork skills, or the ability to collaborate with others, as:

> (a) Demonstrate ability to work effectively and respectfully with diverse teams;
>
> (b) Exercise flexibility and willingness to be helpful in making necessary compromises to accomplish a common goal; and (c) Assume shared responsibility for collaborative work, and value the individual contributions made by each team member. (pp. 4-5)

It is unclear if the definitions were based on theoretical or empirical literature as the framework lacked citations and a description of how the framework was developed.

The Assessment and Teaching of Twenty-First Century Skills (ATC21S) project reviewed 12 frameworks of 21st century skills from around the world and combined their definitions to operationalize teamwork skills (Binkley et al., 2012). Teamwork, which they identified as important to the globalization of the workplace, requires skills such as the ability to "interact effectively with others, work effectively in diverse teams, manage projects, and guide and lead others" (Binkley et al., 2012, p. 47). Table 4 provides indicators for each of these four skills.

Table 4

The ATC21S' Collaboration and Teamwork Skills (Binkley et al., 2012, p. 23)

Skills	Indicators
Interact effectively with others	Speak with clarity and awareness of audience and purpose. Listen with care, patience, and honesty. Conduct themselves in a professional manner.
Work effectively in diverse teams	Leverage social and cultural differences to create new ideas and increase both innovation and quality of work.
Manage projects	Prioritize, plan and manage work to achieve the intended group result.
Guide and lead others	Use interpersonal and problem-solving skills to influence and guide others toward a goal. Leverage strengths of others to accomplish a common goal. Inspire others to reach their very best via example and selflessness. Demonstrate integrity and ethical behavior in using influence and power.

As part of a 2012 National Research Council convening around 21st century skills, a content analysis was carried out of eight reports that defined the 21st century skills. The convening identified common terms that were used to describe teamwork skills: communication, coordination, interpersonal skills, empathy/perspective taking, conflict resolution, negotiation, trust, and service orientation (National Research Council, 2012a).

Comparison of literatures. The lists of teamwork skills based on organizational psychology and 21st century skills literatures provided a framework to inform the operationalization of teamwork skills in STEM OST programs. Table 5 compares the skills within each of the literatures to highlight commonalities and differences. There were eight skills that came up in both sets of definitions of teamwork skills. Overall, the organizational psychology literature described a wider range of skills necessary for teamwork in comparison to the 21st century skills literature.

Table 5

Comparison of Teamwork Skills in the Organizational Psychology and 21st Century Skills Literature

Skills	Organizational psychology	21st century skills
Backup/supportive behavior	X	
Communication	X	X
Conduct themselves in a professional manner		X
Conflict resolution/negotiation	X	X
Coordination	X	X
Empathy/perspective taking		X
Flexibility/adaptability	X	X
Goal setting	X	
Interpersonal skills	X	X
Intrateam feedback	X	
Leadership	X	X
Mission analysis	X	
Motivation of others	X	
Mutual performance monitoring	X	
Problem detection	X	
Problem solving/decision making	X	X
Project management/planning	X	X
Shared responsibility		X
Task-related assertiveness	X	
Value individual contributions		X
Work with diverse teams		X

Note. Definitions are from Binkley et al. (2012); Cannon-Bowers et al. (1995); National Research Council (2012a); Organization for Economic Cooperation and Development (2005); Partnership for 21st Century Learning (2015); Salas et al. (2009); Stevens and Campion (1994).

Measuring Teamwork Skills

A review of the organizational psychology, 21st century skill, and STEM OST literatures provided insight into how evaluators and researchers have started to measure teamwork skills in both youth and adults. A variety of data collection methods have been used to measure these skills, which include self-

report surveys, situational judgment tests, ratings from other people, and observations. This section will describe how each of these methods has been used to measure teamwork skills, strengths and limitations of each data collection method, and examples of currently available instruments.

Self-report surveys. The most commonly used method for measuring these skills is a self-report survey, where individuals report on their own skill level. Self-report surveys are easy to administer and low in cost (Lai & Viering, 2012). An additional benefit of self-report measures is that the individual completing the survey can respond based on all of his/her teamwork experiences, whereas other methods, such as other-reports and observations, are limited to the experiences the person reporting or observing has with the individual (Zhuang et al., 2008). A limitation of self-report surveys is that youth may not realize how good they really are at teamwork skills and not rate themselves appropriately (National Research Council, 2011). To help ensure the validity of the self-report measure, survey ratings can be compared to ratings another person, such as an educator, might give for a youth. An additional limitation is the potential for someone to figure out what the socially desirable response might be and respond in that way instead of responding based on their true skill level (Bedwell et al., 2011). They may end up faking their skills, which introduces measurement error and affects the validity and reliability of the results. The fact that evaluations of STEM OST programs are typically not seen as "high-stakes" (youth are not graded or scored as they are in a formal education setting) may decrease incentives for youth to exaggerate or fake their skill level.

Evaluators of STEM OST programs often create teamwork self-report measures specific to their evaluation needs. They may develop them from scratch or pull items from other measures (for example, Klem & Nicholson, 2008; Norland, Foutz, & Krabill, 2009). Survey questions may also come across informal assessments, quizzes, or checklists such as those in the textbook *Joining Together: Group Theory and Group Skills* that are meant for instructional purposes but could be adapted for evaluation (Johnson & Johnson, 2013). A limitation of evaluators developing their own instruments is that these instruments may not have validity evidence that they are actually measuring the construct of interest and

they lack evidence for their use beyond the program being evaluated.

There are a few self-report instruments available that have been validated with youth: the Youth Outcomes Battery Teamwork scales, the Youth Experience Surveys, and the Educational Testing Service's Teamwork and Collaboration assessment. The strengths and limitations of each of these instruments in relation to their appropriateness for measuring teamwork skills in STEM OST programs are described below.

The Youth Outcomes Battery Teamwork scales were developed to measure teamwork skills in middle and high school residential and day camp programs (American Camp Association, 2013). The American Camp Association developed both a basic and detailed version of the scales. Both scales contained the same eight items, but had differing response options. The basic version has youth respond to eight statements by answering the question, "How much, if any, has your experience as a camper in this camp changed you in each of the following ways?" using a five-point scale with the options: (a) Decreased; (b) Did not increase or decrease; (c) Increased a little bit, maybe; (d) Increased some, I am sure; and (e) Increased a lot, I am sure (American Camp Association, 2013, p. 27). The basic scale had high reliability with a coefficient alpha of .94 and confirmatory factor analysis supported that the scale was one factor (Sibthorp et al., 2013). The detailed version of the scale asked youth to rate how true each statement was for them when they leaved the camp using a six-point scale of (a) False, (b) Somewhat false, (c) A little false, (d) A little true, (e) Somewhat true, and (f) True. They are also asked a reflective question, "Is the above statement more or less true today than before camp?" using a six-point scale: (a) A lot less, (b) Somewhat less, (c) A little less, (d) A little more, (e) Somewhat more, and (f) A lot more (American Camp Association, 2013). Reliability information for the detailed Teamwork scale was unavailable, but the American Camp Association (2013) stated that their suite of detailed scales has coefficient alphas that range between .84 to .93. There was a lack of internal structure validity evidence from confirmatory factor analysis for the detailed scale.

There were several important reasons why the Youth Outcomes Battery Teamwork Scales were

not a good fit for STEM OST programs. The Teamwork Scales have only been tested in one type of informal education program setting, camps, so they lack validity evidence that the instruments would work with a range of STEM OST programs. There is a fee for use ($30 if you are not a member of the American Camp Association). Although the fee is small, some evaluators may be hesitant to pay for measures they cannot see ahead of time and do not have validity evidence for the educational context they are evaluating.

There were also some survey design issues with the scales. The basic version of the scale did not have balanced response options, a common guideline for survey design of ordinal scales (Dillman, Smyth, & Christian, 2014). There was one negative option, a middle or neutral type of option, and three positive options. There were also issues with the response options for the detailed scale. There was no option for youth to indicate if they had no change from before to after camp. They could only respond if a statement was more or less true than before camp. If youth did not feel like there was any change, they would have to choose a response that was not actually indicative of their true response, which would introduce measurement error. The follow up statement also seemed like it would be cognitively difficult for youth to complete and difficult to interpret in reporting. If a youth responded "false" to a statement such as, "I can be a good group leader" at the end of camp, and then responded that the statement is "a lot less" true today than before camp it was unclear how the amount should be interpreted other than saying the rating decreased.

A second pair of instruments, the Youth Experience Survey versions 1.0 and 2.0, included items to measure high school youth's teamwork and social skills experiences in a broad range of organized OST programs (Hansen & Larson, 2002, 2005). The surveys included measures beyond just teamwork skills. Version 1.0 measured teamwork skills using 12 items in a 89-item survey (Hansen & Larson, 2002). Version 2.0 was slightly shorter, measuring teamwork using 10 items in a 71-item survey (Hansen & Larson, 2005). The coefficient alpha for the teamwork scales on both versions of the survey was .93 (Hansen & Larson, 2002, 2005). The Youth Experience Survey is free of cost and tested in informal

settings (Hansen & Larson, 2002, 2005).

A major limitation of the Youth Experience Surveys was they do not actually measure the development of skills. Instead they measure the extent of exposure to certain experiences during the program that may lead to skill development. Survey instructions said to "rate whether you have had the following experiences" on a four-point scale with the response options of (a) Yes, definitely, (b) Quite a bit, (c) A little, and (d) Not at all (Hansen & Larson, 2005, p. 13).

An additional issue was the response options did not seem to work well for some of the items. For instance, youth were asked to rate whether they had the experiences, "I became better at giving feedback" or "Learned to be patient with other group members" (Hansen & Larson, 2005, pp. 14-15). The authors acknowledged that some of the items may "reflect the adolescent's assessment of outcome" in addition to being a survey of experiences (Hansen & Larson, 2002, p. 4). The multiple interpretations of items affected the validity of the interpretations gathered from the measure.

The third instrument reviewed, the Educational Testing Service's Teamwork and Collaboration assessment for high school students, included a student survey composed of 57 self-report rating scale items, along with a situational judgment test and teacher survey that will be described later in this chapter (Wang, MacCann, Zhuang, Liu, & Roberts, 2009). The self-report survey was composed of 30 items that measured youth's views about teamwork and leadership. An example item was, "I enjoy bringing team members together," and youth responded based on how they thought or felt about the item using a six-point scale from never to always. The scale measured three dimensions of teamwork as evident from their confirmatory factor analysis results, and each scale had high reliability: cooperation ($\alpha = .88$), advocate/guide ($\alpha = .80$), and negotiation ($\alpha = .78$).

There were a number of limitations of the Educational Testing Service's Teamwork and Collaboration assessment. The survey lacked validity evidence for use outside a formal school setting. Also, there was not an actual survey available to just pick up and use. Zhuang et al.'s (2008) paper only

included a copy of an earlier version of the survey they used for psychometric testing with 57 items. The paper had a table with the final 30 items and their factor loadings under each dimension if someone wanted to create their own survey, however, people may have difficulty figuring out how to compute the individual factor scores. For two items, they included factor loadings for two dimensions, making it unclear which dimension the item was meant to fit under if someone was unfamiliar with factor analysis and what factor loadings mean, which is very likely for some evaluators and STEM OST educators.

One adult self-report survey was also reviewed. The Teamwork Competency Test was designed based on Steven and Campion's (1994) teamwork knowledge, skills, and ability model and definitions previously described. The Teamwork Competency Test was composed of 36 items with a four-point frequency ratings scale from never/almost never to always/almost always (Aguado et al., 2014). The scale was tested with psychology students at a university in Spain. A five-factor solution was proposed. Four of the scales had coefficient alpha values over .80, while the other scale had a coefficient alpha of .71. However, the confirmatory factor analysis did not point to a five factor solution as suggested by the Stevens and Campion (1994) model. The scale does not have validity evidence for youth populations or English-speaking populations in the United States as it was translated into Spanish before testing, so it would be inappropriate for measuring the construct of teamwork skills in the context of STEM OST programs in the United States. It is also unclear how to obtain the final instrument as the items are included in the article, but all four response options are not – only the labels for the first and fourth response option as described above.

Situational judgment tests. Situational judgment tests are a unique way to measure teamwork skills. These tests are composed of detailed teamwork scenarios with multiple-choice responses. Individuals are asked to choose what they would be likely to do in the described situation or what might be the best response to the situation presented (Lai & Viering, 2012; National Research Council, 2011). A benefit of situational judgment tests is that they can be a more authentic measure of skills since they are created based on potential real-world experiences (Lai & Viering, 2012). They also require someone to

apply their knowledge and skills of teamwork to a new situation, as opposed to rating their skill level in a self-report survey. Similar to self-report surveys, faking can be an issue with situational judgment tests. Individuals may be able to figure out what the most socially desirable choice would be and fake their response (Lai & Viering, 2012; National Research Council, 2011). Additionally, the test requires reading comprehension and language fluency skills. If someone is lacking in these skills, confounding variables are introduced, making it difficult to know if an individual's incorrect response is because they lack the teamwork skill or if the item is instead measuring their language skills (Lai & Viering, 2012).

The Educational Testing Service's Teamwork and Collaboration assessment for high school students included a situational judgment test in addition to the self-rating survey and teacher-rating survey (Zhuang et al., 2008). The situational judgment test was composed of eight scenarios. For each scenario, students are given four different courses of action they could take in response to the scenario. They rate the effectiveness of each reaction to the situation on a five-point scale from very ineffective to very effective. Scenarios are then scored based on the rating they give the item that was considered to be the most effective course of action out of the four items for each scenario. Coefficient alpha was .71 for the situational judgment test. The full situational judgment test is freely available in the Zhuang et al. (2008) paper. The major limitation of the situational judgment test scenarios is that they are school-based, classroom situations, which means they are not relevant to scenarios youth would experience in STEM OST learning environments, making them a poor measure of those experiences.

There were also a few situational judgment tests to measure teamwork skills in adult populations. The Teamwork Knowledge, Skills, and Abilities Test is a 35-item situational judgment test developed to aid employers with hiring decisions and creating work teams (Stevens & Campion, 1999). The instrument was composed of 35 multiple-choice items. The article included some sample items and the survey is available for purchase online ($25 for the administration manual and $17 per survey). Since the test did not have any internal validity evidence around its dimensionality, Aguado et al. (2014) conducted a series of confirmatory factor analyses of the test and found that they were not able to find a model (using both

correlated and uncorrelated models) to adequately fit the five-factor model proposed by Stevens and Campion (1999). They also cited a number of studies that had used the test, but reported low reliabilities in the $\alpha = .60$ range and lower (Aguado et al., 2014). The international Adult Literacy and Life Skills Survey includes a 40-item situational judgment test to measure knowledge of teamwork skills (Baker et al., 2004). The test is composed of five work and non-work based scenarios that include a toy manufacturing team, marketing team, customer service team, and community-based teams. The publication only included one vignette as an example of the instrument, and it is unclear from the report and online searches if the survey was ever finalized and how one might obtain it. The two adult situational judgment test instruments would be inappropriate for measuring teamwork skills in the context of STEM OST programs because the scenarios are not relevant to the experiences youth have in STEM OST programs, and the instruments lack validity evidence for youth populations.

Ratings from others. Some researchers have created other-report instruments in order to triangulate findings with other measures, such as self-report surveys and situational judgment tests. Ratings from others include ratings from teachers or peers. Other ratings are beneficial because they can have less bias than self-report measures (Connelly & Ones, 2010). When used with self-reports, they provide a more holistic view of an individual's skills. These measures are also less likely to be influenced by faking than self-report measures (Lai & Viering, 2012).

There are, however, a number of limitations of ratings by others. Others' ratings are based on a limited number of interactions the rater may have with the individual. The rater does not always know the thoughts behind someone's actions and reasons for participating in the team in a particular way, and they may make incorrect assumptions based on the behaviors they are viewing (Kyllonen, 2012). Another potential limitation is that measurement error can come from raters who tend to rate people at certain points of the scale (low, mid, or high points of the scale) regardless of their skill level, leading to little variance in a rater's scores and the instrument measuring the attitude of the rater instead of the behavior

being observed (Bedwell et al., 2011). As with any rating measure, steps need to be taken to reduce subjectivity. It is important that the rating scale is clear and understandable to raters, training on how to use the instrument is provided if necessary, and the raters are committed to focusing only on the criteria laid out in the instrument (Lai & Viering, 2012).

Of the student instruments reviewed, one included an educator rating instrument. The Educational Testing Service's Teamwork and Collaboration assessment for high school students included a teacher-rating survey in addition to the student self-report survey and situational judgment test already described (Wang et al., 2009). The teacher-rating survey includes 10 behaviorally anchored rating scale items to measure teamwork skills. Each item had a five-point scale with behavioral labels for the 1, 3, and 5-point rating. An example item is "When working on a group goal or project, this student: (a) acts without regard to others' interests or suggests, (b) No label, (c) Attempts to adjust his/her actions to achieve the group goal, (d) No label, or (e) Takes the relevant and appropriate actions to achieve the group goal" (Zhuang et al., 2008, p. 49). The scale has high reliability ($\alpha = .98$), and exploratory factor analysis results showed that the scale was unidimensional. Although this survey is freely available in Zhuang et al. (2008), it only has validity evidence for formal education settings.

There is one other-rating instrument developed for measuring teamwork skills in adult populations. The Comprehensive Assessment of Team Member Effectiveness (CATME) Peer Evaluation is meant for college faculty to use with students who are working on teams in their classes (catme.org, 2013). The assessment includes both a self-evaluation and peer-evaluation for each individual on the team. The instrument uses behaviorally anchored rating scales to assess five areas of student performance on their classroom team. Data are then reported based on how students rated themselves, how their teammates rated them, and an average rating for the student and their team. The instrument went through an extensive development and validation process, including gathering confirmatory factor analysis and reliability data (Loughry et al., 2007). All five factors had reliabilities of $\alpha = .90$ or higher (Loughry et al.,

2007). The CATME is meant for college students within a college classroom context so it does not have the validity evidence necessary for use with youth populations in STEM OST programs.

Observations. Observations are similar to ratings by others as they both rely on observing teamwork behavior in action. Here observations are distinguished from other-reports as they are observations and ratings done by the evaluator, not educators or peers. They may also have more specific protocols to use while observing, while other-reports may rely on someone simply remembering what happened over a number of situations. Observations are a way to measure demonstration of skills during actual collaborative experiences. Observations are beneficial for a variety of reasons. They can help to explain the reasons people may have responded in a certain way to a survey or situational judgment test. Observations can be a more reliable measure than self-reports since there are not issues of satisficing and faking if individuals are unaware of the reason they are being observed. An additional benefit of observations is that the evaluator can understand the circumstances in which certain behaviors are present or absent for an individual, providing more detailed information than other means of measurement (National Research Council, 2011). A limitation of observations is that to get a holistic view of an individual's teamwork skills, an evaluator needs to observe him/her on multiple occasions and ideally in different teamwork situations, which can be time intensive and costly. If individuals happen to know the reason the evaluator is observing them, then issues of faking would need to be considered as they could affect the trustworthiness of what is being observed.

Of the instruments reviewed, one is an observational instrument for youth. The Human Achievement Quotient (HAQ) assessment is an observational checklist meant for formative assessment (MHA Labs, n.d.). The checklist includes six 21st century skill categories, including collaboration. The collaboration checklist includes six items to observe. Evaluators rate each of the items on a five-point rubric: (a) novice, (b) emerging, (c) capable, (d) skilled, and (e) expert. There is also an option to check that the skill was not observed. The assessment does not include information about the observation protocol including what age the instrument is meant for, how long someone should observe, or what kinds

of situations should be observed. However, the website refers to a use case guide for the instrument with examples of how it is used in various youth development programs, but it is not clear what kind of information is included in there about the actual protocol itself (Crown, Fine, & Beller, n.d.). There is also no psychometric information readily available for the assessment from a search online, although the assessment notes that in 2013-2014 a wide variety of validity evidence was going to be gathered in addition to reliability data including inter-rater reliability (Crown et al., n.d.). From the MHA Labs website it is also unclear how to obtain the instrument (MHA Labs, 2017).

An observational instrument for adults was also reviewed. The Center for Health Science Interprofessional Education, Research, and Practice developed three Performance Assessment for Communication and Teamwork (PACT) observational tools: a novice tool, expert tool, and video tool (Chiu, 2014). The tools were developed for use with pre-licensure health professional students and were composed of simulated scenarios carried out by the students. The novice tool focused on observing teamwork skills in a real-time scenario and used a rubric to rate performance across five behavioral markers. The expert tool focused on observing communication on teams in a real-time scenario and had rating on both the frequency and quality of behavior across 13 behavioral markers. The video tool focused on observing communication on teams based on a videotaped scenario and had ratings on both frequency and quality of behavior across 26 behavioral markers. Since the observational tool is meant for a specific health science student audience and are based on medical scenarios, it would not be an appropriate measure for STEM OST programs.

Appropriateness of Current Measures for STEM OST Programs

Nine measures of teamwork skills were reviewed in this paper. As illustrated in Table 6, only four of these instruments are appropriate for middle or high school audiences (a summary of the instruments reviewed for adults are in Table 7). Although these four instruments were tested with youth, they each have limitations that made them inadequate for STEM OST program evaluations. The Youth Outcomes Battery was only tested in one kind of informal education setting, so validity evidence is lacking for the

wide range of settings in which STEM OST programs take place (Sibthorp et al., 2013). Additionally, the instrument uses a pre-retrospective measure for looking at change over time, which might be fine for a weeklong camp, but may not be as reliable and valid of a measure of pre/post changes for STEM OST programs that last months or even years. There were also some concerns about the quality of the items. The Youth Experience Surveys were tested in OST programs and are freely available, making them possibly ideal for a STEM OST program (Hansen & Larson, 2002, 2005). However, the surveys do not actually measure skills. Instead, they measure exposure to experiences, and some of the items appear to be measuring outcomes in addition to experiences, calling into question the validity evidence of the scale for measuring teamwork skills. The Educational Testing Service's Teamwork and Collaboration instruments were developed for a formal education setting and include some items and scenarios that are specific to school experiences and not relevant to STEM OST programs (Wang et al., 2009). Additionally, all three of the Education Testing Service's instruments (self-report survey, situational judgment test, and teacher-ratings) have only been tested in formal education settings, meaning they lack validity evidence for use with STEM OST programs. The HAQ is an observational checklist that OST programs have used. However, information is lacking online about how to use the instrument, any validation work that has been done on it, and where to obtain the assessment, which makes STEM OST programs less likely to use it.

Table 6

Characteristics of the Youth Teamwork Instruments Reviewed

Name of instrument	Type of instrument	Intended age	Setting tested in	Cost
Youth Outcomes Battery (American Camp Association, 2013)	Self-report surveys	Middle and high school	Camp settings	$10 members/ $30 non-members
Youth Experience Survey 1.0 & 2.0 (Hansen & Larson, 2002, 2005)	Self-report survey	High school	Organized OST programs	Free
ETS Teamwork & Collaboration Assessment (Wang et al., 2009; Zhuang et al., 2008)	Situational judgment test, Self-report survey, Teacher-report survey	High school	Formal education setting	Free
Human Achievement Quotient (HAQ) Assessment (MHA Labs, n.d.)	Observation	Intended age is unclear	OST programs	Unclear how to obtain

Table 7

Characteristics of the Adult Teamwork Instruments Reviewed

Name of instrument	Type of instrument	Intended age	Setting tested in	Cost
Teamwork Competency Test (Aguado et al., 2014)	Self-report survey	Adults	Higher Education (Spain)	Unclear how to obtain
The Teamwork-KSA Test (Stevens & Campion, 1999)	Situational judgment test	Adults	Workplace setting	$17 per survey, $25 for manual
Adult Life & Literacy Skills Teamwork Framework (Baker et al., 2004)	Situational judgment test	Adults	Unclear if it was tested	Unclear how to obtain
Comprehensive Assessment of Team Member Effectiveness (Loughry et al., 2007)	Self-rating, Peer-rating	College students	Higher education	Free to college faculty
Performance Assessment for Communication and Teamwork observational tools (Chiu, 2014)	Observation	College students	Higher education	Free

Summary

Twenty-first century skills are important workforce readiness skills for today's youth. Their importance has been stressed at both the national and international level in both formal and informal education settings. Along with the emphasis on the importance of these skills, there has been a call for more measures to evaluate and research these skills. One of the skills that has been found to be "very important" for the workforce, more so even than the basic skills, is teamwork skills (Casner-Lotto & Barrington, 2006). In response to the importance of these skills and the national and international call for more measures, this chapter examined the ways teamwork skills are operationalized and how they are currently measured.

STEM OST programs offer fertile environments for developing the essential 21st century skills of teamwork. However, as evident from the literature review, the field lacks easily accessible and appropriate measures to evaluate teamwork skills in these programs. Research is necessary to first define the construct of teamwork skills so it aligns with the organizational psychology and 21st century skills literatures, but is also relevant to the outcomes of STEM OST programs and their evaluation needs. Instruments then need to be created that have the validity evidence necessary for use with the diverse participants and in the wide ranges of settings of STEM OST middle and high school programs.

Chapter 3: Methodology

As evident in the literature review, there was a need to create an instrument that evaluators can use to measure teamwork skills for middle and high school youth in STEM OST programs. This study developed such an instrument, guided by the following research questions.

1. What skill area of the broad range of teamwork skills best aligns with middle and high school youth outcomes in STEM OST programs and the evaluation needs of these programs?
2. Is the construct of team communication skills unidimensional or multidimensional?[3]
3. To what extent does the developed instrument gather reliable data from youth in STEM OST programs?
4. To what extent is there adequate validity evidence for the developed instrument?

As will be described in this chapter, the teamwork skill area of focus for the final instrument was team communication. The survey items were developed to measure the underlying construct of team communication skills. This meant various types of validity evidence needed to be collected in order to make the argument that the interpretation of scores on the survey were indeed a measure of someone's team communication skill level. The validity argument included construct validity evidence based on content, response process, and internal structure (American Educational Research Association et al., 2014; Messick, 1995). Content-related validity evidence was gathered by reviewing the literature and gathering feedback from STEM OST providers to ensure that the content of the survey aligned with the construct of team communication skills. Think-aloud interviews were carried out to gather response process validity evidence (Wilson, 2005). Internal structure validity evidence came from factor analysis and differential item functioning (DIF) analysis (American Educational Research Association et al., 2014; Wilson, 2005).

[3] This research question emerged after answering Research Question 1 (as described later in this chapter) by identifying team communication skills as the construct area of focus for the instrument.

The research questions were answered and validity evidence was gathered through a four-phase instrument development and validation process. Each phase built on the last leading to the final phase of a national field test of the instrument. The process ensured that the final instrument gathered reliable data and had validity evidence for use with the diverse middle and high school youth population that participates in STEM OST programs.

This research with human subjects was approved by the University of Minnesota's Institutional Review Board. Consent forms and letters are referred to in the dissertation and included in the appendices. Consent forms and letters sent home to parents and guardians of youth were made available in English, Spanish, and Hmong.

This chapter is laid out by phase (see Figure 1). The first three phases are described in their entirety with descriptions of the data collection method, sample, and findings that informed the next phase of instrument development and validation. The methodology of the fourth phase, the national field test, is described in this chapter, and the psychometric results of the field test are shared in Chapter 4.

Overview of Methodology

Phase 1: Defining the Construct
- Literature review.
- In-depth interviews with STEM OST program providers.

Phase 2: Item Development and Expert Review
- Development of test blueprint and items.
- Expert review by STEM OST program providers.
- Revisions to items.
- Expert review by grant advisors in teamwork science, measurement, and youth development evaluation.
- Revisions to items.

Phase 3: Piloting and Revisions
- Iterative cycle of think-aloud interviews with youth and revisions to items.
- Pilot test with middle and high school youth in STEM OST programs.
- Conduct exploratory factor analysis, item analysis, descriptive statistics, and reliability.
- Revision and deletion of items.

Phase 4: Field Test
- Field test with middle and high school youth in STEM OST programs.
- Conduct confirmatory factor analysis, reliability, item analysis, descriptive statistics, and DIF analysis.
- Finalize instrument.

Figure 1. Overview of methodology in each phase of the research.

Phase 1: Defining the Construct

Phase 1 focused on identifying and operationalizing the teamwork skill area that would be the focus of the instrument. During this phase, data were gathered from STEM OST programs to understand what teamwork skills were important within the context of their middle and high school programs. This work was informed by the literature review described in Chapter 2 to ensure the skill area aligned with the literature and was relevant to youth outcomes in STEM OST programs and their evaluation needs.

Method. Data collection methods in Phase 1 included interviews and content analysis. In-depth semi-structured interviews were conducted with STEM OST professionals to understand how they operationalized teamwork skills in their programs. The interview protocol was piloted with a pair of staff

from the Science Museum of Minnesota's Kitty Andersen Youth Science Center. As a result of the pilot, some questions were added, removed, or changed to better reflect teamwork processes in STEM OST programs and better align with the language used in these programs (see Appendix A for the final interview protocol). Interviews took place in-person or over the phone and lasted up to an hour. Prior to participating in an interview, interviewees completed a consent form (see Appendix B for the consent form). All interviews were audio recorded and transcribed for analysis. During the interviews, programs were asked if they had any supplemental information they could share. Programs provided a variety of documents, including evaluation tools and reports, logic models, annual reports, and curriculum. Additional relevant information was also downloaded from programs' websites.

Sample. The goal for sampling was to gather data from a diverse range of programs to look at how the construct of teamwork skills was operationalized in middle and high school STEM OST programs. To reach this goal, a maximum variation sampling technique (Patton, 2002) was used, looking for variation across STEM topic areas, types of programs (camps, internship programs, youth development programs, clubs, etc.), locations of programs across the United States, program setting (library, museum, community site, university, school, etc.), and grades reached. In terms of grade ranges, middle school was considered 5^{th} – 8^{th} grade and high school was 9^{th} – 12^{th} grade. While sampling, an effort was also made to ensure that similar kinds of programs were not selected. For instance, if the sampling frame included various Lego robotics programs, only one was selected.

A wide variety of recruitment strategies were used to develop the sampling frame. Emails were sent to individuals who expressed interest in measuring teamwork skills in their STEM OST program on a needs assessment survey that was administered in preparation for the National Science Foundation proposal for this project (Grack Nelson, 2013). Information was shared about the study through a variety of listservs, discussion boards, blogs, forums, and Facebook pages (see Appendix C for a list of the places recruitment information was shared during each phase of the research). Individuals who participated in an interview were also asked to share information with their own afterschool networks and contacts.

Individuals expressed their interest in participating in an interview by filling out an online interest form where they described their program, how they used teams in their program, and if they have goals or outcomes related to teamwork skills. This information was used to inform sampling decisions as well as provide background information about individuals who ended up being interviewed. The sampling frame consisted of 83 programs.

Of the 83 programs that completed an interest form, 34 programs were chosen for the final sample. Of these, 12 were interviewed about their middle school programs, 10 were interviewed about their high school programs, and 12 were interviewed about both their middle and high school programs, although these interviews were analyzed separately by middle and high school. This resulted in data from 22 middle school programs and 24 high school programs. Programs were located in Washington, DC and across 16 states, including California, Colorado, Connecticut, Florida, Georgia, Illinois, Indiana, Minnesota, North Carolina, North Dakota, New York, Ohio, Pennsylvania, Tennessee, Utah, and Washington. Experts worked at a wide range of organizations, including community-based organizations, an aquarium, an arboretum, a zoo, a research institute, schools, science museums, natural history museums, 4-H, a library, and universities. Some experts came from programs that covered a variety of STEM topics, while others came from programs focused on specific areas of STEM, including aquatic science, astronomy, biotechnology, design, energy, engineering, environmental science, conservation science, making[4], microbiology, and robotics. STEM OST program types included camps, clubs, volunteer programs, 4-H, youth development programs, youth employment programs, internships, research experience programs, afterschool classes, apprenticeships, workshops, drop-in programs, leadership programs, competition-type programs (robotics or design), and more.

[4] Making activities are at "the intersection of computer science, design, art, and engineering" (Rosenfeld Halverson & Sheridan, 2014, p. 501).

Discussion. All of the data (educator interest forms, interviews, program website information, and documents provided by interviewees) were uploaded, coded, and analyzed in NVivo 10. The data were coded based on a coding framework developed from the working list of teamwork skills and indicators identified in the literature (see Appendix D for the coding framework). The coding framework included the following list of teamwork skills:

- Leadership.
- Communication.
- Task-related assertiveness.
- Goal setting.
- Mission analysis.
- Situational analysis.
- Shared responsibility.
- Coordination, project management, and planning.
- Problem detection.
- Problem solving/decision making.
- Flexibility/adaptability.
- Mutual performance monitoring/intrateam feedback.
- Backup/supportive behavior.
- Motivation of others.
- Value individual contributions.
- Conflict resolution/negotiation.
- Work with diverse teams.
- Empathy/perspective taking.
- Professionalism.

As the data were coded, two additional teamwork-related skills emerged that were not present in the literature. These included the skills of "respecting others" on the team and "self-efficacy," where self-efficacy was related to youth recognizing their own strengths and what they can contribute to a team. These two skills were added to the coding framework, along with an "other" code for skills that did not fall into any of the codes in the framework.

As illustrated in Table 8, all of the teamwork skills identified in the literature review were addressed by at least one of the programs in the sample. The most common skill was communication on teams, mentioned by most of the programs (85%). Over half of the programs also discussed shared responsibility and leadership.

Table 8

Teamwork Skills Addressed by Interviewed STEM OST Programs (n = 34)

Teamwork skill	Percentage of programs
Communication	85%
Shared responsibility	59%
Leadership	56%
Task-related assertiveness	44%
Coordination, project management, and planning	44%
Value individual contributions	44%
Problem solving/decision making	41%
Motivation of others	41%
Respect others	32%
Mutual performance monitoring/intrateam feedback	29%
Work with diverse teams	29%
Backup/supportive behavior	24%
Flexibility/adaptability	18%
Other	15%
Conflict resolution/negotiation	12%
Professionalism	12%
Self-efficacy	12%
Goal setting	9%
Empathy/perspective taking	9%
Problem detection	6%
Mission analysis	3%
Situational analysis	3%

Overall, middle and high school programs addressed similar teamwork skills. As illustrated in Table 9, most teamwork skills had similar occurrence in both the middle and high school data, with communication skills most frequently mentioned. The largest differences were middle school programs placing more of an emphasis on the teamwork skills of problem solving/decision making and goal setting, while high school programs placed more of an emphasis on the teamwork skills of coordination, project management, and planning.

Table 9

Teamwork Skills Addressed by Interviewed STEM OST Programs by Grade Range

Teamwork skill	Middle school (n = 24)	High school (n = 22)
Communication	75%	91%
Shared responsibility	54%	55%
Problem solving/decision making	50%	23%
Leadership	46%	55%
Value individual contributions	46%	36%
Motivation of others	42%	36%
Task-related assertiveness	38%	45%
Respect others	29%	23%
Mutual performance monitoring/intrateam feedback	29%	23%
Backup/supportive behavior	25%	27%
Coordination, project management, and planning	21%	50%
Work with diverse teams	21%	36%
Flexibility/adaptability	17%	14%
Other	13%	14%
Goal setting	13%	0%
Conflict resolution/negotiation	8%	9%
Professionalism	8%	14%
Self-efficacy	8%	14%
Empathy/perspective taking	8%	14%
Problem detection	4%	5%
Situational analysis	4%	0%
Mission analysis	0%	5%

As was evident from the literature and results from STEM OST programs, a broad range of teamwork skills were important for STEM OST programs. Many of the individual skills could be their own construct and thus scale (for instance, team leadership skills or problem solving/decision making in a team). To focus instrument development on one skill area within the larger construct of teamwork skills, the construct of communication in teams was selected for the instrument because communication was the skill most frequently addressed by both middle and high school STEM OST programs.

The next step was to operationalize the construct of team communication skills. To do so, the STEM OST interviews and literature were revisited. The literature was reviewed to gain a deeper understanding of team communication skills and how they were measured. Interview data were coded in NVivo based on skills that aligned with how the literature talked about team communication skills, while also allowing for the emergence of skills that were not mentioned in the literature. This led to three areas of focus for team communication skills: (a) task-related assertiveness, (b) closed-loop communication, and (c) listening. These areas are defined in detail below, pulling from both the interviews and the literature. For the purpose of the instrument, the construct of team communication skills did not include a number of general communication skills such as public speaking, reading, writing, and communicating in a foreign language.

Task-related assertiveness. Task-related assertiveness was defined as asserting oneself by speaking up and sharing ideas, opinions, and knowledge with teammates. Many STEM OST providers talked about the desire for youth to be able to communicate their ideas or express their opinions with their team, even if they have a different opinion or idea than someone else. Some STEM OST providers stressed the need for youth to not be shy and speak up in an assertive manner. This was echoed in the literature, which talked about sharing ideas and opinions related to the task at hand, with an emphasis on doing this openly and without hesitation (Baker et al., 2004; Organization for Economic Cooperation and Development, 2005; Pearsall & Ellis, 2006; Salas et al., 2009). As a result of the grant advisor expert

review described in Phase 2, this focus area was later replaced with the team communication skill area of "information exchange."

Closed-loop communication. The skill area of closed-loop communication was defined primarily from the literature. The STEM OST providers spoke generally about youth being able to communicate ideas clearly and ask questions, but the literature provided more specificity to help operationalize the construct area. Closed-loop communication is the process between the sender and receiver of a message to make sure a message is communicated, received, and understood (Johnson & Johnson, 2013; McIntyre & Salas, 1995; Salas et al., 2009). Overall effective communication, as described by Johnson and Johnson (2013, p. 130), "exists among group members when the receivers interpret the sender's message in the same way the sender intended it." During the process of closed-loop communication, the sender ensures that the other person received the message, interpreted it correctly, and encourages the receiver to clarify understanding by repeating back what they heard and asking clarifying questions (Johnson & Johnson, 2013; McIntyre & Salas, 1995; Rosen et al., 2013; Salas et al., 2009). The receiver acknowledges they have received the information, repeats back what they heard, and, if necessary, asks clarifying questions to make sure they fully understand what is being communicated (Dickinson & McIntyre, 1997; McIntyre & Salas, 1995; Salas et al., 2009). At times, it may be necessary for the sender and receiver to negotiate meaning of the information being communicated when it seems that the receiver may not have interpreted the information or message correctly (Johnson & Johnson, 2013). Some definitions of closed-loop communication also talked about team members providing summaries to one another related to what the team has discussed (Aguado et al., 2014; Rosen et al., 2013).

Listening. Listening skills emerged as another area of importance related to team communication. As described by Baker et al. (2004), the ability to listen effectively is an important part of strong team communication skills. Active listening is key, and on a team this means focused listening to the ideas and information discussed by individual team members (Organization for Economic Cooperation and Development, 2005; Stevens & Campion, 1994). One STEM OST provider described listening as "this

idea of one mic, which is like this idea that whoever is talking is the focal point and all of your attention is directed there." Another provider talked about youth needing to know when to listen and the importance of listening to teammates: "I think the biggest skill is communication and knowing when to listen, because a lot of times nobody wants to listen. They don't want to listen to people in their group, they don't want to listen to their group leaders, they just want to do their own thing." Another provider talked about what it means to be an active listener and engage with what you are hearing.

> Learning how to actually listen when someone is giving their opinion or idea rather than just waiting for their turn to talk … Helping them figure out what that difference is and how to listen and give feedback not just listen and then move on to what you want to say or what they want to say.

This idea relates to the skill of balancing listening and speaking, which was stressed in both the interviews and literature (Greenstein, 2012). This also relates to refraining from interrupting teammates, something brought up by a few STEM OST providers.

Phase 2: Item Development

Phase 2 included development of the blueprint and items; expert review with STEM OST providers; and expert review with experts in youth development evaluation, teamwork science, and measurement. After each of the expert reviews, there was refinement of the construct and instrument.

After the construct of team communication skills was defined, a test blueprint was created. The test blueprint served as a guide for item development. The items were developed by looking at each skill area's interview data, literature definitions, and items from other measures with an effort to make sure the items encompassed the range of skills that made up each skill area. The initial plan was to create items to equally cover the three areas. As the items were developed, it became clear that some skill areas needed more items than others to help ensure that each skill area was sufficiently represented. The first version of the instrument ended up with 32 task-related assertiveness items, 19 close-looped communication items, and 9 listening items (see Table 10). Listening had fewer items because it was difficult to create self-

report items to measure listening without the items having an obvious "right" answer. All of the items were self-report rating scale items. This first version of the instrument was created for use with both middle and high school youth with the idea that testing through the think-alouds and pilot test would indicate if separate middle and high school versions of the instrument were needed.

Table 10

Test Blueprint for Team Communication Skills Instrument Version 1

Skill area	Number of items	Percent of total items	Item type	Grade range
Task-related assertiveness	32	53%	Self-report scale item	Middle & high school
Closed-loop communication	19	32%	Self-report scale item	Middle & high school
Listening	9	15%	Self-report scale item	Middle & high school

The instrument had to be developed so it could work for a wide variety of programs. Since STEM OST programs differ widely in terms of their content and activities, a context needed to be created for the items that would span a wide range of programs. Thinking about the teamwork skill instruments reviewed in the literature review, the team communication instrument ended up being a cross between a situational judgment test and a self-report survey. Like a situational judgment test, the instrument had a teamwork situation, or scenario, that youth were asked to imagine themselves in while responding to the items. The teamwork scenario was created to be general enough that youth in a wide variety of programs could see themselves in the scenario. To help ensure this, the scenario was created based on how team experiences were described in the Phase 1 STEM OST provider interviews. The scenario provided a similar context for all youth to keep in mind when responding to the survey items.

Unlike a situational judgment test, the assessment did not use multiple choice items with a correct answer for the behavior youth should take. Instead, the instrument used self-report items related to the scenario. This choice was made because it was important that the assessment matched the informal learning environment and did not feel like a test (National Research Council, 2009b). However, similar to

a situational judgment test, there was a series of items that measured behavioral tendencies or the likelihood that youth would perform a behavior in the proposed scenario (McDaniel, Hartman, Whetzel, & Grubb, 2007). The instrument was not a direct measure of team communication skills, but instead was composed of self-report items that measured youths' perceived comfort level of doing a skill, their likelihood of exhibiting a behavior related to that skill, and their knowledge of appropriate behaviors or demonstration of a skill. All items had a six-point scale to capture the range of variability in youths' responses.

Phase 2: STEM OST Provider Expert Review

The second activity of Phase 2 involved collecting content-related validity evidence from STEM OST providers since they were experts in what takes place on the ground in STEM OST programs. Their feedback was important to help ensure that the survey's scenario and items were relevant to the way teams were used in STEM OST programs, the team communication skills addressed in these programs, and programs' evaluation needs. STEM OST providers' feedback was also a first step to make sure the questions were appropriate for middle and high school youth.

Method. STEM OST providers were sent a document with review instructions and questions embedded within Version 1 of the team communication skills survey (see Appendix E). They were also sent a summary of findings from the Phase 1 STEM OST provider interviews to provide context for the decision to focus on team communication skills. Providers were instructed to focus on either their middle or high school programs since some programs reached both middle and high school youth. They were given two weeks to complete the review and send back the completed review document. Upon completion of review, they were given a $25 VISA prepaid gift card in appreciation for their time

The review asked for feedback on the scenario, the items, and programs' potential use of the survey. Questions about the scenario focused on general comments about the scenario as well as relevance of the scenario to their program. Providers were asked about the relevance of the skills described in the items to the team communication skills and youth in their program. In addition to

reviewing the survey, providers were asked to comment on the usefulness of the survey for evaluation since they would be one set of intended users.

Sample. The 34 STEM OST providers who participated in the interviews during Phase 1 were invited to participate in the expert review. Of the 34 providers who participated in Phase 1, fifteen agreed to participate in the expert review. Of the 15 who agreed to participate, 11 completed the review. The final expert review sample included six middle school and five high school program providers. They were from programs in Washington DC, and across the United States in California, Georgia, Idaho, Indiana, Minnesota, North Carolina, North Dakota, and Utah. Experts worked at a wide range of organizations, including a community-based organization, 4-H, a historic home, a research institute, a school, science museums, a library, and a university. Some experts came from programs that covered a variety of STEM topics, while others came from programs focused on specific areas of STEM, including astronomy, engineering, environmental science, making, microbiology, and robotics. STEM OST program types included camps, clubs, a teen volunteer program, 4-H, youth development programs, a youth employment program, research experience programs, afterschool classes, apprenticeships, drop-in programs, competition-type programs (robotics or design), and general afterschool programs.

Discussion

Changes to the scenario. The first set of expert review questions were about the scenario (see scenario in Appendix E). STEM OST providers were asked if the scenario seemed like something that would happen in their program and, if not, what they would change about the scenario. They were also asked if the scenario was missing anything important. Many of the providers said the scenario was similar to their program, but they had a number of suggested changes that were incorporated to help ensure that the scenario was relevant for a wide range of STEM OST programs.

- The group size was changed from five to three youth. This change was based on a review of the 34 STEM OST provider interviews from Phase 1. The most common team size for both middle

and high school programs was teams of three (74% of the STEM OST interview sample used teams of three).

- Instead of having all of the youth on the imaginary team in the same grade, the scenario was changed so that teams could be youth of mixed ages in either middle or high school to be more inclusive of programs that use mixed grade teams.
- Since STEM OST programs included those that were not considered afterschool, the reference to afterschool was removed from the scenario.
- The original scenario said the team worked together over a few weeks, which was not the case for a variety of STEM OST programs. The language was changed so that the team experience did not have a given timeframe. Instead, the scenario focused on the process of the work the team was doing together. The assumption was that keeping the scenario more general would help youth see themselves in the scenario and put their own timeframe to the scenario so that it would work for one-day to multi-day team experiences.
- The original scenario said, "Solve a challenge," and one reviewer asked why the scenario had youth solving a challenge instead of solving a problem. To inform revisions to the language, the STEM OST interviews were reviewed to see how programs talked about challenges. This brought to light that of the 34 STEM OST programs that were interviewed, nine had team challenges, 25 had team projects, and four had some other kind of team activity (four programs had both challenges and projects, which is why these numbers add to over 34). None of them talked about solving a problem. For this reason, the language in the scenario and items were changed from thinking about solving a challenge to instead thinking about completing a "challenge" or "project" with the idea that programs could select the term that fit best for their program.
- An icebreaker activity was added to the scenario to address suggestions from a number of reviewers. The widespread use of "get to know you" activities was also evident after reviewing

the STEM OST interviews and seeing that 29 of 31 programs (three of the 34 interviewees were not asked a question about icebreakers) included some kind of team builder or icebreaker activity.

- To address the fact that some programs did not have access to computers or that youth were not allowed to use them, researching information was expanded to include not only looking online, but also reading books and magazines or using information provided by the program (something one of the reviewers said they did in their program).
- The scenario was rewritten so it was clear that each team member was participating and contributing to the team as well as checking in with each other throughout the process, something suggested by one of the reviewers.

Some of the specifics suggested by the reviewers were not added to the scenario, mainly because they did not seem to be typical of a wide range of STEM OST programs and the idea was to keep the scenario as broad as possible so many programs could see themselves represented in it. For instance, one reviewer mentioned adding the use of roles on teams. However, when looking at the STEM OST interview data, the use of roles, how roles were assigned, and if they were even used varied widely among programs, which would have made it difficult to create a general scenario with roles. Another suggestion was to add an adult staff person to the scenario. In the STEM OST interview data, this varied for programs so details about an adults' role was not included. Another recommendation was the use of a presentation at the end of the project or challenge. Presenting to a group was also more of a general communication skill instead of a team specific communication skill. This was similar to suggestions about including written communication skills for both the scenario and items, which were more general communication skills and not teamwork specific. Also, the scenario had youth working together in person so there would not be the need for written communication in their particular interactions. There was a suggestion to add that the program was a safe space to express ideas with zero tolerance for bullying. The scenario states that youth should imagine themselves in their own STEM OST program, so if a program

has made those values and norms explicit in their program, the assumption was that youth would keep that in mind when thinking about the scenario and responding to the items. Finally, there was a suggestion to include conflict in the scenario. Conflict resolution was considered a separate 21st century skill and not specific to team communication skills, so it was not added to the scenario or items.

Changes to individual items. STEM OST providers were asked about the difficulty of the items and relevance of the skills described in the items to the team communication skills and youth in their program. The following changes were made to items based on providers' feedback.

- When asked about the difficulty of the items, all of the high school providers and some of the middle school providers felt items would not be difficult for youth in their program to understand. Some of the middle school providers thought youth might have difficulty distinguishing between ideas and opinions, since both terms were used in the survey. One high school reviewer did not think the items were difficult, but mentioned later in the review that "a description of the difference between ideas and opinions might be helpful" (ID37). Because of these reviewers' comments, items with the word "opinion" were removed or reworded to avoid confusion between the terms opinion and idea. When rewording, the word "opinion" was replaced with "idea" or "what you think."
- For Item 9 (Listening to a teammate's opinions before evaluating their positions as good or bad), the wording was changed from "good" or "bad" to "strong" or "weak" because of a comment from a reviewer that in their program, they "prefer to use 'strong' or 'weak.' 'Good' and 'bad' connect to moral meanings" (ID3). The term "evaluating" was also changed to "deciding" per a different reviewer's suggestion.
- Some reviewers mentioned a number of topics they felt were missing in the items. These included feedback, support, and conflict resolution. A few items were added that got at communicating

around help. However, feedback, support, and conflict resolution were all considered distinct teamwork skills or constructs in the literature, so they were not added to the survey.

Some suggestions from reviewers were not incorporated into the survey.

- One reviewer mentioned that they did not have computers at their site, so Item 15 (Share with your team information you found about the topic of the challenge) did not apply since they would not be able to conduct research. As mentioned earlier, the scenario was changed so that youth could find information various ways, other than on the computer. The language for Item 15 stayed the same since the change in the scenario addressed the reviewer's comment.
- Reviewer 15 had a number of comments about items saying they were skills that youth cannot do, are bad at, or should not be expected to do. This reviewer had a view of youth that does not align with a positive youth development way of thinking that many STEM OST programs follow. Additionally, Reviewer 15 said that active listening was not taught, but that item came directly from interviews with STEM OST providers from Phase 1 of the study. When the STEM OST providers were asked if the statements aligned with skills they work on in their program, Reviewer 15 was the only one to state that there were items that did not align. Items were not changed or removed based on the comments from this reviewer since his/her views of what youth could not do did not align with the literature or other reviewers' way of thinking about the possibilities of youth's abilities.
- One of the providers felt that Item 52 (Asking a teammate for their opinion if they are quiet and not speaking up) seemed more like a leadership skill and something only the leader of the team would do. This item was kept in for the expert review with grant advisors to see if they felt the item fell under the team communication skill area of task-related assertiveness or if it was some other skill as suggested by the reviewer.

- One middle school provider commented that a number of the items related to youth asking their teammates questions to clarify ideas (Items 5, 6, and 7) were things that middle school youth would not do, but the adult in the program would. These items were kept in since they were important parts of the closed-loop communication skill area. If the instrument ended up having sub-scales with closed-loop communication being one of them, the program the reviewer was a part of could decide not to use this sub-scale with their program if it was not relevant to how youth communicate in the program's teams.
- Some reviewers had suggestions for additional questions to add to the survey that were not directly related to the construct being measured by the items. This included questions such as a youth's experience in the program, their prior experience with teams, and how they can improve their team communication skills. These questions were not added since they did not relate directly to creation of a scale to measure the construct of communication on teams. Instead, these are explanatory questions that programs can add to their survey if they would like that information for their own evaluation purposes.

Some comments that were not addressed were tested in later phases of the project.

- A few reviewers brought up cultural issues around the items. "The statements in general may not be difficult, however the majority of our kids are Latino. In general, they tend to be introverts or shy to share ideas. Although we encourage sharing ideas, listening to other's ideas, and providing constructive feedback, most of our kids (grades 5-9) will struggle in achieving these goals" (ID 42). Think-alouds and DIF analyses later in the instrument development process helped to ensure that the items were not biased based on a youth's race or ethnicity.
- One of the reviewers stated a concern that youth in the South "are taught to 'please' others, so they don't say things that are contrary to someone else's opinion, don't disagree, and so on"

(ID15). This was tested through the think-alouds where programs were specifically recruited from southern states.

Potential usefulness of survey. Reviewers were asked about the usefulness of the survey for their evaluation purposes. Most reviewers (64%) felt they would use the survey with their STEM OST program, while over a third (36% maybe use) felt they would maybe use it. The seven providers who said they would use the survey were asked to indicate how they might use it to evaluate their program. They were provided with the options to use it as a pre-, mid-, and/or pre-post measure. They all said they would use it as a pre-/post- and one said they would also use it as a mid-point measure. One provider said they would also use some of the survey questions "as part of the application process to assist with the selection of program participants" (ID3).

STEM OST providers that would use the survey were asked to explain what about the survey made it useful for their program. In addition to measuring impact, providers talked about how the survey could help inform programming. Two providers said they would use results to plan activities and help ensure program impact.

> Everything you've included [in the survey] is something that we value about teamwork. I think it would help us assess where to place more emphasis or time on at the beginning of the program when reviewing expectations and training our volunteers with how to facilitate teamwork among the teens they are working with. It's also a good indication, when done pre-and post, what they have taken away from their experience. (ID3)

One provider talked about using the skills covered in the survey as a kind of checklist. "It is nice to have a list of communication skills to work on with the kids. We can refer to the list to make sure that all of the skills are addressed" (ID141). This value of the survey for program providers was echoed by another reviewer: "I learned a lot about the skills involved in teamwork from this document" (ID 37).

The STEM OST providers that said "maybe" were asked why it might not be useful. One provider was concerned about the time it would take to complete because they already had a pre-/post-

survey for their program. Another provider said they already evaluate teamwork skills using a different tool. One provider was worried about the items that were problem behaviors associated with team communication. "I worry that they would alter their actions negatively based on the survey not realizing that some of the questions pose problems with teamwork" (ID 20).

Phase 2: Grant Advisor Expert Review

The third activity of Phase 2 involved collecting content-related validity evidence from the advisors for the National Science Foundation grant funding this work. The advisory board was composed of three experts: (a) Dr. Eduardo Salas, a professor at Rice University, contributed his expertise in teamwork science, (b) Dr. Daniel Bolt, a professor at University of Wisconsin – Madison, contributed his expertise in educational measurement; and (c) Beki Sato, Senior Research Associate at Rainbow Research, contributed her expertise in youth development evaluation. The advisors' feedback was important for gathering evidence related to the construct of interest (team communication skills) and the content (scenario and items) being used to measure that construct.

Method. The grant advisor expert review involved advisors reviewing the construct and instrument and then coming together for a two-hour conference call to discuss the review results. For the review, advisors were provided with an overview of the construct definitions and a worksheet to gather their feedback on aspects of the instrument and its alignment with the construct. The worksheet was informed by Newman, Lim, and Pineda's (2013) content validation process that uses a table of specifications for an instrument to judge the alignment of each item to the various aspects of a construct. Their process uses a mixed-methods approach where reviewers indicate the extent to which an item is sufficiently measuring a skill and then provide qualitative feedback for ratings. The grant advisor expert review instrument, which used a table of specifications approach to gain feedback on Version 2 of the survey, can be found in Appendix F. Advisors indicated if an item aligned, did not align, or kind of aligned with a particular skill area. They were also asked to provide comments or suggestions on the alignment or wording of the individual item, particularly thinking about if the way the item was phrased

was adequate to get at the particular skill area that was being measured and what changes could be made to an item to better measure that area of the skill. In addition to reviewing the table of specifications, advisors were also asked for feedback about the overall scenario and its ability to provide effective grounding for the survey items, if the items for a particular skill area (such as close-loop communication) seemed to fully represent the skill area and if not what was missing from the items, if there was anything missing from the definitions of the individual skill areas and overall construct of team communication skills, and overall comments about the instrument related to their area of expertise.

Discussion.

Feedback on the scenario. The advisors had a number of suggestions about the scenario. The addition of a team leader and specialized roles for youth were suggested for the scenario's team. In the STEM OST provider interviews the teams did not always have a leader or assigned roles so adding them to the scenario would make it less applicable to some STEM OST programs. For this reason, these changes were not made to the scenario. Another suggestion was around the addition of task interdependency. The following sentences were added to the scenario to stress the interdependency of the team's project or challenge. "Team members work on tasks both together and alone. The final [challenge/project] depends on everyone's contributions so team members are constantly checking in with each other to make sure the team is on track to reach their goal." A final suggestion for the scenario was to make it more problem-oriented. The scenario had been written as completing a challenge or project after suggestions from the STEM OST provider reviews and revisiting the STEM OST interview data where 25 of the 34 STEM OST programs described their team activities as projects and nine described them as challenges. For this reason, the language of solving a problem was not added to the scenario.

Feedback on the construct. The advisor discussion around the three skill areas of the team communication construct (task-related assertiveness, closed-loop communication, and listening) focused mostly around clarifying and revising construct areas. The teamwork science expert said that task-related assertiveness was seldom measured and it was more about psychological safety. There was a discussion

and agreement to remove this skill area from the construct of team communication skills. When discussing closed-loop communication, the teamwork science expert said some of the items were actually closer to the skill area of information exchange, an area that was missing from the definition of team communication skills. The teamwork science expert provided feedback in the expert review worksheet to indicate which items in the three different skill areas actually fell more closely under information exchange instead of where they had been aligned in the table of specifications. As a result of the advisor expert review, the skill area of task-related assertiveness was removed and the skill area of information exchange was added.

The literature and STEM OST provider interview data were reviewed to define the skill area of information exchange and develop related items. Per clarification from the teamwork science expert and the literature, much of what was previously part of the task-related assertiveness skill area was moved to the definition of information exchange. As defined by McIntyre and Salas (1995, p. 25), "communication refers to the exchange of information between a sender and a receiver." The exchange of information and ideas is key for a team to effectively work together toward a goal (Aube, Brunelle, & Rousseau, 2014). Team members each have important knowledge and ideas, and part of the skill is knowing when to share it, what is important to share, and doing so without being asked (Salas, 2013; Smith-Jentsch, Cannon-Bowers, Tannenbaum, & Salas, 2008; Smith-Jentsch, Johnston, & Payne, 1998). As stated by Smith-Jentsch, Zeisig, Acton, and McPherson (1998, p. 287), "information exchange includes when and what to communicate." Of particular emphasis was recognizing and sharing "unique" information with the team (Salas, 2013). The ability for youth to share and explain their ideas was also a key skill brought up by close to half (44%) of the STEM OST providers during their interviews. Many STEM OST providers talked about the desire for youth to be able to communicate their ideas or express their opinions with their team, even if they had a different opinion or idea than someone else. This can be a difficult skill as sharing an idea can "pose a risk because of potential errors and possible negative reactions to those ideas" (Gong, Kim, Lee, & Zhu, 2013, p. 830). An additional information exchange skill discussed in the

literature was the ability to provide the team with "big picture" updates or summaries to help make sure everyone was on the same page, something that was also part of the definition of closed-loop communication (Smith-Jentsch et al., 2008). This was only one instance where there was overlap between the information exchange and closed-loop communication definitions. In some cases closed-loop communication was actually included in the definition of information exchange, which suggests that they may not be distinct construct areas (American Institutes for Research, n.d.; Salas, 2013).

There was some discussion about the inclusion of listening skills as part of the team communication skills construct. The teamwork science expert felt the listening items should not be included unless they are related to information exchange or the teamwork skill area of mutual support. After going back to the literature, most of the definitions around listening came from the 21st century skills literature that focused on youth teamwork skills and not the organizational psychology literature that focused on adult skills (and the teamwork science expert's area of expertise). Additionally, when looking at the STEM OST provider interviews, almost half the providers (47%) said listening skills were important on their teams. The decision was made to keep listening as part of the construct of team communication skills since it was part of the 21st century skills literature and aligns with what was important for STEM OST programs.

Feedback on the overall instrument. The advisors talked about some of the overall features of the items on the survey, specifically the three question types (likelihood, comfort, and agreement), the response options, and the use of reverse scored items. The measurement expert felt the agreement items might have less variability compared to the other two kinds of items since it might be more obvious what the "right" answer is. He talked about the possibility of writing items about comfort and likelihood using the same item stem, which would allow for parallelism of the items. He said being able to pair likely and comfort allows someone to tell a richer story when looking at differences between pre-/post-. The advisors discussed the use of six points on the scale, questioning whether six points were necessary. After consideration, it was decided a four-point scale was sufficient. The advisors discussed the benefits and

limitations of the reverse scored items. The measurement expert stressed the challenges they present in factor analysis and reporting scores for a scale. Based on this recommendation, the reverse scored items were rewritten or removed from the instrument. Since most of the agreement items were reverse scored, the agreement scale was removed.

Based on the revised construct and feedback from advisors, the items and response options were revised. A total of 10 item stems were written for each of the three skill areas. For the information exchange and closed-loop communication items, an anchor statement was written with three questions following each statement. The questions asked about youth's perceived skill level related to the skill described in the statement, how comfortable or uncomfortable they would be doing that skill, and the likelihood they would do it. For the listening items, youth were asked how easy or hard it was to do that skill. Table 11 includes the revised test blueprint for Version 3 of the survey (see Appendix G for the think-aloud instrument that includes Version 3 of the instrument).

Table 11

Test Blueprint for Team Communication Skills Instrument Version 3

Skill area	Number of items	Percent of total items	Response options	Grade
Closed-loop communication	30 (10 for each response option)	43%	Bad/Good Uncomfortable/Comfortable Unlikely/Likely	Middle & high school
Information exchange	30 (10 for each response option)	43%	Bad/Good Uncomfortable/Comfortable Unlikely/Likely	Middle & high school
Listening	10	14%	Hard/Easy	Middle & high school

Phase 3: Think-aloud Interviews with Youth

Phase 3 began with think-aloud, or cognitive, interviews to gather response process validity evidence (American Educational Research Association et al., 2014). During a think-aloud interview,

individuals respond to survey questions aloud, verbally describing the thinking process behind each response (Beatty & Willis, 2007). This method allowed researchers to identify if items were interpreted correctly and measuring what was intended. Data from think-alouds helped to inform decisions on how to increase the validity and reliability of responses through item revisions and possibly construct revisions.

Method. During Phase 3, the instrument was revised through an iterative process of conducting think-aloud interviews, making revisions, and carrying out more think-aloud interviews. The think-aloud interview process tested the survey instructions, scenario, response options, and individual items (see Appendix G for the think-aloud instrument that includes Version 3 of the survey). The interview began by describing and modeling the think-aloud process to the youth. The youth then read the survey's scenario out loud, which brought to light if youth had problems reading the phrasing or any words in the scenario. Follow up questions were asked about the scenario before the youth began to think out loud about the items. The youth read each item out loud and then verbalized how they thought through their answers to the question. Youth were also encouraged to indicate what they did not understand about an item or why they might have been unsure about how to answer a question. The interview also included questions to gauge the youth's comprehension of the response options used throughout the survey.

Think-aloud interviews took place in two rounds. The first round occurred in April and May of 2016. The second round was in October 2016. Before youth could participate in an interview, consent was secured from a parent or guardian. STEM OST providers were sent a letter, consent form, and assent form to send home to parents/guardians (see Appendix H for these documents). The STEM OST providers then mailed or emailed back the signed consent and assent forms. Most of the think-aloud interviews took place over the phone, except for five in-person interviews with local youth. The youth completed the survey via the online survey platform, SurveyGizmo, while participating in the interview. An adult at the STEM OST program site set up the computer and got the survey started. The adult then left the room so the youth could do the interview in a private, confidential setting to help them feel okay saying they were

uncomfortable doing something or felt like they may not be very good at a skill. The interviews lasted around 45 to 60 minutes. In appreciation for their time, youth were given a $10 pre-paid VISA card.

Sample. A diverse range of youth were needed to ensure that the items worked for the variety of youth who participate in STEM OST programs. Recruitment efforts were such to ensure that diversity in terms of youth's age (grade 5 to 12), gender, racial/ethnic background, and team communication skills. Recruitment was also targeted to ensure that youth were interviewed from different geographic regions in the United States. During recruitment, STEM OST providers completed an information form about their program and each of the youth participating in an interview. The form asked for information about the STEM OST program, how the program used teams, youths' demographic information, and how youth tended to communicate on teams (with the options: shy/does not talk much, average/talks every so often, talks a lot, tends to dominate a conversation). The original intent was to have a maximum of four youth from a program (two middle school and two high school) and a pool of interested programs to be able to purposefully select a diversity of youth. There was difficulty finding programs and youth interested in participating in the interviews, so the sample ended up being any eligible programs and youth who were interested. See "sampling characteristics" below for additional details of the final think-aloud sample.

Youth were recruited for the think-alouds by connecting to STEM OST providers across the country. Invitations to participate were sent to STEM OST providers along with information about the think-aloud process. Individuals from the Phase 1 sampling frame were emailed, as well as people who had originally expressed interest during the original grant proposal needs assessment. A Facebook page for the project was created in April 2016 to share information and updates about the project and serve as an additional means of recruitment throughout the study. Recruitment information was also shared through a wide variety of listservs, discussion boards, LinkedIn groups, Twitter feeds, and Facebook pages (see Appendix C for a list of the places recruitment information was shared during each phase of the research).

Sample characteristics. A total of 30 youth from 11 STEM OST programs were interviewed (22 youth in the first round of think-alouds and eight in the second round). The think-aloud sample included a diversity of youth. Of the 30 youth, there were 12 boys and 18 girls. Most of the youth were in middle school (21), and nine were in high school. Youth spanned a variety of grades as illustrated in Table 12, with the highest number of youth in 7th grade.

Table 12

Grade Level of Interviewees (n = 30)

Grade	Number of youth
5th grade	1
6th grade	4
7th grade	11
8th grade	5
9th grade	3
10th grade	2
11th grade	3
12th grade	1

The sample included youth from a variety of racial/ethnic groups. Less than half of the sample self-identified as white (see Table 13). The "other" category included Indian.

Table 13

Race/Ethnicity of Interviewees (n = 30)

Race/Ethnicity	Number of youth
White or Caucasian	10
African-American/Black	6
Asian	4
Hispanic/Latino	3
Multi-racial	3
Native American/American Indian/Alaskan Native	0
Other	1
Prefer not to answer	3

When STEM OST providers completed the interest form for youth in their program, they indicated youth's typical communication style on a team. It was important that the sample was not composed of only youth who talked a lot on teams, but included youth with various communication styles so the items could be tested with a diverse range of youth. As illustrated in Table 14, the sample included a variety of youth with around half having average levels of communication or talking every so often within a team. Note that some providers did not fill out the interest form so information was missing for five youth.

Table 14

Team Communication Styles of Interviewees (n = 25)

Team communication	Percentage of youth
Shy/Does not talk much	12%
Average/Talks every so often	52%
Talks a lot	20%
Tends to dominate a conversation	16%

Programs were recruited from across the country, and 11 STEM OST programs chose to participate. Programs were located in seven states: California, Florida, Minnesota, New York, North Carolina, Pennsylvania, and Utah. Although there was some focused recruitment in regions of the country where there were gaps, there was not always interest from programs in those areas or a program ended up not able to participate. Youth participated in programs from a wide range of organizations, including science museums, a community-based organization, an aquarium, an art museum, schools, a natural history museum, and a university. Some youth came from programs that covered a variety of STEM topics, while others came from programs focused on specific areas of STEM such as computer programming, environmental science, marine science, engineering, natural science, and robotics. STEM OST program types included clubs, a teen volunteer program, youth development programs, a youth

employment program, design or robotics competition-type programs (including FIRST Lego League), and general afterschool programs.

Discussion. Over the course of the think-aloud interviews, data were gathered about the scenario, individual items, and response options. The think-aloud data and the resulting changes to the survey are described below.

Scenario. Overall, the teamwork scenario worked well for youth. After reading the scenario, youth were asked, "Was there anything confusing or hard to understand about this scenario?" None of the 22 youth who were asked this question (15 middle school and seven high school) found anything confusing or hard to understand about the scenario. These youth were then asked, "Can you imagine yourself in this situation? If no, what about this scenario makes it hard for you to imagine yourself in it?" Twenty-one youth said they could imagine themselves in the situation and that it sounded like something that could happen in their program. One high school youth said they had not done what was described in the scenario in their program before but could probably imagine it. When asked to explain the reason for saying "probably," they said, "Well, I mean it seems too perfect…Everyone is doing exactly what they need to do. I mean, there's always that like, one kind of person that doesn't stay on track all the time."

The only change that was deemed necessary to the teamwork scenario came out of youth's responses when talking about some of the items. The scenario was meant to be a team of youth who did not know each other very well. However, when thinking-aloud about questions, some youth referred to their teammates in the scenario as their friends. When it was clarified that the scenario was supposed to be about people they had just met, some youth said they would answer a different way, saying how they answered depended on how well they knew their teammates. For this reason, language was added to stress that youth had just met their teammates. This helped to make sure all youth were interpreting their relationship with their teammates in the same way.

After youth finished thinking-aloud about the closed-loop communication and information exchange items, they were asked if they were able to keep the scenario's imaginary team in mind or if

they started to think about other teamwork experiences. The first three youth participating in the think-alouds said they had a hard time remembering to think about the imaginary team. To help youth remain focused on the imaginary team described in the scenario, the survey instructions and items were revised to say "the imaginary team" instead of "the team."

At the beginning of Phase 3 think-alouds, the survey items were grouped by construct with all the closed-ended communication items first, then all the information exchange items, and then all the listening items. After 10 think-alouds, the closed-loop communication and information exchange items were mixed together and grouped so they made more sense cognitively in an effort to help youth better focus on particular aspects of the scenario and tasks at hand. Item groupings were based on sharing information, sharing their own idea, and teammates sharing ideas. A sentence was added before each grouping to provide some context about the next set of statements they were going to be thinking about and to help youth transition cognitively.

Individual items. Over the course of think-aloud testing, items were revised to ensure that they were being interpreted as intended, were easy to understand, and were clearly measuring the skill area of interest. A total of nine items ended up being removed before the pilot test. Eight items (7aGood, 7aComfort, 7aLikely, 11bGood, 11bComfort, 11bLikely, 4c, 7c[5]) were removed because how youth responded to the items did not align with the skill areas they were meant to measure meaning the items were not valid measures of the construct. One item (8c) went through multiple revisions but was ultimately removed because it remained confusing to youth or continued to be misinterpreted. Individual item changes and removal over time, along with reasons for changes and removal, can be found in Appendix I.

[5] Information exchange items are labeled with an "a," closed-loop communication items are labeled with a "b," and listening items are labeled with a "c." Information exchange and closed-loop communication items are also named based on if they are a good/bad, comfortable/uncomfortable, or likely/unlikely item.

Response options. There were four different rating scales used in the survey. For the closed-loop communication and information exchange items, the options for these ratings scales were the following.

- How good or bad do you think you would be at doing this on the imaginary team? (Bad at this, Kind of bad at this, Kind of good at this, Good at this)
- How comfortable or uncomfortable do you think you would be doing this on the imaginary team? (Uncomfortable doing this, Kind of uncomfortable doing this, Kind of comfortable doing this, Comfortable doing this)
- How likely or unlikely would you be to actually do this with the imaginary team? (Unlikely to do this, Kind of unlikely to do this, Kind of likely to do this, Likely to do this)

During the first 12 think-alouds, youth were asked if they could explain the difference between being good at something, comfortable with it, and the likelihood of doing it. Youth were able to recognize the differences among the three options. Youth were also asked if they understood the differences among the four response options for each of the three questions. All 25 youth (18 middle school, seven high school) who were asked this question said they understood the differences. Youth were also asked if they understood the differences among the four response options for the listening items (Hard to do this, Kind of hard to do this, Kind of easy to do this, Easy to do this). All 14 youth (nine middle school, five high school) who were asked about the listening response options said the options made sense to them.

Using the survey for middle and high school youth. Throughout Phases 2 and 3, findings were reviewed to see if separate middle and high school surveys were necessary. From the think-aloud results, items were revised so they could work for both age ranges. There was not a strong reason from the think-aloud data to create separate surveys since the items appeared to work well for both age ranges, so the same survey was used for middle and high school youth during the pilot and field tests.

Phase 3: Pilot Test

The second half of Phase 3 was pilot testing Version 4 of the survey with youth from STEM OST programs across the country. The purpose of the pilot test was to look at the reliability of the responses and gather internal structure validity evidence from exploratory factor analysis, item analysis, and, if a big enough sample could be obtained, DIF analysis (American Educational Research Association et al., 2014). Since there was overlap in definitions between information exchange and closed-loop communication, it was unclear if the three construct areas of team communication skills (information exchange, closed-loop communication, and listening) fell under one dimension or multiple dimensions. For this reason, exploratory factor analysis was carried out in Phase 3 to identify the model structure to later test through confirmatory factor analysis in Phase 4. The pilot test was also used to identity items to remove from the instrument to decrease the length of the survey so it only took 5 to 10 minutes to complete, which was more in line with the length of instruments typically used to evaluate STEM OST programs.

Method. The pilot test consisted of STEM OST programs across the country testing the survey in their program. The STEM OST program provider was sent a passive consent letter to send home to each youth's parent/guardian before the survey was administered (see passive consent letter in Appendix J). The survey was personalized for each program by including the name of the program in the scenario, indicating in the scenario if they were supposed to imagine they were working with middle or high school youth, and throughout the survey referring to the work they were doing together on the imaginary team as either a "program" or "challenge" to align with how youth worked on teams in that program (see Appendix K for Version 4 of the survey used for the pilot test). Each program got a personalized online survey that was created through the online survey software, SurveyGizmo, or a paper version. In some cases, programs administered the survey both online and on paper depending on the access they had to computers or tablets. If the survey was administered on paper, the surveys were sent back and entered manually into SurveyGizmo. The survey took youth around 20 minutes to complete. Data were collected

from December 2016 to March 2017. In appreciation for their time, programs received a pre-paid VISA card ($25 for up to 20 complete surveys, $50 for more than 20 complete surveys) as well as a report of the aggregate survey results for their program that was generated by SurveyGizmo. Data from the pilot survey were used to conduct exploratory factor analysis, item analysis, and reliability analysis. The pilot test sample size of 310 was not large enough to meet the minimum sample size requirements for DIF analysis of at least 100 in the focal group and 400 in the reference group (Paek & Guo, 2011).

Sample. A diverse range of youth and programs from across the country were needed for the pilot test. Recruitment efforts were such to help obtain a diversity of middle and high school STEM OST programs in terms of content areas covered, program type, grade ranges covered, and geographic region of the country. During recruitment, STEM OST providers completed an information form about their program. The form asked for information about the STEM OST program, how the program used teams, and the number of youth in each grade range that might complete the survey. Programs had to include some kind of team-based activities to be included in the sample. There was difficultly finding programs interested in participating in the pilot test to create a sampling frame in which to draw programs from, so the sample ended up being any eligible program that was interested. See "sample characteristics" below for additional details of the final pilot test sample of youth and programs.

Programs were recruited for the pilot test by connecting to STEM OST providers across the country. Individuals from the Phase 1 sampling frame were emailed, as well as people who had originally expressed interest during the NSF grant proposal needs assessment. Participants from previous phases of the research were contacted about the pilot study. The National Afterschool Association invited the project to write up a blog post about the study that also included an invitation for afterschool programs to participate in the pilot test (Grack Nelson, 2016). The project presented at the American Evaluation Association Conference in October 2016, and session participants signed up to be contacted about the pilot study. Additionally, recruitment information was shared through a wide variety of listservs, discussion boards, LinkedIn groups, Twitter feeds, the project's Facebook page, and other Facebook

pages of other relevant organizations (see Appendix C for a list of the places recruitment information was shared during each phase of the research).

Sample characteristics. A total of 378 youth completed the survey during the pilot test. However, the data were cleaned so that any youth who did not answer all of the team communication survey items were removed from the sample (this removal rule did not apply to how youth answered the demographic questions). There were also four youth whose surveys were removed from the sample because it was clear they did not take the survey seriously from the comments they put in the demographic open-ended questions, such as "dog" for race ethnicity and "smart kids" for grade, in addition to the other indicators like choosing the lowest options for each question and finishing the survey in under 5 minutes (which can be calculated from the survey software). After removing these youth from the sample, the final sample size for the pilot test was 310 youth. Minimum sample size recommendations for exploratory factor analysis vary, and the sample size of 310 youth, although low, falls within some of the guidelines that recommend a minimum of five participants per scale item, so the sample size was sufficient for the exploratory nature of the analyses of the pilot test data (Osborne, 2014).

The pilot test sample included a diversity of youth (see Table 15). Youth spanned all grade levels, with the fewest percentage of youth in 5th grade (7%). The youth who indicated "other" for grade completed the high school survey, but did not explain his/her response. There were more girls in the sample than boys. The sample included youth from a variety of racial/ethnic groups with just over a third of the youth (36%) identifying as white. Two youth only checked "other," one wrote in Arctic and the other youth did not identify their race/ethnicity. Over half the youth had a mother who had graduated college, a question taken from the National Assessment of Educational Progress survey that can serve as a proxy for socioeconomic status (National Center for Education Statistics, 2012, 2015).

Table 15

Demographic Characteristics of Youth in the Pilot Test Sample

Demographic variable	Percentage of youth
Grade (*n* = 309)	
5th grade	7%
6th grade	9%
7th grade	18%
8th grade	10%
9th grade	10%
10th grade	17%
11th grade	16%
12th grade	12%
Other	<1%
Gender (*n* = 279)	
Female	56%
Male	43%
Transgender	2%
Race/Ethnicity (*n* = 258)	
White or Caucasian	36%
African-American/Black	26%
Hispanic/Latino	14%
Asian/Asian American	12%
Multi-racial	11%
Native American/American Indian/Alaskan Native	1%
Other	1%
Mother's education level (*n* = 305)	
Did not finish high school	7%
Graduated high school	9%
Some education after high school	13%
Graduated college	57%
I don't know	14%

Youth were also asked how long they had been in their STEM OST program. This question was used to get a sense of the exposure youth had to the team-based experiences in their program and gauge the extent to which the survey was tested with youth of varying levels of skill, with the assumption that youth who have less exposure to the STEM OST program may perceive themselves as having lower levels of team communication skills, be less comfortable with the skills, and be less likely to exhibit them

on a team. As illustrated in Table 16, the sample included youth with a wide range of experiences from less than a month to over a year, with half being in the programs six months or less and the other half over six months.

Table 16

Length of Time Youth Participated in Their STEM OST Program (n = 305)

Length of time	Percentage of youth
Less than a month	12%
1 to 6 months	38%
7 to 12 months (a year)	13%
Over a year	38%

A total of 19 STEM OST program sites from across the country chose to participate in the pilot test. Four of these sites had two different programs participate in the pilot study so there were 23 total programs in the sample. Program sites were located in Washington DC and across 16 states: Arizona, California, Connecticut, Florida, Illinois, Minnesota, Missouri, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Pennsylvania, Washington, and Wisconsin. Youth participated in programs from a wide range of organizations including science museums, community-based organizations, a Boys and Girls club, 4-H, schools, a nature center, libraries, and zoos. Some youth came from programs that covered a variety of STEM topics, while others came from programs focused on specific areas of STEM such as math, gardening, robotics, marine science, engineering and design, making, agriculture, natural science, and conservation. STEM OST program types included clubs, a teen volunteer program, youth development programs, youth employment programs, robotics competition-type programs (including FIRST Lego League), internship programs, a Future Farmers of America program, and general afterschool programs.

Discussion

Comparing good, comfort, and likely question types. To first aid in decisions of item removal, data from the three follow-up question types (good, comfort, and likely) used across the information

exchange and closed-loop communication items were reviewed to see if one of the response types could possibly be removed across the items as a way to shorten the survey. Correlations were calculated between the sum scores of the three different question types. As illustrated in Table 17, the three item types were highly correlated, but none of them had correlations of .90 or higher.

Table 17

Correlations Matrix for the Three Question Types

	Good items	Comfort items	Likely items
Good items	1.00		
Comfort items	.86	1.00	
Likely items	.83	.78	1.00

It was ultimately decided that shortening the survey by removing one of the three question types (good, comfort, or likely) would not do much to decrease the cognitive load on youth completing the survey. Having more anchor statements, as opposed to removing entire item sets, would require more cognitive load as youth would need to imagine themselves in more situations. This meant that even though the survey might have fewer items, it may take longer to complete than a survey of equivalent length but with fewer anchor statements (or item sets) they would need to consider.

Exploratory factor analysis. Exploratory factor analysis was carried out to understand how many factors, or dimensions, were measured by the items and if the items loaded well on the factor(s) (Brown, 2015). The dimensionality of the construct was examined to understand if the construct of team communication skills was unidimensional or multidimensional. Any items that did not load well suggested that the way youth responded to that item would not be influenced by the underlying construct of team communication skills (or one of three skill areas if a multidimensional construct), hence not a valid measure of the construct. Items that did not load well were removed from the instrument.

Exploratory factor analysis was carried out for ordinal data in R Studio using a polychoric correlation matrix, weighted least squares extraction method, and oblique rotation. To guide factor

selection, eigenvalues were calculated and interpreted using the Kaiser-Guttman rule. There were 13 eigenvalues above 1, which according to the rule would suggest 13 factors (Brown, 2015). Parallel analysis scree plots were also reviewed. As illustrated in Figure 2, the plots suggest a solution of seven factors based on the number of points that fall above the eigenvalue line of 1.

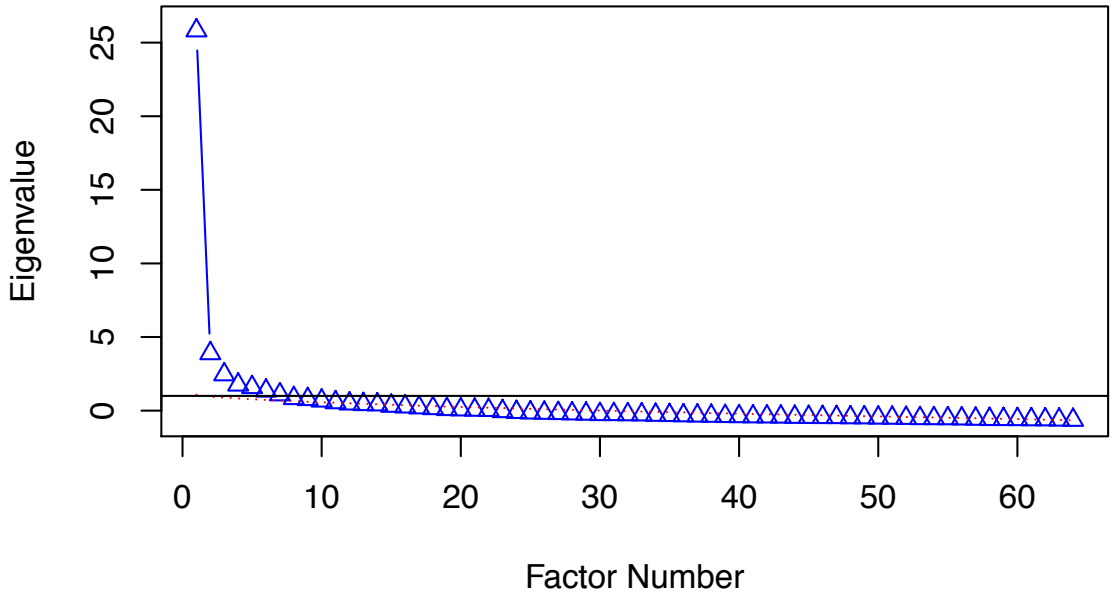

Figure 2. Parallel analysis scree plots for the pilot test data. The triangles represent the actual data and the dotted line represents the simulated data.

Exploratory factor analyses were run with one through seven factors. The five-factor model worked best both empirically and theoretically. Below is a summary of results from the exploratory factor analyses results that did not work well, and then the five-factor model is discussed.

- With a seven-factor solution, one factor was made up of only one item set (9b) so there were only three items for the factor, which was not ideal. A clear seventh factor also did not emerge by looking at the factor loadings. Factor loadings across the seventh factors ranged from $\lambda = -.32$ to $\lambda = .44$, with only four items having factor loadings higher than the recommended .30 cutoff

(Furr & Bacharach, 2014). Three of these four items also loaded much higher on a different factor.

- With a six-factor solution, a clear sixth factor did not emerge. Factor loadings across the sixth factor ranged from $\lambda = -.32$ to $\lambda = .45$, with only three items having factor loadings higher than .30 and two of these items loaded much higher on a different factor.
- With a four-factor solution, item sets 8a and 9a did not load strongly on a factor (factor loadings ranged from $\lambda = .02$ to $\lambda = .37$). These two-item sets were an important part of the closed-loop communication construct area related to youth asking their teammates if they understood their idea, which would not be covered if these items were removed.
- With a three-factor solution, items representing important parts of the team communication skills construct areas were not loading well on a factor. The listening factor disappeared as none of the items related to the listening construct area loaded strongly on a factor (factor loadings ranged from $\lambda = -.01$ to $\lambda = .43$, with only two items having a factor loading above .30). Like the four-factor solution, item sets 8a and 9a did not load strongly on a factor (factor loadings ranged from $\lambda = .05$ to $\lambda = .42$, with three items having a factor loading above .30). As previously mentioned, these two item sets were an important part of the closed-loop communication construct area.
- With a two-factor solution, the listening items did not load strongly on a factor (factor loadings ranged from $\lambda = .02$ to $\lambda = .43$, with only two items having a factor loading above .30). Listening is an important construct area that would not be covered if the items were removed.
- With a one-factor solution, the listening items did not load strongly on a factor (factor loadings ranged from $\lambda = .36$ to $\lambda = .48$). A one-factor solution also was not supported by the scree plot.

A five-factor solution was chosen because it was most interpretable. Even though the eigenvalues and parallel analysis scree plots pointed to more factors, the larger factor solutions, as described above, did not make sense empirically or theoretically for the constructs being measured. As illustrated Table 18,

the five-factor model had five distinct factors that spanned the three construct areas and the items that did not load well on any of the factors were not key areas of the construct that would be missing if the items were removed from the instrument. Figure 3 outlines the item sets that loaded strongly on each factor along with the four item sets that did not clearly load on a factor. In all instances in which there was an anchor statement with three items, the good, comfort and likely items all loaded on the same factor.

Table 18

Factor Loadings for the Five-Factor Exploratory Factor Analysis Solution

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Q5bGood	.63	-.09	-.01	.22	.05
Q5bComfort	.72	-.08	.12	.02	.10
Q5bLikely	.84	-.06	-.06	-.07	-.04
Q6bGood	.79	-.01	.02	-.01	-.06
Q6bComfort	.73	-.10	.14	.03	.01
Q6bLikely	.83	-.10	-.06	-.04	.05
Q7bGood	.68	.18	.01	.00	.00
Q7bComfort	.73	.09	.00	.08	-.03
Q7bLikely	.75	.18	-.26	.04	.00
Q10bGood	.65	.07	.00	.10	.10
Q10bComfort	.72	.12	.07	.04	-.04
Q10bLikely	.70	.10	-.21	.02	.16
Q1bGood	.53	-.02	.16	.13	.15
Q1bComfort	.64	.02	.27	.06	-.06
Q1bLikely	.67	-.05	.13	.06	.13
Q10aGood	.43	.02	.07	.21	.22
Q10aComfort	.48	.10	.20	.17	.10
Q10aLikely	.47	.18	-.09	.13	.18
Q1aGood	.05	.62	.09	.18	.01
Q1aComfort	.09	.61	.22	.11	-.17
Q1aLikely	.04	.70	-.15	.02	.05
Q2aGood	.04	.75	.17	-.02	.00
Q2aComfort	.06	.72	.20	-.02	-.08
Q2aLikely	.09	.82	-.09	-.07	.07
Q3aGood	-.10	.75	.15	-.04	.11
Q3aComfort	-.02	.71	.26	.06	-.04
Q3aLikely	-.04	.90	-.11	-.08	.11
Q4aGood	.05	.69	.07	.09	.09
Q4aComfort	.11	.68	.11	.14	-.07

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Q4aLikely	.04	.81	-.20	.08	.02
Q5aGood	-.08	.70	.03	.21	.13
Q5aComfort	-.03	.70	.20	.14	-.09
Q5aLikely	.03	.75	-.18	-.02	.19
Q3bGood	.06	.09	.75	.03	.02
Q3bComfort	-.01	.13	.76	.07	-.02
Q3bLikely	-.01	.21	.68	.07	-.04
Q4bGood	.00	-.10	.72	.04	.33
Q4bComfort	-.01	.01	.73	.02	.22
Q4bLikely	.03	.05	.56	.02	.28
6c	.00	.09	-.09	.66	-.09
2c	-.01	.01	-.03	.78	.03
5c	.06	.12	-.01	.62	-.07
3c	.03	-.03	.02	.78	-.01
9c	.20	-.08	-.08	.68	.02
1c	-.07	-.04	.05	.76	.10
10c	-.18	.14	.03	.50	.12
Q8aGood	.17	.05	.08	.06	.64
Q8aComfort	.23	.16	.24	-.06	.47
Q8aLikely	.22	.12	-.10	.04	.67
Q9aGood	-.01	.06	.13	.08	.73
Q9aComfort	.11	.10	.26	.00	.56
Q9aLikely	-.08	.06	.00	.07	.81
Q8bGood	.32	.23	.46	-.06	-.12
Q8bComfort	.29	.10	.47	.15	-.07
Q8bLikely	.33	.33	.37	-.04	-.16
Q6aGood	.35	.21	.13	-.03	.29
Q6aComfort	.35	.23	.22	.02	.16
Q6aLikely	.34	.38	.05	-.10	.23
Q2bGood	.42	.09	.40	.10	.04
Q2bComfort	.31	.04	.51	.15	.01
Q2bLikely	.45	.10	.30	.01	.11
Q9bGood	.16	.24	.33	.21	.13
Q9bComfort	.19	.18	.36	.19	.16
Q9bLikely	.16	.29	.12	.17	.20

Factor 1 *Information exchange items related to youth sharing their information/idea with the team.*
1b. Explaining an idea you have to the team.
5b. Sharing information you found about the topic of the [project/challenge] with your team.
6b. Sharing information you found about the topic of the [project/challenge] that none of your teammates have mentioned yet.
7b. Sharing information you found about the topic of the [project/challenge] when you notice it might be useful to the team, instead of waiting for a teammate to ask you for the information.
10b. Updating the team with new information you found related to the [project/challenge].
10a. Explaining your idea in more than one way if a teammate is confused about your idea.

Factor 2 *Closed-loop communication items related to a teammate's idea.*
1a. Repeating back your teammate's idea to make sure you understood it correctly.
2a. Letting a teammate know that you are having trouble understanding his or her idea.
3a. Asking your teammate to explain his or her idea in a different way so you can understand it better.
4a. Asking a question if there is something confusing about your teammate's idea to make sure you understand it.
5a. Asking your teammate to repeat his or her idea because you are unsure if you understood it correctly.

Factor 3 *Information exchange items about bringing up an idea that might be more difficult to share.*
3b. Sharing an idea even if you think your team might dislike it.
4b. Bringing up an idea for the [project/challenge] that is different from the idea the team just finished discussing.

Factor 4 *Listening items.*
1c. Stay focused on what a teammate is saying when you would rather be working on your part of the team project.
2c. Listen closely to a teammate share an idea instead of focusing on what you are going to say to the team about your own idea.
3c. Fully focus on what a teammate is saying instead of thinking about what you are going to say next to the team.
5c. Listen to a teammate that has an idea of what the team should do that is different from your idea.
6c. Keep listening to what a teammate is saying even if you disagree with the teammate's idea.
9c. Stay focused on the conversation your team is having instead of letting your mind wander.
10c. Stop working on the team's project while a teammate is updating the group with new information.

Factor 5 *Closed-loop communication items related to youth's own idea.*
8a. Asking your teammates if they understand your idea.
9a. Encouraging your teammates to ask you questions about your idea to make sure they understand it correctly.

Did not load clearly on a factor
2b. Sharing an idea about the [project/challenge] that is different from a teammate's.
8b. Sharing information with the team even if you are unsure if it will be useful to the [project/challenge].
9b. Summarizing out loud the ideas your team discussed about the [project/challenge].
6a. Asking a question to better understand the information your teammate found about the topic of the [project/challenge].

Figure 3. Exploratory factor analysis five-factor solution and associated items.

Item analysis and descriptive statistics. After the five-factor solution was chosen, item analysis was carried out to see how well each item was discriminating between individuals with low and high levels of team communication skills within the factor and identify items to remove from the instrument. Discrimination (item-total correlation) was calculated based on the total score of the factor the item belonged to (Meyer, 2014). The four item groups (2b, 8b, 9b, 6a) that did not load clearly on one factor were not included in the calculation of discrimination or difficulty since those items were already tagged to be removed from the instrument. Item analysis was conducted in the software JMetrik Version 4.0.4 using the "Pearson Correlation" and "Correction for Spuriousness" settings. Discriminations ranged from .36 to .77 (see full results in Table 19). The lowest discriminations were for items 10c (.36), 5c (.46), and 6c (.48). Since all of the discriminations were above .30, the items were discriminating well between youth with high and low scores within a given factor (Meyer, 2014), although, the items with discrimination values closer to the cutoff value could be considered for possible removal.

Item difficulty, or mean of the item, was calculated for each item to indicate how difficult an item was where the lower the value, the more difficult the item was to endorse (Meyer, 2014). As illustrated in the Table 18, difficulties ranged from 2.75 to 3.78. The highest difficulties (or easiest items to endorse) were items 6bLikely (3.78) and 10bGood (3.51), while the lowest difficulties (or hardest items to endorse) were 3bComfort (2.75) and 3bLikely (2.82). These findings were also reflected in the frequencies of each response option for an item (see Table 19). Items 10bGood (59%), 10bComfort (59%), and 10bLikely (57%) had the highest percent of youth choosing the highest response option. Items 5bGood and 10bGood had items with the lowest percentage of youth (1%) choosing the lowest response option. As evident by the pattern of responses, all of the items were negatively skewed. However, even though they were skewed, none of the items had what might be considered a ceiling effect, with 59% being the highest percentage of youth selecting the highest rating option. There were 11 items that had a higher frequency of response to the third response category than to the fourth (highest) response category (all three 3b items, all three 4b items, 8aGood, 8aLikely, 9aGood, 9aLikely, and 1c).

Table 19

Pilot Test Item Analysis Results and Frequency for Each Response Option (n = 310)

Item	Discrimination	Difficulty (*M*)	*SD*	Response options 1	2	3	4
Factor 1 (Information exchange)							
1bGood	.65	3.29	0.79	3%	11%	39%	47%
1bComfort	.68	3.28	0.85	4%	13%	33%	49%
1bLikely	.69	3.38	0.78	3%	10%	34%	53%
5bGood	.62	3.44	0.63	1%	5%	43%	51%
5bComfort	.68	3.36	0.76	2%	10%	37%	51%
5bLikely	.59	3.40	0.78	3%	9%	32%	56%
6bGood	.65	3.42	0.71	2%	8%	37%	54%
6bComfort	.63	3.33	0.79	4%	10%	37%	49%
6bLikely	.63	3.78	0.74	2%	9%	38%	51%
7bGood	.67	3.38	0.77	3%	10%	35%	53%
7bComfort	.73	3.33	0.76	2%	11%	38%	48%
7bLikely	.65	3.34	0.76	2%	11%	38%	49%
10bGood	.65	3.51	0.66	1%	6%	34%	59%
10bComfort	.71	3.47	0.72	2%	7%	32%	59%
10bLikely	.61	3.44	0.74	2%	8%	33%	57%
10aGood	.63	3.20	0.85	5%	14%	38%	44%
10aComfort	.67	3.32	0.80	4%	9%	38%	49%
10aLikely	.60	3.37	0.81	4%	10%	32%	54%
Factor 2 (Closed-loop communication)							
1aGood	.67	3.35	0.78	4%	7%	39%	50%
1aComfort	.63	3.26	0.83	5%	11%	38%	47%
1aLikely	.55	3.19	0.89	7%	11%	38%	44%
2aGood	.77	3.23	0.84	4%	16%	35%	46%
2aComfort	.73	3.10	0.91	6%	20%	33%	41%
2aLikely	.71	3.17	0.90	6%	16%	33%	45%
3aGood	.70	3.13	0.92	7%	16%	35%	43%
3aComfort	.74	3.09	0.93	6%	21%	31%	42%
3aLikely	.73	3.08	0.94	8%	15%	37%	40%
4aGood	.73	3.34	0.78	3%	10%	37%	50%
4aComfort	.74	3.30	0.81	3%	13%	35%	49%
4aLikely	.65	3.35	0.80	4%	8%	37%	52%
5aGood	.74	3.34	0.80	4%	9%	36%	51%
5aComfort	.72	3.21	0.86	4%	16%	35%	45%
5aLikely	.62	3.27	0.85	6%	8%	38%	47%

Item	Discrimination	Difficulty (*M*)	*SD*	Response options 1	2	3	4
Factor 3 (Information exchange)							
3bGood	.75	2.90	0.93	9%	21%	40%	30%
3bComfort	.70	2.75	0.99	13%	27%	34%	27%
3bLikely	.72	2.82	0.98	12%	22%	38%	28%
4bGood	.74	3.05	0.89	7%	18%	39%	36%
4bComfort	.72	2.91	0.93	8%	25%	37%	31%
4bLikely	.68	2.87	0.95	11%	21%	40%	29%
Factor 4 (Listening)							
1c	.59	3.05	0.85	5%	18%	43%	34%
2c	.60	3.18	0.85	5%	13%	40%	42%
3c	.60	3.08	0.84	5%	16%	44%	51%
5c	.46	3.37	0.76	3%	8%	38%	51%
6c	.48	3.12	0.83	3%	20%	39%	38%
9c	.53	3.16	0.90	7%	15%	36%	43%
10c	.36	3.12	0.88	5%	19%	36%	40%
Factor 5 (Closed-loop communication)							
8aGood	.67	3.25	0.76	3%	12%	44%	42%
8aComfort	.63	3.23	0.80	3%	16%	39%	43%
8aLikely	.64	3.21	0.82	5%	11%	43%	41%
9aGood	.71	2.97	0.94	9%	19%	38%	34%
9aComfort	.70	3.10	0.88	5%	20%	36%	40%
9aLikely	.64	2.93	0.94	10%	18%	41%	31%

Note. Response options for Good items: (1) Bad at this, (2) Kind of bad at this, (3) Kind of good at this, (4) Good at this. Response options for Comfort items: (1) Uncomfortable doing this, (2) Kind of uncomfortable doing this, (3) Kind of comfortable doing this, (4) Comfortable doing this. Response options for Likely items: (1) Unlikely to do this, (2) Kind of unlikely to do this, (3) Kind of likely to do this, (4) Likely to do this. Response options for Listening items: (1) Hard to do this, (2) Kind of hard to do this, (3) Kind of easy to do this, (4) Easy to do this.

Revision of the survey. Part of the Phase 3 pilot test was to identify items to remove so that the instrument could be shortened while still adequately measuring the construct areas identified by the five factors. The goal was to create a final survey that would take youth 5 to 10 minutes to complete. To guide removal of items for each factor, a list of decision rules was created.

1. Look at exploratory factor analysis results. Remove items that do not clearly load on one factor or load on a factor that does not align with the theory and definitions of the three construct areas.

2. Look at descriptive statistic results. Remove items with low variability of responses as defined by having 1% of responses at the lowest response option and the highest percent of youth responding to the highest response option.
3. Look at item discrimination results. If necessary, remove items with the lowest discrimination values within a factor.
4. Finally, look at think-aloud results and grant advisor expert review comments to help identify additional items to remove.

If one item within an item set (an anchor statement with three follow up questions) was tagged for removal, the entire item set was removed. The goal for the information exchange and closed-loop communication factors was to end up with six items (two item sets) in each factor.

Removing items based on exploratory factor analysis results. As illustrated in the exploratory factor analysis results in Table 18, there were two item sets loading on two factors. The three items for statement 2b were loading on both Factor 1 (2bGood $\lambda = .42$, 2bComfort $\lambda = .31$, 2bLikely $\lambda = .45$) and Factor 3 (2bGood $\lambda = .40$, 2bComfort $\lambda = .51$, 2bLikely $\lambda = .30$). Looking at the items and the factors, the items related to both of these information exchange factors since the statement talked about sharing an idea in general and sharing something different. This statement was similar to 4b as both were about bringing up a different idea, which provided additional support for removing the item because the area of the construct would still be covered. The three items for statement 8b were also loading onto both Factor 1 (8bGood $\lambda = .32$, 8bComfort $\lambda = .29$, 8bLikely $\lambda = .33$) and Factor 3 (8bGood $\lambda = .46$, 8bComfort $\lambda = .47$, 8bLikely $\lambda = .37$). The items were related to both sharing information (Factor 1) and the uncertainty of sharing the information (Factor 3), which could account for why they loaded on the two different information exchange factors.

There were two item sets (9b and 6a) that did not have any items loading well on any of the five factors (see Table 18). Statement 9b also caused problems during think-alouds as there were still some

youth who were not interpreting the statement as intended after three rounds of revision. Statement 6a was the only closed-loop communication item that talked about "information" instead of "ideas" which might be why it was not factoring with the other closed-loop communication items.

One of the item sets, 10a, was not loading onto the factor that aligned with the construct area it was meant to measure. 10a was meant to measure closed-loop communication but it was loading on an information exchange factor (although the loadings were low for the three 10a items ranging from $\lambda = .43$ to $\lambda = .48$). The part of the item "explaining an idea" aligned with the information exchange items, but the overall item was meant to get at clarifying understanding if a teammate is confused, which was part of closed-loop communication.

Removing items based on descriptive statistics. Another way to identify items to remove was by looking at the distribution of responses and tagging items that had low variability of responses as defined by having lowest percentage of responses for the lowest option and highest percentage of responses for the highest option. As illustrated in Table 19, two items (5bGood and 10bGood) had only 1% of youth choosing the lowest response option. Both of these item sets also had the items with the highest percentage of youth choosing the highest response option (10bGood 59%, 10bComfort 59%, 10bLikely 57%, 5bLikely 56%).

Removing items based on item analysis results. To aid in item removal, the discrimination indices were also reviewed. As stated earlier, all of the items had discrimination indices above .3, so those closest to .3 within a factor were reviewed for removal. Factor 4 had some of the lowest discrimination indices of all the items. The items with the three lowest discrimination indices in Factor 4 were removed (10c = .36, 5c = .46, 6c = .48). Items still needed to be removed from Factor 2, so the item set with the lowest discrimination in the factor was chosen (1aLikely = .55). Additional items still needed to be removed from Factors 1 and 2, so the think-aloud results and grant advisor expert review results were reviewed to help aid in that decision since the rest of the items in those factors had discriminations above .60.

Removing items based on think-aloud and grant advisor expert review results. More items had to be identified for removal for Factors 1 and 2, so the think-aloud results (see Appendix I) and comments from the grant advisor expert review were reviewed. One more item set needed to be removed for Factor 1, so 7b was removed since some youth had issues interpreting the item as intended even after multiple revisions to the wording during the think-alouds. Youth also had some problems with statement 2a during think-alouds, which was about letting other youth know if they have trouble understanding their idea. During the grant advisor expert review, the youth development evaluation expert said youth can be hesitant to say they are confused by what someone said because some youth may find that offensive. This provided additional reason to remove statement 2a. Statement 4a was also about youth vocalizing confusion about someone's idea so that item set was removed as well.

Summary of item removal. The decision rules were used to inform removal of items from the survey. A total of 36 items were removed. The a and b statements (closed-loop communication and information exchange items) each had three items (good, comfort, and likely) associated with each statement. Table 20 outlines a summary of the items and item sets that were removed and the main reason for their removal. This left a total of 28 items on the survey with the information exchange and closed-loop communication factors each having six items (two item sets) and the listening factor having four items. The resulting five factors and associated items that were then tested in Phase 4 are illustrated in Figure 4 and survey Version 5 that was tested during the field test can be found in Appendix L.

Table 20

Items Removed After Phase 3 and Reasons for Their Removal

Item removed	Associated factor	Main reason for removal
2b	N/A	Loaded on two factors
8b	N/A	Loaded on two factors
9b	N/A	Did not load on any factors
6a	N/A	Did not load on any factors
10a	Factor 1	Item loading onto wrong construct area
5b	Factor 1	Only 1% chose lowest rating
10b	Factor 1	Only 1% chose lowest rating, 59% highest rating
6c	Factor 5	Lowest discrimination within factor
10c	Factor 5	Lowest discrimination within factor
5c	Factor 5	Lowest discrimination within factor
1a	Factor 2	Lowest discrimination within factor
7b	Factor 1	Think-aloud results
2a	Factor 2	Think-aloud and expert review results
4a	Factor 2	Think-aloud and expert review results

Closed-Loop Communication Factor 1

Closed-loop communication items related to a teammate's idea.

3a. Asking your teammate to explain his or her idea in a different way so you can understand it better.

5a. Asking your teammate to repeat his or her idea because you are unsure if you understood it correctly.

Closed-Loop Communication Factor 2

Closed-loop communication items related to youth's own idea.

8a. Asking your teammates if they understand your idea.

9a. Encouraging your teammates to ask you questions about your idea to make sure they understand it correctly.

Information Exchange Factor 1

Information exchange items that are about youth sharing their information/idea with the team.

1b. Explaining an idea you have to the team.

6b. Sharing information you found about the topic of the [project/challenge] that none of your teammates have mentioned yet.

Information Exchange Factor 2

Information exchange items about bringing up an idea that might be more difficult to share.

3b. Sharing an idea even if you think your team might dislike it.

4b. Bringing up an idea for the [project/challenge] that is different from the idea the team just finished discussing.

Listening Factor

1c. Stay focused on what a teammate is saying when you would rather be working on your part of the team project.

2c. Listen closely to a teammate share an idea instead of focusing on what you are going to say to the team about your own idea.

3c. Fully focus on what a teammate is saying instead of thinking about what you are going to say next to the team.

9c. Stay focused on the conversation your team is having instead of letting your mind wander.

Note. Information exchange and closed-loop communication statements each had three follow up questions: 1) How good or bad do you think you would be doing this on the imaginary team? (Bad at this, Kind of bad at this, Kind of good at this, Good at this) 2) How comfortable or uncomfortable would you be doing this on the imaginary team? (Uncomfortable doing this, Kind of uncomfortable doing this, Kind of comfortable doing this, Comfortable doing this) 3) How likely or unlikely would you be to actually do this with the imaginary team? (Unlikely to do this, Kind of unlikely to do this, Kind of likely to do this, Likely to do this)

Listening statements were grouped with one overarching question. How easy or hard would it be for you to do each of these things with the imaginary team? (Hard to do this, kind of hard to do this, kind of easy to do this, easy to do this)

Figure 4. Five-factor solution with items related to each factor after the Phase 3 pilot test.

Reliability. The reliabilities of the factor scores for the proposed five-factor solution were calculated in SPSS Version 24 by computing coefficient alpha (Thorndike & Thorndike-Christ, 2010). As illustrated in Table 21, the factors were reliably measuring the associated construct area as evidenced by the coefficient alpha estimates ranging from $\alpha = .78$ to $\alpha = .89$. All of the coefficient alpha values were above $\alpha = .70$, which is considered an acceptable value (DeVellis, 2012).

Table 21

Coefficient Alpha Estimates for Each Factor

Factor	Coefficient alpha
Closed-loop Communication 1	.89
Closed-loop Communication 2	.87
Information Exchange 1	.85
Information Exchange 2	.89
Listening	.78

Additional validity evidence. Before Phase 4, the grant's expert advisor in teamwork science was consulted to gather additional content validity evidence. He was sent the five-factor factor analysis results, the decision rules, the definitions of the three construct areas, and the resulting five factors after items were removed. Two main questions were posed to him:

1. Does each individual item align with the construct area we identified?
2. Did we propose removing any items that covered a key area of the construct that is no longer being measured?

He felt all of the items aligned with the construct areas identified and the way he defined the three construct areas in his work was represented by the items. He felt that even though the scales were short, each of the items was a characteristic of the construct, and he did not feel anything was missing. He felt the instrument was "robust" and informed well by the literature.

Phase 4: Field Test

The final phase of the research, Phase 4, was a national field test of Version 5 of the survey. The purpose of the field test was to test the survey with a larger cross-section of STEM OST programs and youth. The field test data were used to look at the reliability of the responses and gather additional internal structure validity evidence through confirmatory factor analysis, item analysis, and DIF analysis (American Educational Research Association et al., 2014).

Method. The field test consisted of STEM OST programs across the country testing the survey in their program. The methodology was identical to the pilot test. The STEM OST program provider was sent a passive consent letter to send home to youth's parent/guardian before the survey was administered (see passive consent letter in Appendix J). The survey was personalized for each program by including the name of the program in the scenario, indicating in the scenario if they were supposed to imagine they were working with middle or high school youth, and throughout the survey referring to the work they were doing together on the imaginary team as either a "project" or "challenge" to align with how youth worked on teams in that program (see Appendix L for Version 5 of the survey used for the field test). Each program got either a personalized online survey that was created through the online survey software, SurveyGizmo, or a paper version. In some cases, programs administered the survey both online and on paper depending on the access they had to computers or tablets. If the survey was administered on paper, the surveys were sent back and entered manually into SurveyGizmo. The survey took youth around 5 to 10 minutes to complete. Data were collected from August to October 2017. In appreciation for their time, programs received a pre-paid VISA card ($25 for up to 20 complete surveys, $50 for more than 20 complete surveys) as well as a report of the aggregate survey results for their program that was generated by SurveyGizmo.

Confirmatory factor analysis, item analysis, reliability analysis, and DIF analysis were conducted on the field test data. Confirmatory factor analysis was run for the proposed five-factor model in R Studio. The "lavaan" package for latent variable analysis was used and the R script specified that the data

were ordinal so that the diagonally weighted least squares (WLSMV) estimator would be used as has been suggested for confirmatory factor analysis of ordinal data (Johnson, 2011). Item analyses (calculations of item discrimination and difficulty) were conducted in the software JMetrik Version 4.0.4 using the "Pearson Correlation" and "Correction for Spuriousness" settings. Descriptive statistics and reliability (coefficient alpha) were conducted in SPSS Version 24. Fairness of the survey items was investigated based on race/ethnicity and gender using DIF analysis.

The Mantel-Haenszel procedure for DIF analysis was run in SPSS Version 24. The only racial/ethnic groups with large enough sample sizes for DIF analyses were African-American/Black youth (n = 113), Asian youth (n = 168), and white youth (n = 396). The gender comparison was between youth who identified as a boy or a girl. Matching variables were developed for each factor based on total factor scores. In SPSS using the ranking option "Ntiles" to specify the formation of six approximately equal groups, youth in the focal and reference groups were matched based on six levels of the trait measured by the overall factor. These matching variables were then used for the DIF analyses.

Sample. A diverse range of youth and programs from across the country were needed for the field test. The goal was to get between 800 to 1,000 youth to ensure that there were sufficient sample sizes of racial/ethnic subgroups to meet the minimum sample size requirements for DIF analysis of at least 100 in the focal group and 400 in the reference group (Paek & Guo, 2011). Similar to the pilot test, recruitment efforts were such to help obtain a diversity of middle and high school STEM OST programs in terms of content areas covered, program type, grade ranges covered, and geographic region of the country. During recruitment, STEM OST providers completed an information form about their program. The form asked for information about the STEM OST program, how the program used teams, and the number of youth in each grade range that might complete the survey. Programs had to include some kind of team-based activities to be included in the sample. Again, there was difficultly finding programs interested in participating in the field test to create a sampling frame in which to draw programs from, so the sample ended up being any eligible program that was interested. Because of the experience during the

pilot test with some programs ending up not administering the survey, a larger sample was recruited for the field test than the target sample size of 800 to 1,000 youth. The survey was sent to the 65 program sites that signed up for the field test. These programs provided approximate sample size estimates, which equaled a total sample size estimate of over 2,700 youth. As described in the "sample characteristics," only 40 program sites completed the field test, and within those programs a total of 1,041 youth took the survey (which ended up being 959 surveys after the data were cleaned). These 40 program sites had alone estimated a total of around 2,000 youth, but did not reach those numbers for various reasons such as program estimates not matching the number of youth who actually ended up participating in the program, varying attendance by youth at STEM OST program, and most likely other reasons not mentioned by programs. See "sample characteristics" below for additional details of the final field test sample.

Programs were recruited for the pilot test by connecting to STEM OST providers across the country. Individuals from the Phase 1 sampling frame were emailed, as well as people who had originally expressed interest during the NSF grant proposal needs assessment. Participants from previous phases of the study were contacted about the field test, as well as people who had expressed interest participating in the study during previous phases but ended up not being able to participate. Additionally, recruitment information was also shared through a wide variety of listservs, discussion boards, LinkedIn groups, Twitter feeds, the project's Facebook page, and other Facebook pages of other relevant organizations (see Appendix C for a list of the places recruitment information was shared during each phase of the research).

Sample characteristics. A total of 1,040 youth completed the survey during the field test. However, the data were cleaned so that any youth who did not answer all of the team communication survey items were removed from the sample (this removal rule did not apply to how youth answered the demographic questions). After removing these youth from the sample, the final sample size for the pilot test was 959 youth.

The pilot test sample included a diversity of youth (see Table 22). Youth spanned all grade levels, with the fewest percent of youth in 5th grade (5%). Overall, more high school youth participated in the

field test than middle school youth. The three youth who indicated “other” for grade had the following responses, “But I do 7th grade math,” “I am advanced,” and “I’m in 10th for math and science but in different grades for other subjects.” There were more youth in the sample who identified as girls than boys. The sample included youth from a variety of racial/ethnic groups, with white youth being the largest group (45%). The youth who only checked “other” wrote in a wide variety of race/ethnicities including Middle Eastern, Indian, Italian, Irish, Pacific Islander, and Russian. Around two-thirds of the youth had a mother that had graduated college.

Table 22

Demographic Characteristics of Youth in the Field Test Sample

Demographic variable	Percentage of youth
Grade (*n* = 951)	
5th grade	5%
6th grade	10%
7th grade	9%
8th grade	10%
9th grade	13%
10th grade	16%
11th grade	20%
12th grade	17%
Other	<1%
Gender (*n* = 908)	
Female	54%
Male	45%
Transgender	1%
A gender identity other than male or female	1%
Race/Ethnicity (*n* = 876)	
White or Caucasian	45%
Asian	19%
African-American/Black	12%
Multi-racial	13%
Hispanic/Latino	8%
Native American/American Indian/Alaskan Native	1%
Other	2%
Mother's education level (*n* = 942)	
Did not finish high school	4%
Graduated high school	9%
Some education after high school	10%
Graduated college	67%
I don't know	10%

A change to a survey question happened during the field test as a result of an email from a field test site. A program provider sent a message from a youth in her program that included a thoughtful discussion around the gender question for transgender youth. The youth was concerned about error that could be introduced into the survey by transgender youth who were unsure how to respond to the

question. For instance, a transgender girl who identified as a girl would not have known if she should check girl or transgender. Since the intent of the question was to make it inclusive by ensuring that youth who may not identify as a boy or girl could see themselves in the options, the transgender option was replaced with "an identity other than male or female" for the rest of the field test.

Youth were also asked how long they had been in their STEM OST program. As described in the pilot test, this question was used to get a sense of the exposure youth had to the team-based experiences in their program and gauge the extent to which the survey was tested with youth of varying levels of skill, with the assumption that youth who have less exposure to the STEM OST program may perceive themselves as having lower levels of team communication skills, be less comfortable with the skills, and be less likely to exhibit them on a team. As illustrated in Table 23, the sample included youth with a wide range of experiences from less than a month to over a year.

Table 23

Length of Time Youth Participated in Their STEM OST Program (n = 947)

Length of time	Percentage of youth
Less than a month	29%
1 to 6 months	24%
7 to 12 months (a year)	7%
Over a year	40%

One-way ANOVAs comparing the total factor scores of youth with varying lengths of participation in their STEM OST program were computed. A significant difference was found among length of time for four out of the five factors (Information Exchange Factor 2 was not significant). Tukey's *HSD* was used to determine the nature of differences between length of time participating in a STEM OST. The analysis revealed that across the four factors, youth who had been in their programs for less than a month has significantly ($p < 0.05$) lower factor scores than youth who had been in their program for over a year.

A total of 40 STEM OST program sites from across the country chose to participate in the field test. One sites had four different programs participate in the field study so there were 43 total programs in the sample. Program sites were located in 18 states across the United States: California, Connecticut, Florida, Illinois, Indiana, Massachusetts, Michigan, Minnesota, Missouri, New Mexico, New York, North Carolina, North Dakota, Pennsylvania, South Carolina, Utah, West Virginia, and Wisconsin. The sample also included two program sites in British Columbia, Canada. Youth participated in programs from a wide range of organizations, including science museums, community-based organizations, a history museum, natural history museums, a research institute, Boys and Girls clubs, 4-H, a statewide FIRST robotics organization, schools, nature/environmental centers, a library, a university, aquariums, zoos, and botanical gardens. Some youth came from programs that covered a variety of STEM topics, while others came from programs focused on specific areas of STEM such as computer science, genetics, space science, robotics, marine science, engineering and design, making, health science, natural science, and conservation. STEM OST program types included clubs, citizen science programs, camps, teen volunteer programs, youth development programs, mentorship programs, youth employment programs, robotics and science competition programs (including FIRST Lego League, FIRST Robotics, and Science Olympiad), internship programs, teen advisory groups, teen leadership programs, and general afterschool/out-of-school programs.

Results. Phase 4 was the final phase of validation for the instrument. All of the results for the Phase 4 field test are presented in the following Results chapter. This includes the results from the confirmatory factor analysis, item analysis, reliability analysis, and DIF analysis. The meaning and implications of these results as well as areas of potential additional study are discussed in Chapter 5.

Chapter 4: Results

This chapter lays out the results from the field test of the team communication skills instrument. Analyses included confirmatory factor analysis, reliability analysis, item analysis, descriptive statistics, and DIF analysis.

Confirmatory Factor Analysis

As suggested from the exploratory factor analysis described in Chapter 3, the hypothesized structure was a five-factor model (see Chapter 3 for more details of this model). The five-factor structure was fit by conducting confirmatory factor analysis. As illustrated in Figure M1 in Appendix M, the model had 66 freely estimated model parameters (23 factor loadings, 28 error variances, five factor variances, and 10 factor covariances) and was overidentified with 340 *df*. The results for the fit indices are reported in Table 24. The fit indices reported were chosen based on guidance from Johnson (2011) around indices to pay attention to for confirmatory factor analysis of ordinal data.

Table 24

Fit Indices for the Hypothesized Five-Factor Model

Fit index	Value
Absolute fit	
Chi-square value	2,447.26
df	340
p-value	.00
SRMR	.063
Parsimony correction	
RMSEA	.080
90% confidence interval	.077 to .083
Comparative fit	
CFI	.92

The following guidelines for absolute fit, parsimony correction, and comparative fit were used to interpret the results of the fit indices. For absolute fit, the standardized root mean square residual (SRMR) was used. For the SRMR, good model fit is defined as ≤ .08 (Hu & Bentler, 1999). The χ^2 value was also

reviewed, but as stated in Brown (2015), large sample sizes, such as the sample size (n = 959) in this validation study, can cause the χ^2 value to be inflated and thus significant, which would suggest rejecting the model. For this reason, researchers tend to refrain from relying on the χ^2 value to evaluate model fit and put more weight in the other fit indices (Brown, 2015). The parsimony correction index used was the root mean square error of approximation (RMSEA). As suggested by Hu and Bentler (1999), good fit is defined as RMSEA ≤ .06. The 90% confidence interval for the RMSEA value also provides support for good fit when the upper number of the interval is below .06 (Brown, 2015). Browne and Cudeck (1993) provide additional guidance that a RMSEA value less than .08 is considered adequate fit. For comparative fit, the comparative fit index (CFI) was used. Hu and Bentler (1999) define good fit as CFI ≥ .95. However, values between .90 and .95 may suggest that the model's fit could be considered acceptable (Brown, 2015).

Using these guidelines, the five-factor model had overall good to fair fit. The SRMR value suggested good model fit. However, the RMSEA value was on the edge of being adequate fit, and when looking at the confidence interval, the upper limit (.083) was above the cutoff value for adequate fit. The CFI value suggested acceptable, but not good, model fit.

Although the SRMR suggested good fit, the RMSEA and CFI did not. This meant the next step was to identify areas of potential misfit by reviewing the standardized residuals and modification indices (Brown, 2015). The standardized residuals ranged from to -0.27 to 0.15. None of the residuals were above or equal to the absolute value of 1.96, which is considered the cut off for identifying model parameters that may substantially underestimate or overestimate the relationship between two indicators (Brown, 2015). The residuals did not help to identify areas of misfit so the modification indices were reviewed. Modification indices suggest where the fit of the model can be improved by allowing parameters whose modification indices are at or above 3.84 to be freely estimated (Brown, 2015). As illustrated in Appendix N, there were 147 modification indices above 3.84. The largest modification index was 144.38 for the

relationship between the error variances of 9aGood and 9aLikely. These two items had a similar anchor statement, "Encouraging your teammates to ask you questions about your idea to make sure they understand it correctly." 9aGood asked youth how good or bad they would be at doing this, while 9aLikely asked youth how likely or unlikely they would be to do this on the imaginary team. To improve model fit, the correlated error between these two items could be freely estimated (Brown, 2015). However, they were not the only items that had a structure with a similar anchor statement. All of the information exchange and closed-loop communication items had an anchor statement with three follow up questions (good, comfort, likely) related to the anchor statement. These results suggested a method effect for items that had overlap in content that needed to be accounted for in the model (Brown, 2003). This is referred to as shared method variance and can be accounted for by allowing errors to covary (Cole, Ciesla, & Steiger, 2007). The addition of correlated errors to a model can account for these method effects and help to explain covariation between indicators that is not accounted for by covariation explained by the latent dimension (Brown, 2003, 2015; Gerbing & Anderson, 1984). The five-factor confirmatory factor analysis model was run again allowing not only the errors for 9aGood and 9aLikely to be correlated, but the errors of all items that had the method effect of a shared anchor statement. This is illustrated in the revised path diagram in Figure M2 in Appendix M. This revised model had 90 freely estimated model parameters (23 factor loadings, 28 error variances, five factor variances, 10 factor covariances, and 24 error covariances) and was overidentified with 316 *df*.

When correlated errors were added to the model, the overall fit of the model improved. As illustrated in Table 25, the fit indices all suggested good model fit with the SRMR < .08, the RMSEA value and the 90% confidence interval less than .06, and CFI > .95. The standardized residuals ranged from to -0.09 to 0.12, which did not suggest any item misfit. The modification indices were reviewed again. Although there were 89 modification indices above 3.84 (the highest modification index was 36.75), there was not a theoretical reason to include any additional freely estimated parameters in the model. Since the shared method variance had already been accounted for with the correlated errors among

items sharing an anchor statement and the resulting model had good fit, additional changes were not made to the model.

Table 25

Comparison of Fit Indices for the Five-Factor Model Without and With Correlated Errors

Fit index	Without correlated errors	With correlated errors
Absolute fit		
Chi-square value	2,447.26	846.23
df	340	316
p-value	.00	.00
SRMR	.063	.038
Parsimony correction		
RMSEA	.08	.04
90% confidence interval	.077 to 0.083	.038 to .045
Comparative fit		
CFI	.92	.98

Parameter estimates were then calculated for the five-factor model with the correlated errors. As illustrated in Table 26, factor loadings ranged from .59 to .82 and were statistically significant ($p < .001$). All of the factor loadings were above the recommended cutoff of .30, with 21 of the 28 factor loadings at or above .70, which can be considered a "very strong" relationship between the item and its associated factor (Furr & Bacharach, 2014). As illustrated by the R^2 values (.34 to .67), an adequate amount of the variance of each item was explained by the associated factor (Brown, 2015). The variance-covariance matrix that includes the parameter estimates for the model's error variance and covariance is included in Appendix O.

Table 26

Standardized Parameter Estimates for the Five-Factor Model with Correlated Errors

Factors and Items	Factor loading (λ)	*SE*	R^2
Closed-loop Communication 1			
3aGood	.78	.02	.60
3aComfort	.77	.02	.60
3aLikely	.68	.03	.47
5aGood	.82	.02	.67
5aComfort	.82	.02	.68
5aLikely	.75	.03	.56
Closed-loop Communication 2			
8aGood	.73	.02	.53
8aComfort	.74	.02	.55
8aLikely	.69	.03	.48
9aGood	.70	.02	.49
9aComfort	.76	.02	.58
9aLikely	.66	.02	.44
Information Exchange 1			
1bGood	.73	.03	.53
1bComfort	.79	.02	.62
1bLikely	.74	.03	.55
6bGood	.65	.03	.42
6bComfort	.70	.03	.50
6bLikely	.59	.03	.34
Information Exchange 2			
3bGood	.71	.02	.51
3bComfort	.71	.02	.51
3bLikely	.67	.03	.45
4bGood	.72	.02	.52
4bComfort	.79	.02	.62
4bLikely	.69	.03	.48
Listening			
1c	.74	.03	.55
2c	.77	.03	.60
3c	.77	.02	.60
9c	.76	.03	.58

Discriminant validity was also reviewed by looking at the correlations between the five factors. As illustrated in Table 27, correlations between factors ranged from $r = .46$ to $r = .78$. Poor discriminant

validity is defined as a factor correlation that is larger than $r = .80$ or $r = .85$ (Brown, 2015). Since all of the factor correlations were below this, the discriminant validity of the five factors was good.

Table 27

Correlation Matrix for the Five Factors

	CLoop1	CLoop2	InfoE1	InfoE2	Listen
Closed-loop Communication 1	1.00				
Closed-loop Communication 2	.73	1.00			
Information Exchange 1	.66	.78	1.00		
Information Exchange 2	.63	.71	.75	1.00	
Listening	.50	.54	.48	.46	1.00

Reliability

Reliability was calculated using coefficient alpha. Coefficient alpha was first calculated using the traditional method where the observed total score variance used in calculating alpha was based on the sum score of the individual items in the test or in this case the factor. However, the traditional calculation of coefficient alpha does not take into account the correlated errors that were added to the model. Davenport, Davison, Liou, and Love (2016) propose using total scores based on parcels instead of individual items to calculate coefficient alpha when there are correlated errors in the model. In the case of this validation study, the parcel is the sum of three items with a shared anchor statement and the total score for the factor that is used in calculating coefficient alpha is the sum of the parcels. For example, for Information Exchange Factor 1 the sum of the three 6b items is parcel 1, the sum of the three 1b items is parcel 2, and the total score for a person for that factor is computed by adding the two parcel scores.

As illustrated in Table 28, both with and without parcels, each factor score was reliably measuring the associated construct area as evidenced by the coefficient alpha estimates ranging from $\alpha = .79$ to $\alpha = .88$ without parcels and $\alpha = .70$ to $\alpha = .79$ with parcels. All of the coefficient alpha values were above $\alpha = .70$, which is considered an acceptable value (DeVellis, 2012).

Table 28

Coefficient Alpha Estimates for Each Factor Without and With the Use of Parcels

Factor	Coefficient alpha without parcels	Coefficient alpha with parcels
Closed-loop Communication 1	.88	.77
Closed-loop Communication 2	.86	.71
Information Exchange 1	.83	.72
Information Exchange 2	.88	.70
Listening	.79	.79

Note. Parcels were not computed for the Listening Factor since the factor did not have any correlated errors between items, which is why coefficient alpha did not change for the Listening Factor.

Although factor analysis did not find the construct of team communication to be unidimensional and it is recommended to report results of the survey based on the five factors, some evaluators may still decide to report a total score for the survey. Reliability for the entire instrument was high both without parcels ($\alpha = .93$) and with the use of parcels ($\alpha = .85$).

Item Analysis and Descriptive Statistics

Item analysis was carried out for the items in each factor. Discrimination (item-total correlation) was calculated based on the total score of the factor the item belonged to (Meyer, 2014). Discriminations ranged from .54 to .72 (see full results in Appendix P). Since all of the discriminations were above .30, the items were discriminating well between youth with high and low scores within a given factor (Meyer, 2014). Item difficulty, or mean of the item, was calculated for each item to indicate how difficult an item was where the lower the value, the more difficult the item was to endorse (Meyer, 2014). Difficulties ranged from 2.69 to 3.41. The highest difficulties (or easiest items to endorse) were in the Information Exchange 1 factor, while the lowest difficulties (or hardest items to endorse) were in the Information Exchange 2 factor. These findings were also reflected in the frequencies of each response option for an item (see Appendix P). The Information Exchange 1 factor had items with the lowest percentage of youth choosing the lowest response option (1% for 6bGood and 6bComfort). The Information Exchange 2

factor had the item (3bComfort) with the lowest percentage of youth choosing the highest rating (25%). As evident in the pattern of responses (see Appendix P), all of the items were negatively skewed. However, even though they were skewed, none of the items had what might be considered a ceiling effect, with 54% being the highest percentage of youth selecting the highest rating option (for item 5aGood).

Factor Descriptive Statistics

Total score descriptive statistics and histograms were run for each of the five factors. The information exchange and closed-loop communication factors all had six items, which meant they had a possible total score range of 6 to 24. The Listening factor had four items, so a possible total score range of 4 to 16. As illustrated in Table 29, the mean and median total scores were similar within each factor and fell toward the higher end of the total score range. However, for all factors, the median was slightly higher than the mean sum score, suggesting negatively skewed distributions. The skewness was verified by looking at histograms of the total scores for each factor (see Figures 5 through 9). All of the histograms were negatively skewed, with the Information Exchange Factor 2 distribution closest to normal.

Table 29

Total Score Descriptive Statistics by Factor (n = 959)

Factor	*M*	*SD*	*Mdn*	Min	Max
Closed-loop Communication 1	19.68	4.00	21	6	24
Closed-loop Communication 2	18.64	4.07	19	6	24
Information Exchange 1	20.03	3.29	21	6	24
Information Exchange 2	17.29	4.42	18	6	24
Listening	12.79	2.57	13	4	16

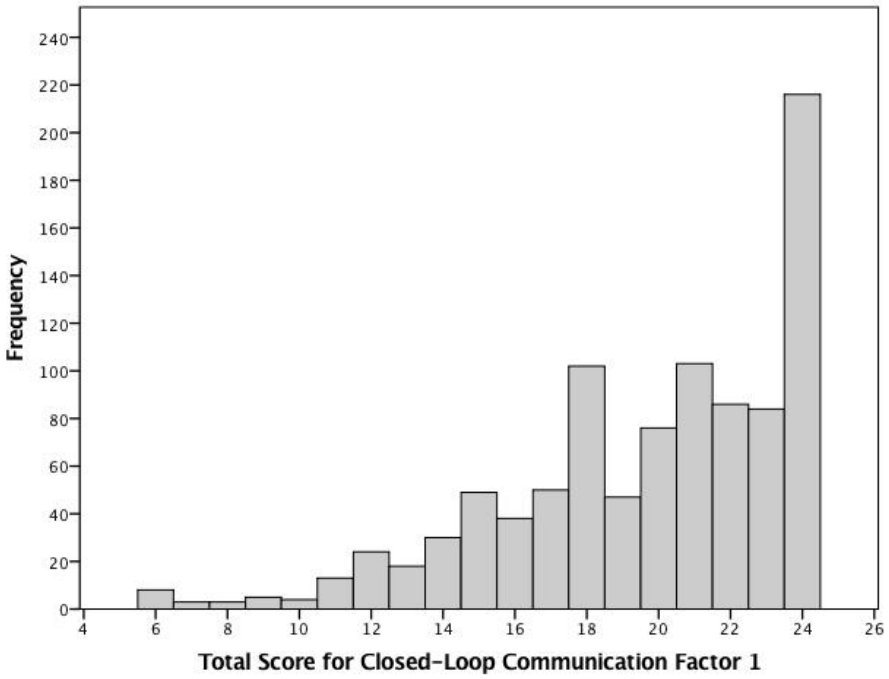

Figure 5. Total score histogram for Closed-Loop Communication Factor 1 (n = 959).

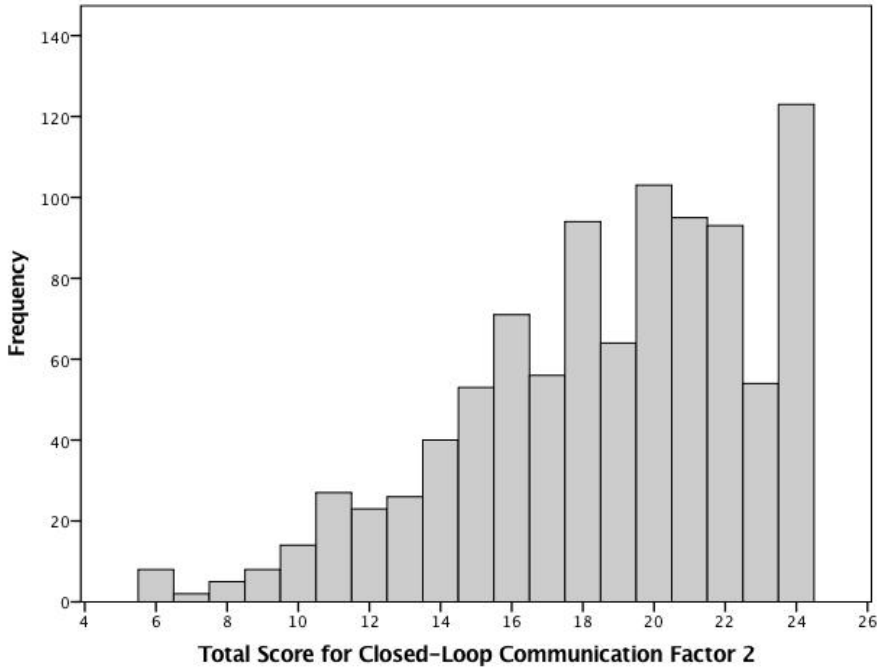

Figure 6. Total score histogram for Closed-Loop Communication Factor 2 (n = 959).

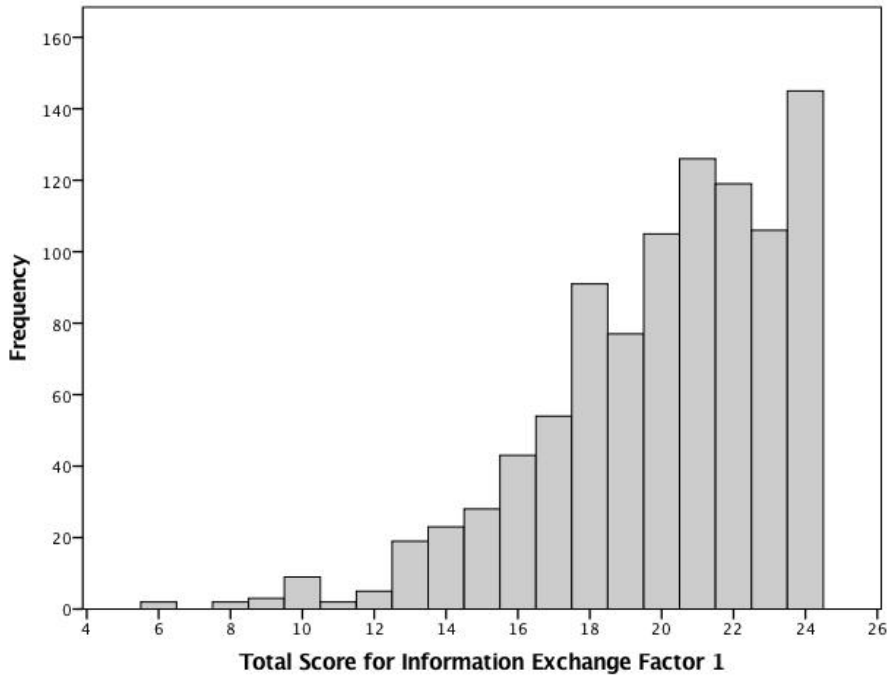

Figure 7. Total score histogram for Information Exchange Factor 1 (n = 959).

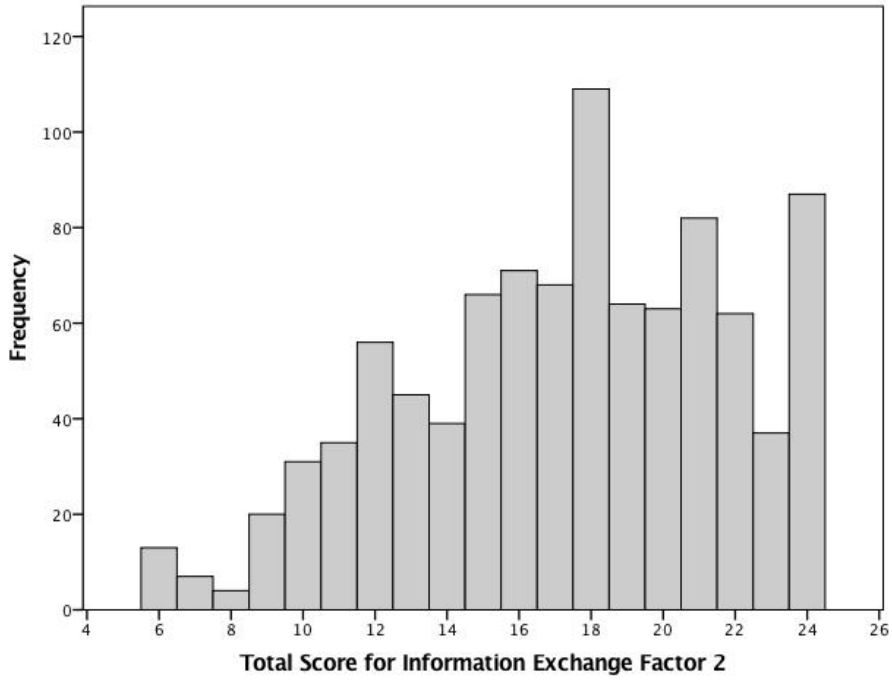

Figure 8. Total score histogram for Information Exchange Factor 2 (n = 959).

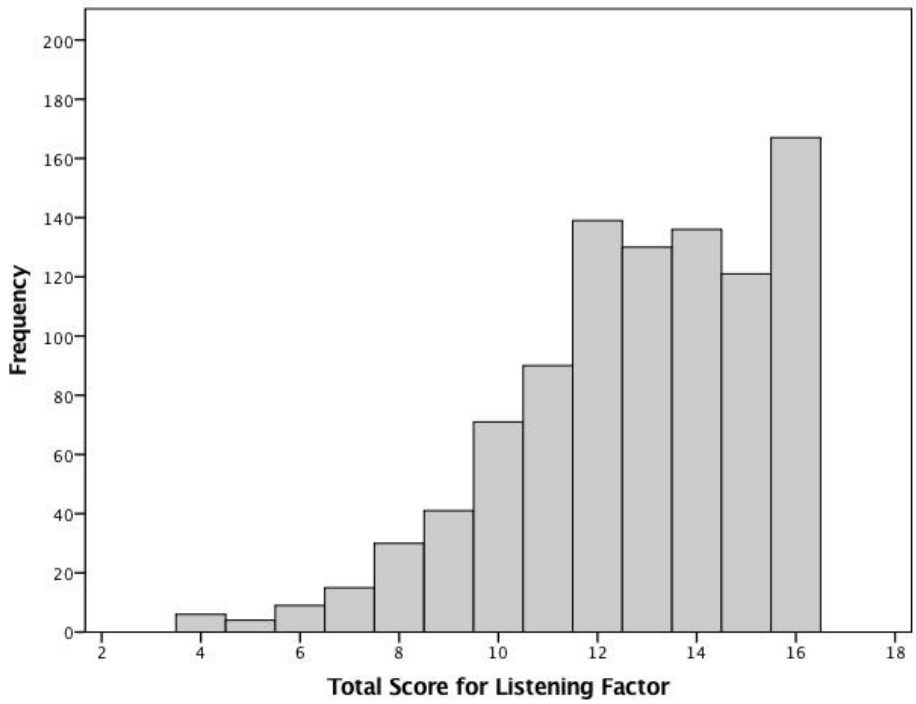

Figure 9. Total score histogram for Listening Factor (n = 959).

Differential Item Functioning

The final analysis was DIF analysis using the Mantel-Haenszel procedure to look at item fairness for different groups of youth. The analysis was first run inputting the data as four response options, however, SPSS produced the warning that "the Mantel-Haenszel Common Odds Ratio Estimate table are not computed because either (a) the group variable does not have exactly two distinct non-missing values or/and (b) the response variable does not have exactly two distinct non-missing values" (IBM Corp., 2016). For this reason, the data had to be dichotomized in order to run the Mantel-Haenszel procedure. DIF analyses were run for the variables described in Table 30.

Table 30

Variables Used for DIF Analysis and Sample Sizes for Each Group

Demographic variable	Focal group	Reference group
Gender	Girls (n = 487)	Boys (n = 410)
Race/ethnicity	African-American/Black youth (n = 113)	White youth (n = 396)
Race/ethnicity	Asian youth (n = 168)	White youth (n = 396)

DIF results were interpreted by looking at the significance of the Mantel-Haenszel χ^2 statistic and, when significant, computing the value of the ETS delta value. The first step was to identify any items in which the Mantel-Haenszel χ^2 statistic was significant ($p < .05$). Items without a significant Mantel-Haenszel χ^2 statistic were considered to not have DIF, while the ETS delta metric was then calculated for items with a significant Mantel-Haenszel χ^2 statistic to understand the extent of potential DIF (de Ayala, 2009; Zwick, 2012). The next step was to compute the ETS delta statistic, $\Delta = -2.35\ln(\alpha_{MH})$, where α_{MH} is the Mantel-Haenszel common odds ratio estimate (Zwick, 2012). The ETS uses a classification system to interpret delta (Sinharay & Doran, 2010; Zwick, 2012):

- A (negligible or non-significant DIF): non-significant Mantel-Haenszel χ^2 or the absolute value of $\Delta < 1.00$ if Mantel-Haenszel χ^2 is significant.
- B (slight to moderate DIF): $1.00 \leq$ absolute value of $\Delta < 1.50$.
- C (moderate to large DIF): absolute value of $\Delta \geq 1.50$.

The sign of the delta statistic indicates which group the item favors; a negative sign means the item favors the reference group while a positive sign means it favors the focal group (Zieky, 2003). Table 31 identifies the items that exhibited DIF (see Appendix Q for DIF results for all items based on gender and race/ethnicity). When looking at gender, DIF was only slight to moderate in relation to favoring boys. Three items had moderate to large (C) levels of DIF. In two cases the items favored the white youth and in one case the item favored Asian youth. Per guidance from ETS, for items with high levels of DIF, the

Mantel-Haenszel procedure was rerun removing the item with DIF from the calculation of the factor total score used to create the grouping variable (Zieky, 2003). This did not make a difference in the interpretation of DIF for the item for the African-American/Black youth and white youth comparison (Item 9aComfort $\chi^2 = 6.73, p = .01, \Delta = -2.14$, ETS Category = C-) or the items for the Asian youth and white youth comparison (Item 3aLikely $\chi^2 = 5.57, p = .02, \Delta = -1.85$, ETS Category = C-; Item5aGood $\chi^2 = 7.65, p = .01, \Delta = 2.29$, ETS Category = C+).

Table 31

Results for Items that Exhibited DIF

	χ^2	*p*	Δ	ETS category
Girls (n = 487) & Boys (n = 410)				
9c. Stay focused on the conversation your team is having instead of letting your mind wander. (Listening Factor)	4.81	.03	-1.29	B-
African-American/Black (n = 113) & White (n = 396)				
9aComfort: Encouraging your teammates to ask you questions about your idea to make sure they understand it correctly. (CLC Factor 2)	5.34	.02	-2.25	C-
Asian (n =168) & White (n = 396)				
3aLikely: Asking your teammate to explain his or her idea in a different way so you can understand it better. (CLC Factor 1)	8.08	.004	-2.46	C-
5aGood: Asking your teammate to repeat his or her idea because you are unsure if you understood it correctly. (CLC Factor 1)	7.46	.01	2.57	C+

Note. The negative sign means the item is more difficult for the focal group (African-American/Black youth, Asian youth, or girls) to endorse, while the plus sign means the item is more difficult for the reference group (boys or white youth) to endorse (Zieky, 2003).

A way to help gauge if the DIF items should be considered for removal is to compare the mean factor scores with and without the DIF item(s). Independent sample t-tests were carried out to identify if there was a statistically significant difference between the mean factor scores of the focal and reference

groups with and without the C category item included. If there was a significant difference, effect sizes would be compared and a dramatic change in effect size would be additional evidence to remove the DIF item from the scale. As illustrated in Table 32, there was not a statistically significant difference between focal and reference group means either with or without the DIF items included in the total factor score.

Table 32

Mean Comparison of Factors With and Without Items Exhibiting DIF by Subgroups

	Focal M	*SD*	*Ref. M*	*SD*	*t*	*p*
African-American/Black (n = 113) & White (n = 396)						
Closed-loop Communication Factor 2 with 9aComfort	19.01	3.93	18.47	4.22	-1.22	.22
Closed-loop Communication Factor 2 without 9aComfort	15.84	3.27	15.48	3.47	-0.99	.32
Asian (n = 168) & White (n = 396)						
Closed-loop Communication Factor 1 with 3aLikely and 5aGood	19.42	3.91	19.86	3.78	1.24	.22
Closed-loop Communication Factor 1 without 3aLikely & 5aGood	12.79	2.68	13.16	2.60	1.53	.13

Note. The focal groups were African-American/Black youth and Asian youth, while the reference group was white youth.

An additional consideration is that differential item functioning analysis requires running multiple statistical tests of the same data set. Running multiple tests can increase the occurrence of Type 1 errors and lead to a false significant finding, which in this case would mean finding DIF for an item when there is no DIF (Kim, 2010).

One way to account for this is to use the Bonferroni correction. The Bonferroni correction takes into account the number of tests that are run. In this case, there were 28 tests for each of the three DIF analyses for a total of 84 tests. The Bonferroni correction uses a new significance level that is calculated by dividing the original alpha value ($\alpha = 0.05$) by the total number of tests, which in this case would be 0.05/84 = 0.0006 (Thompson, 2006). Using this new significance value where the *p*-value must be *p*

≤ .0006 to be considered significant, none of the Mantel-Haenszel X^2 statistics were significant, meaning none of the items across the three different comparisons exhibited DIF. Additionally, with a $p = 0.05$ significance level, 5% of the tests would be significant simply by chance. As illustrated in Table 31, there were four items with DIF which is the number we would expect to be significant by chance alone (5% of 84 tests is four). This provides additional evidence for using the Bonferroni correction which shows there are no items with DIF.

Chapter 5: Discussion

There is a national call for more measures to assess the wide variety of 21st century skills necessary to prepare today's youth for the STEM workforce. One skill area of particular importance is teamwork skills as STEM fields have become increasingly collaborative (National Research Council, 2015). STEM OST programs are key environments for youth from diverse backgrounds to learn and practice teamwork skills, but as described in Chapter 2 there is a need for tools appropriate for assessing these skills. This National Science Foundation-funded research project addressed the lack of instruments with validity evidence to measure teamwork skills of middle and high school youth who participate in STEM OST programs by developing and validating an instrument focused on the construct of team communication skills. The project was guided by four research questions that were answered through a rigorous four-phase instrument development and validation process based on standards from the field of educational measurement (American Educational Research Association et al., 2014).

1. What skill area of the broad range of teamwork skills best aligns with middle and high school youth outcomes in STEM OST programs and the evaluation needs of these programs?
2. Is the construct of team communication skills unidimensional or multidimensional?
3. To what extent does the developed instrument gather reliable data from youth in STEM OST programs?
4. To what extent is there adequate validity evidence for the developed instrument?

The process ensured that the team communication skills survey gathered reliable data and had validity evidence for use with the diverse middle and high school youth populations participating in STEM OST programs that have team-based activities.

The final instrument is a self-report survey that measures youths' perceived team communication skill level and their comfort, ease, and likelihood of using the skill. The instrument is composed of an imaginary teamwork scenario that provides the framing for responding to the survey's questions. The

survey includes 28 items that fall into five factors based on the team communication skill construct areas of closed-loop communication, information exchange, and listening. All items have response options based on a four-point scale. The closed-loop communication and information exchange items are structured so that there is an anchor statement with three follow up questions where youth rate how good or bad they would be at doing the skill, how comfortable or uncomfortable they would be doing it, and the likelihood they would do it on the imaginary team. For the listening factor, youth respond to each statement in terms of how easy or hard it would be to do that skill. What follows is a summary of the findings that answer the research questions and led to the development of this instrument.

Summary of Findings

Research question 1: Teamwork skill area of focus. The first research question was, "*What skill area of the broad range of teamwork skills best aligns with middle and high school youth outcomes in STEM OST programs and the evaluation needs of these programs?"* As described in Chapter 2, most of the literature detailing teamwork skills comes from fields outside of out-of-school time such as formal education and organizational psychology. In order to create a measure with content validity evidence for the way teamwork skills are operationalized in STEM OST programs and ensure its usefulness, it was important to ground the instrument and construct of measurement in the way teamwork skills are operationalized in STEM OST program. The construct of team communication skills was selected because communication was the skill most frequently addressed during interviews with middle (75%) and high school (91%) STEM OST programs. The definition of team communication was then operationalized by pulling from the STEM OST interviews and literature. The final definition of team communication skills was composed of three construct areas: (a) closed-loop communication, (b) information exchange, and (c) listening.

Closed-loop communication. Closed-loop communication is the communication process between the sender and receiver of a message to make sure a message is communicated, received, and understood (Johnson & Johnson, 2013; McIntyre & Salas, 1995; Salas et al., 2009). The sender needs to ensure that

the other person received the message, interpreted it correctly, and encourage the receiver to clarify understanding by repeating back what they heard and asking clarifying questions; while the receiver needs to acknowledge they have received the information, repeat back what they heard, and if necessary, ask clarifying questions to make sure they fully understand what is being communicated (Dickinson & McIntyre, 1997; Johnson & Johnson, 2013; McIntyre & Salas, 1995; Rosen et al., 2013; Salas et al., 2009).

Information exchange. The exchange of information and ideas is key for a team to effectively work together toward a common goal (Aube et al., 2014). Team members each have important knowledge and ideas and part of the skill is knowing when to share them, what is important to share, and doing so without being asked (Salas, 2013; Smith-Jentsch et al., 2008; Smith-Jentsch, Johnston, et al., 1998). Of particular emphasis is recognizing and sharing "unique" information with the team, even if it differs from what someone else has shared (Salas, 2013). In some cases, closed-loop communication is actually included within the definition of information exchange, which suggests that they may not be distinct construct areas (American Institutes for Research, n.d.; Salas, 2013).

Listening. As described by Baker et al. (2004), the ability to listen effectively is an important part of strong team communication skills. Listening includes the skills of knowing when to listen, being an active listener, balancing listening and speaking, and avoiding interrupting teammates (Greenstein, 2012; Organization for Economic Cooperation and Development, 2005). Listening was an important part of the construct for STEM OST programs, as it came up in many of the interviews but listening did not have much of an emphasis as a distinct skill in the organizational psychology literature.

Research question 2: Dimensionality. While defining the team communication construct, a research question around internal structure validity emerged: *"Is the construct of team communication skills unidimensional or multidimensional?"* Since team communication skills had three construct areas, it raised the question if team communication was one construct or multiple constructs, and, if it were multiple constructs, were information exchange and closed-loop communication distinct? It was also

unclear if the good, comfort, and likely items would load together or fall into different factors. These questions were answered through both exploratory and confirmatory factor analyses. From the exploratory factor analysis results, a five-factor solution was selected because it made the most sense empirically and theoretically for the constructs being measured. The five factors aligned with the three construct areas and had two factors each for closed-loop communication and information exchange.

1. Closed-loop Communication Factor 1 related to a teammate's idea.
2. Closed-loop Communication Factor 2 related to youth's own idea.
3. Information Exchange Factor 1 related to youth sharing their information/idea with the team.
4. Information Exchange Factor 2 related to youth bringing up an idea that might be more difficult to share.
5. Listening Factor.

In all instances in which there was an anchor statement with three items, the good, comfort and likely items all loaded on the same factor.

Confirmatory factor analysis was then run to gather internal structure validity evidence around the dimensionality of the construct. The five-factor model, with the addition of correlated errors between the good, comfort, and likely items for the closed-loop communication and information exchange anchor statements, was found to have good fit as suggested by the values of the fit indices (SRMR < .08, the RMSEA value and the 90% confidence interval less than .06, and CFI > .95). Although there were still high modification indices that suggest possible areas of misfit, there was not a theoretical reason to include any additional freely estimated parameters in the model since the shared method variance had already been accounted for with the correlated errors among items sharing an anchor statement.

Research question 3: Reliability. The third research question asks, *"To what extent does the developed instrument gather reliable data from youth in STEM OST programs?"* Reliability of responses was first addressed through the development and testing of the items and then checked statistically by computing coefficient alpha. An initial means to help avoid sources of measurement error was to develop

items based on survey design guidelines and then test the items with potential audiences (Dillman et al., 2014). STEM OST providers first provided feedback on the items and scenario and brought to light any interpretation problems youth might experience. The survey was then tested with youth from the intended population to ensure that items were clearly interpreted and measuring what was intended. Any items where youth experienced confusion or had multiple interpretations were removed or reworded to help improve the reliability of survey responses. Reliability of responses for the field test data were calculated using coefficient alpha with parcels to account for the correlated errors in the model (Davenport et al., 2016). Both with and without parcels, each of the five factor scores were at or above $\alpha = .70$ (range without parcels $\alpha = .79$ to $\alpha = .88$, range with parcels $\alpha = .70$ to $\alpha = .79$), providing evidence of the reliability of the factor scores for the five factors.

Research question 4: Validity evidence. The final research question was, *"To what extent is there adequate validity evidence for the developed instrument?"* Various types of validity evidence were collected in order to make the argument that the interpretation of factor scores on the survey are indeed measures of youth's perception of their level of the skill area covered by the factor, their comfort and likelihood of performing the skill, and, in the case of the listening factor how easy it is for them to use the skill.

The validity argument includes construct validity evidence based on content, response process, and internal structure (American Educational Research Association et al., 2014; Messick, 1995). The validity argument did not include validity evidence based on relations to other variables (American Educational Research Association et al., 2014). Ideas for gathering this kind of evidence are discussed in the Future Directions section.

Content–related validity evidence. Content-related validity evidence was gathered to look at the alignment of the survey items and scenario with the construct areas of team communication skills. A first step was to review the literature and example items related to the construct areas while developing the

instrument to ensure that the construct was accurately represented. The STEM OST providers then provided feedback to ensure that the survey's scenario and items were relevant to the way teams are used in their programs, the team communication skills their programs address, and their evaluation needs. Many of the STEM OST providers said the scenario was similar to their program, but they had a number of suggested changes that were incorporated to better align the scenario with STEM OST team experiences.

Additional content validity evidence was gathered from the advisors from the grant that funded this research. The advisors had expertise in measurement, teamwork science, and youth development. The advisors' feedback was important for gathering evidence related to the construct areas of interest and the content (scenario and items) being used to measure that construct areas. The advisors provided feedback on the three construct areas of team communication, and because of their feedback the originally proposed team communication construct area of task-related assertiveness was removed and replaced with information exchange. This helped to ensure that the construct of team communication skills was more in line with how the field of teamwork research currently thinks about the operationalization of team communication. In addition, the advisors provided feedback on the alignment of the individual items to the three construct areas. In some cases, this involved revision, removal, or addition of items to better ensure that a construct area, was adequately covered by the items. At other times an item may have been aligned with the wrong construct area and an advisor indicated what area it should be part of instead. The advisors also had suggestions about the scenario to make it more in alignment with key characteristics of teams. Even though the teamwork science expert suggested removing listening skills as it is not an area focused on in the organization psychology field, the construct area stayed in the survey to ensure that the construct of team communication aligned with the way STEM OST programs define team communication skills in their programs. Additional content validity evidence was gathered from the teamwork science expert after the Phase 3 pilot test to ensure that the final items aligned with the construct areas they were intended to measure and the construct areas were adequately covered by the items. The expert felt that all

of the items aligned with the construct areas identified and the way he defined the three construct areas in his work was represented by the items. He felt even though the scales were short, each of the items were a characteristic of the construct and he did not feel anything was missing.

Response process validity evidence. Response process validity evidence was gathered through think-aloud interviews to ensure that the way youth were interpreting the items aligned with the intention of what the item was meant to measure in relation to team communication skills (American Educational Research Association et al., 2014). Data from think-alouds helped to inform decisions on how to increase the validity and reliability of responses through scenario and item revisions. The only necessary change to the scenario was to stress that youth had just met their teammates so all of the youth were imagining the same type of relationship with their imaginary teammates. Items were revised to ensure that they were being interpreted as intended, were easy to understand, and were clearly measuring the skill area of interest. In some cases, items were removed because they did not align with the skill areas they were meant to measure, meaning the items were not valid measures of the construct area.

Internal structure validity evidence. Internal structure validity evidence came from confirmatory factor analysis and DIF analysis (American Educational Research Association et al., 2014). As described for Research Question 2, the team communication construct was found to be multidimensional. Through confirmatory factor analysis, a five-factor model, with the addition of correlated errors between the good, comfort, and likely items for the closed-loop communication and information exchange anchor statements, was found to have good fit as suggested by the values for the fit indices (SRMR $< .08$, the RMSEA value and the 90% confidence interval less than .06, and CFI $> .95$). This provides evidence that the internal structure is five distinct factors related to the construct areas of team communication. This evidence of internal structure supports the use of individual factor scores for each of the five factors, instead of one total score for the survey. Users of the instrument will be provided with instructions on how to calculate an average score for a factor by adding the individual values of the response options (1 to 4) chosen for each item within a factor. This approach of using coarse factor scores is common in

applied research, as well as evaluation of informal education experiences (American Camp Association, 2013; Brown, 2015; Phillips, Porticella, Bonney, & Grack Nelson, 2015; Science Learning Activation Lab, 2016). While there are limitations to using this approach as it treats each item as contributing equally to measuring the construct instead of weighting an item based on the factor loading, it makes it easier for evaluators to interpret evaluation results for the wide range of stakeholders using the findings (Brown, 2015).

Additional internal structure validity evidence was collected using DIF analysis to look at item fairness for different groups of youth. When looking at gender, DIF was only slight to moderate in relation to favoring boys on one item. Three items had moderate to large (C) levels of DIF. In one case an item favored white youth over African-American/Black youth, in a second case the item favored white youth over Asian youth, and in a third case the item favored Asian youth over white youth. However, when looking at a comparison of mean factor scores with and without the DIF items, there was not a statistically significant difference between focal and reference group means either with or without the DIF items included in the total factor score for the three items. Additionally, when accounting for multiple tests using the Bonferroni correction, none of the items were found to have DIF across the three groups of comparisons.

Limitations

Although the instrument was tested by a diverse range of programs, a limitation of the research is that STEM OST programs were not drawn from a sampling frame for the think-aloud interviews, pilot test, or field test. Instead, it was a convenience sample of all the programs that expressed interest in participating. As described in Chapter 3 and Appendix C, there were extensive efforts to recruit programs to participate. The timing of data collection could have affected some programs' ability to participate as a few programs sent emails that they were interested but unable to participate because they were not able to fit it into programming at that time or the program had not started yet or had just ended. The length of the think-aloud interviews (up to an hour) and pilot test administration (around 20 minutes) could have also

played a factor with challenges of recruitment. However, obtaining a sampling frame may not always be a realistic goal for gathering large amounts of data from a diverse range of STEM OST programs. This research brought to light the difficulty in trying to use a sampling frame when large numbers of youth are needed within a timeframe of a few months. Extending the timeline for data collection and cold calling programs to invite them to participate could have helped to increase the sample size and possibility of having a sampling frame to pull programs from. However, a major issue during this research was that there were quite a few programs that ended up having to drop out of data collection or never ended up following through with administering the survey. This significantly affects the final sample size when a program that said they could administer 50 to 100 surveys does not collect any. Program estimates of the number of youth who could take the survey were also quite different from the number of completed surveys for those programs. As described in Chapter 3, the 40 programs that ended up participating in the field test estimated that there could be up to 2,000 youth responses, but only 1,041 took the survey and of these 959 ended up having complete data. It may not be possible or even ideal to collect data from STEM OST programs using a sampling frame knowing that extent of attrition that might happen when the desire is to get a diverse, large sample within a short (2- to 3-month) time frame.

Another sample-related limitation is the low numbers of fifth graders in the sample. Fifth graders were the smallest group of youth in the samples for the think-alouds (1 out of 30), pilot test (7%), and field test (5%). Even with some targeted recruitment messages for middle school programs, the sample for fifth grade remained small. It could be that there are just fewer STEM OST programs for fifth graders or fifth graders participate in lower numbers in STEM OST programs than youth in other graders. From the programs that did end up participating, there were a few emails from STEM OST providers with a concern about the amount of reading for fifth graders, and in some cases providers decided to read the scenario out loud. The limited testing and validation work with fifth graders should be kept in mind if evaluators or STEM OST programs want to use this instrument with fifth grade youth. They should consider doing some additional validation work themselves such as think-alouds with fifth grade youth.

This is also an area of potential future research in general, to collect more validity evidence for the use of this instrument with fifth graders.

An additional limitation is that this is not a direct measure of team communication skills: it is an instrument measuring youths' perceived skill level and their comfort, ease, and likelihood of using the skill. As described in Chapter 2, situational judgment tests are often used to measure skills in a survey format but that format was not chosen for this instrument because a "test" does not fit the informal nature of STEM OST programs. Additionally, STEM OST programs vary widely in their content, activities, and context so to create scenarios relevant to a diverse range of program types would have been difficult. For this reason, a self-report survey was developed and validity evidence was gathered to support the use of it as an instrument for youth to reflect on how they perceive their skills and think about what it might be like for them to participate in an imagined team scenario. Observations are an additional way to measure teamwork skills directly and could be used to triangulate findings obtained from this survey.

An additional limitation of the instrument is that the distribution of responses for the items tends to be negatively skewed. Some items are more difficult to endorse than others, but overall items tend to be easy to endorse as evident by the range of difficulty for the items (2.69 to 3.41 on a four-point scale). Even with the skewness, none of the items have what might be considered a ceiling effect, with only four of the 28 items having 50% or more of the youth selecting the highest response option. This means the scale allows room to see a spread of youth across the five factors (as illustrated in the distribution of total scores for each factor in Figures 5 through 9 in Chapter 3), as well as measure improvement in programs across the factors if using this as a pre/post measure.

Implications

Implications for evaluators. The results of this research will potentially be of great value to the evaluation field. This instrument begins to fill a gap for evaluators wanting to measure teamwork skills in STEM OST programs. Many of the conversations and publications around 21st century skills assessments, which include teamwork, are related to formal education. This presents a challenge for evaluators who are

working with informal education programs because there are few appropriate tools for evaluators to use. This research resulted in an instrument that evaluators can be confident will gather reliable data, has adequate validity evidence for use with STEM OST programs, is grounded in what actually occurs in these programs, and provides data that a wide range of programs will find useful.

The results of defining the construct in Chapters 2 and 3 are also valuable as they provide evaluators with a deep understanding of how STEM OST programs use teams and the teamwork skills they address. As was evident in Chapter 2, definitions of teamwork skills in the 21st century skills and OST literatures are lacking in detail. This research provides a deeper understanding of what kinds of teamwork skill outcomes STEM OST programs have and provides definitions of the areas of team communication skills that evaluators can use to inform their use of this instrument as well as the development of other evaluation instruments. Additionally, operationalizing the definitions of team communications skills within the context of middle and high school STEM OST programs helps evaluators better understand and evaluate the impact programs have on these outcomes areas.

Implications for STEM OST practitioners. The results of this study will also be of great potential value to STEM OST practitioners. The survey findings that evaluators share with STEM OST stakeholders will be useful for their programs. From a formative evaluation perspective, the results will allow programs to identify necessary areas of improvement around team communication that they can then address through program activities. From a summative evaluation perspective, survey findings can be used to understand the overall impact a program has on the development and comfort level of youth in relation to the team communication skill areas.

The instrument will not only be useful for evaluators, it will be useful for STEM OST practitioners to evaluate their own program if they do not have funds for a formal evaluation. The guide created for this survey (see the Future Directions section for more information on the guide) will include detailed instructions for both evaluators and practitioners on how to administer the survey and use the

results. The experience testing the instrument with STEM OST programs throughout this research will inform the development of the guide to ensure that practitioners have the information and support they need to be able to successfully use the instrument.

The results of defining the construct of team communication skills in Chapters 2 and 3 are also valuable for STEM OST programs. As discussed earlier, definitions of teamwork skills in the 21st century skills and OST literatures are lacking in detail. This research provides STEM OST programs with common language and definitions around team communication skills that are grounded in the theoretical and empirical literature and built on extensive input from STEM OST practitioners. As one provider mentioned while reviewing the instrument and construct areas in Phase 2, "I learned a lot about the skills involved in teamwork." A few providers talked about the value of having a sort of "checklist" of communication skills that can guide their work with youth and help them better address teamwork skills in their program.

Future Directions

This instrument will be disseminated broadly to evaluators and STEM OST providers to help ensure its use. A guide will be created to accompany the instrument that outlines the test characteristics, includes definitions of the construct areas, describes details about survey administration, and provides guidance on how to score the survey and interpret the scores. The guide will be written for both an evaluator and STEM OST provider audience. As during the pilot and field test, programs will be able to personalize the survey so they can include their program name and appropriate grade range in the scenario and indicate if youth are working on a project or challenge. The survey and guide will be freely available online so that it is easily accessible and can be widely used.

Validation is an ongoing process and additional validity evidence could still be gathered for this instrument (American Educational Research Association et al., 2014). The instrument lacks validity evidence based on relations to other variables. One way this could be addressed is by gathering

convergent validity evidence to look at the relationship of the instrument's scores to other measures such as observations of youth working in teams or educator ratings (American Educational Research Association et al., 2014). This could happen by comparing scores on the team communication survey to observations of team communication that occurs in STEM OST programs as youth work together on projects and challenges. Team communication survey data could also be compared to ratings educators give youth on their team communication skills. For instance, an educator could complete the team communication survey for a youth by imagining how they think that youth would respond to the items based on their experiences observing that youth's work on teams. Even though the Bonferroni correction indicated that there were no items with DIF, there may be a desire to still gather additional internal structure validity evidence to ensure the fairness of items for Asian and African-American/Black youth. This could be done by gathering larger samples of Asian and African-American/Black youth since the sample sizes were right near the suggested minimum for calculating DIF (minimum for focal group 100, reference group 400) (Paek & Guo, 2011). Think-aloud interviews could also be conducted with youth from the focal and reference groups to look for any differences in how youth interpret and respond to questions based on their race/ethnicity to identify if there actually is any unfairness in the item and if so, what about the item's wording might be causing that unfairness. Additionally, DIF could be calculated using parcels for each group of items in a factor (similar to how reliability was calculated) instead of based on individual items to see the effect this has on detecting DIF. Finally, response process validity evidence could also be collected for the instrument. As described in the Limitations section, because of the small sample of fifth grade youth in this study, additional response process validity evidence could be gathered from this population to help ensure that the scenario and items are appropriate for fifth grade youth and easy for them to understand.

The instrument was developed and tested for use with STEM OST programs. However, there are a wide range of non-STEM OST programs with teamwork skill outcomes that may be interested in using this instrument. The scenario and items are written broadly enough that they may work also work for OST

programs in other content areas. Additional validation work could be done with other types of OST programs to ensure the construct areas, scenario, and items align with how teams are used in those settings and the team communication skills they address.

This research focused on designing a survey and testing it in programs at one time point. Programs decided when during their program schedule they administered the survey so the sample included both youth who had participated in STEM OST programs for a short period of time and youth who had been participating for years. This was intentional for the purpose of testing to help gather variability in responses with the assumption that youth who had been in a program for a short amount of time may have rated themselves lower than someone who had been in a program longer and presumably exposed to more teamwork experiences. Since many STEM OST programs expressed interest in using this as a pre-/post- summative evaluation measure to look at program impact and the instrument was not tested for that use, additional validation work should be done around that particular use of the instrument. This would give insight into how well the instrument measures change over time in youths' perceptions of their skill level, and their comfort, ease, and likelihood of using the skill.

There are various ways evaluators can think about analyzing the data from the survey. One analysis that might be of interest for a multi-site evaluation is the use of hierarchical linear modeling (HLM). Through interviews with STEM OST providers and think-aloud interviews with youth, it was clear that there were differences across programs in how they structure teams and address teamwork skills. However, within programs there were similarities in the way some youth talked about their teamwork experiences during think-aloud interviews, suggesting that youths' team communication factor scores may be dependent on the program they are in. For this reason, a HLM of youth nested within programs may be useful to consider when measuring team communication outcomes across programs for an evaluation.

This instrument is only one measure of the range of teamwork skills that STEM OST providers cover in their programs. There is an opportunity for evaluators and researchers to continue developing

measures to assess the other teamwork skill areas, as well as other 21st century skills. There is a call in both the STEM OST and larger informal science education fields for more measures of shared outcomes (Ellenbogen, 2014; Krishnamurthi, Ballard, & Noam, 2014; Noam & Shah, 2013; Sacco, 2014). Future research can address this need, helping to increase the capacity of evaluators to measure important outcomes of STEM OST programs more rigorously.

Bibliography

21st Century Readiness Act, H.R. 2536, 112th Cong., 1st Sess. (2011).

Afterschool Alliance. (2013). *Defining youth outcomes for STEM learning in afterschool*. Washington, DC: Author.

Afterschool Alliance. (2015). *America after 3PM. Full STEM ahead: Afterschool programs step up as key partners in STEM education*. Washington, DC: Author.

Aguado, D., Rico, R., Sánchez-Manzanares, M., & Salas, E. (2014). Teamwork Competency Test (TWCT): A step forward on measuring teamwork competencies. *American Psychological Association, 18*(2), 101-121.

American Camp Association. (2013). *Camp Youth Outcomes Battery: Measuring developmental outcomes in youth programs*. (2nd ed.). Martinsville, IN: Author.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Institutes for Research. (n.d.). *Evidence-base: Communication*. Washington, DC: Author.

Assessment and Teaching of 21st Century Skills. (2012). About the project. Retrieved from http://www.atc21s.org/about.html

Aube, C., Brunelle, E., & Rousseau, V. (2014). Flow experience and team performance: The role of team goal commitment and information exchange. *Motivation and Emotion, 38*, 120-130.

Baker, D. P., Horvath, L., Campion, M., Offermann, L., & Salas, E. (2004). *Teamwork: Status memorandum*. Ottawa, Ontario: Statistics Canada.

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly, 71*(2), 287-311.

Bedwell, W. L., Fiore, S. M., & Salas, E. (2011). *Developing the 21st century (and beyond) workforce: A review of interpersonal skills & measurement strategies*. Orlando, FL: University of Central Florida.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw & E. Care (Eds.), *Assessment and teaching of 21st century skills*. (pp. 17-66). New York, NY: Springer.

Brown, T. A. (2003). Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects? *Behaviour Research and Therapy, 41*, 1411-1426.

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

Browne, M. W., & Cudeck, R. (1993). Alternate ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. (pp. 136-162). Newbury Park, CA: Sage.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. In R. A. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations*. (pp. 333-380). San Francisco, CA: Jossey-Bass, Inc.

Casner-Lotto, J., & Barrington, L. (2006). *Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century US workforce*. New York, NY: The Conference Board, Partnership for 21st Century Skills, Corporate Voices for Working Families & Society for Human Resource Management.

catme.org. (2013). Comprehensive Assessment of Team Member Effectiveness (CATME) Peer Evaluation. Retrieved from https://info.catme.org/wp-content/uploads/CATME_brochure_-_8_5x11_2013.pdf

Center for Health Science Interprofessional Education Research and Practice. (2012). *Definition of behavioral markers for Performance Assessment for Communication and Teamwork (PACT)*. Seattle, WA: University of Washington.

Chiu, C. J. (2014). *Development and validation of Performance Assessment Tools for Interprofessional Communication and Teamwork (PACT)* (Unpublished doctoral dissertation). University of Washington, Seattle, WA.

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insiduous effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods, 12*(4), 381-398.

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*(6), 1092–1122.

Critical Thinking, Collaboration, Communication, and Creativity for Careers Act, H. R. 5663, 114th Cong., 2nd Sess. (2016).

Crown, L., Fine, W., & Beller, L. (n.d.). A step by step guide to how a community collaboration called MHA Labs is creating comprehensive and engaging assessments to support 21st century youth. Retrieved from http://mhalabs.org/wp-content/uploads/2014/04/mha_steps_poster_v1.pdf

Davenport, E. C., Jr., Davison, M. L., Liou, P.-Y., & Love, Q. U. (2016). Easier said than done: Rejoinder on Sijtsma and on Green and Yang. *Educational Measurement: Issues and Practice, 35*(1), 6-10.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: The Guildford Press.

DeVellis, R. F. (2012). *Scale development: Theory and application*. Thousand Oaks, CA: Sage.

Dickinson, T. L., & McIntyre, R. M. (1997). A conceptual framework for teamwork measurement. In M. T. Brannick, E. Salas & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications*. (pp. 19-43). Mahwah, NJ: Lawrence Erlbaum Assoicates, Inc.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. (4th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Ellenbogen, K. (2014). *Summary of the CAISE convening on building capacity for evaluation in informal science, technology, engineering and math (STEM) education*. Washington, DC: Center for Advancement of Informal Science Education (CAISE).

Every Hour Counts. (2014). *Every Hour Counts measurement framework*. New York, NY: Author.

FIRST LEGO League. (2017). Core values. Retrieved from http://firstlegoleague.org/mission/corevalues

Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics: An introduction* (2nd ed.). Thousand Oaks, CA: Sage.

Gerbing, D. W., & Anderson, J. C. (1984). On the meaning of within-factor correlated measurement error. *Journal of Consumer Research, 11*, 572-580.

Gong, Y., Kim, T., Lee, D., & Zhu, J. (2013). A multilevel model of team goal orentation, information exchange, and creativity. *Academy of Management Journal, 56*(3), 827-851.

govtrack.us. (2014). H.R. 2536 (112th): 21st Century Readiness Act. Retrieved from http://www.govtrack.us/congress/bills/112/hr2536

govtrack.us. (2017a). H.R. 347 (113th): 21st Century Readiness Act. Retrieved from https://www.govtrack.us/congress/bills/113/hr347

govtrack.us. (2017b). H.R. 5663 (114th): Four C's for Careers Act. Retrieved from https://www.govtrack.us/congress/bills/114/hr5663

Grack Nelson, A. (2013). *Pre-proposal needs assessment survey*. Unpublished raw data.

Grack Nelson, A. (2016). Assessing team communication skills in STEM afterschool programs [Blog post]. Retrieved from https://naaweb.org/professional-development/item/587-assessing-team-communication-skills-in-stem-afterschool-programs

Greenstein, L. (2012). *Assessing 21st century skills: A guide to evaluating mastery and authentic learning*. Thousand Oaks, CA: Corwin.

Griffin, P., Care, E., & McGaw, B. (2012). *Assessment and teaching of 21st century skills*. New York, NY: Springer.

Hand, J. (2016). Loebsack, Costello reestablish bipartisan congressional 21st Century Skills Caucus. Retrieved from https://loebsack.house.gov/news/documentsingle.aspx?DocumentID=394855

Hansen, D. M., & Larson, R. (2002). *The Youth Experience Survey 1.0: Instrument development and testing*. Urbana, IL: University of Illinois, Urbana - Champaign.

Hansen, D. M., & Larson, R. (2005). *The Youth Experience Survey 2.0: Instrument revisions and validity testing*. Urbana, IL: University of Illinois, Urbana - Champaign.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

IBM Corp. (2016). IBM SPSS Statistics for Mac (Version 24). Armonk, NY: IBM Corp.

Institute of Museum and Library Services. (2009). *Museums, libraries, and 21st century skills*. Washington, DC: Author.

Johnson, D. W., & Johnson, F. P. (2013). *Joining together: Group theory and group skills*. (11th ed.). Boston, MA: Pearson Education, Inc.

Johnson, G. M. (2011). *A consideration of issues related to the confirmatory factor analysis of the Evaluation Use and Evaluation Involvement scales* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis, MN.

Kim, J. (2010). *Controlling Type 1 error rate in evaluating differential item functioning for four DIF methods: Use of three procedures for adjustment of multiple item testing*. (Unpublished doctoral dissertation). Georgia State University, Atlanta, GA.

Klem, M. D., & Nicholson, D. J. (2008). Proven effectiveness of Missouri 4-H camps in developing life skills in youth. *Journal of Youth Development, 2*(3), 133-139.

Krishnamurthi, A., Ballard, M., & Noam, G. G. (2014). Examining the impact of afterschool STEM programs. Retrieved from http://www.afterschoolalliance.org/ExaminingtheImpactofAfterschoolSTEMPrograms.pdf

Kyllonen, P. (2012). *Measurement of 21st century skills within the common core state standards*. Princeton, NJ: Educational Testing Service.

Lai, E. R., & Viering, M. (2012). *Assessing 21st century skills: Integrating research findings*. Vancover, British Columbia: National Council on Measurement in Education.

Loughry, M. L., Ohland, M. W., & Moore, D. D. (2007). Development of a theory-based assessment of team member effectiveness. *Educational and Psychological Measurement, 67*(3), 505-524.

McDaniel, A. M., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. I. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60*, 63-91.

McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. A. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations*. (pp. 9-45). San Francisco, CA: Jossey-Bass Inc.

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

Meyer, J. P. (2014). *Applied measurement with JMetrik*. New York, NY: Routledge.

MHA Labs. (2017). MHA Labs. Retrieved from http://mhalabs.org/

MHA Labs. (n.d.). *Human Achievement Quotient (HAQ) assessment*. Chicago, IL: Author.

National Center for Education Statistics. (2012). *Improving the measurement of socioeconomic status for the National Assessment of Educational Progress: A theoretical foundation*. Washington, DC: Author.

National Center for Education Statistics. (2015). *Grade 8 sample questions: General information about The Nation's Report Card*. Washington, DC: Author.

National Research Council. (2004). *The engineer of 2020: Visions of engineering in the new century*. Washington, DC: The National Academies Press.

National Research Council. (2009a). *Engineering in K-12 education: Understanding the status and improving the prospects*. Washington, DC: The National Academies Press.
National Research Council. (2009b). *Learning science in informal environments*. Washington, DC: The National Academies Press.
National Research Council. (2010). *Exploring the intersection of science education and 21st century skills: A workshop summary*. Washington, DC: The National Academies Press.
National Research Council. (2011). *Assessing 21st century skills: Summary of a workshop*. Washington, DC: The National Academies Press.
National Research Council. (2012a). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: The National Academies Press.
National Research Council. (2012b). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.
National Research Council. (2012c). *Report brief. Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: The National Academies Press.
National Research Council. (2015). *Enhancing the effectiveness of team science*. Washington, DC: The National Academies Press.
National Science Board. (2012). *Science and engineering indicators 2012*. Arlington, VA: National Science Foundation.
Newman, I., Lim, J., & Pineda, F. (2013). Content validity using a mixed methods approach. *Journal of Mixed Methods Research, 7*(3), 243-260.
NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: The National Academies Press.
Noam, G., & Shah, A. M. (2013). *Game-changers and the assessment predicament in afterschool science*. Cambridge, MA: Program in Education, Afterschool, and Resiliency.
Norland, E., Foutz, S., & Krabill, M. (2009). *Youth astronomy apprenticeship: An initiative to promote science learning among urban youth and their communities*. Edgewater, MD: Institute for Learning Innovation.
Organization for Economic Cooperation and Development. (2005). *The definition and selection of key competencies: Executive summary*. Paris, France: Author.
Osborne, J. W. (2014). *Best practices in exploratory factor analysis*. Retrieved from https://www.researchgate.net/profile/Jason_Osborne2/publication/265248967_Best_Practices_in_Exploratory_Factor_Analysis/links/5405ec140cf23d9765a79c54/Best-Practices-in-Exploratory-Factor-Analysis.pdf
Paek, I., & Guo, H. (2011). Accuracy of DIF estimates and power in unbalanced designs using the Mantel-Haenszel DIF detection procedure. *Applied Psychological Measurement, 35*(7), 518-535.
Partnership for 21st Century Learning. (2015). *P21 framework defintions*. Washington, DC: Author.
Partnership for 21st Century Learning. (2017). Our history. Retrieved from http://www.p21.org/about-us/our-history
Partnership for 21st Century Skills. (2003). *The road to 21st century learning: A policymakers' guide to 21st century skills*. Washington, DC: Author.
Partnership for After School Education. (2010). *Afterschool youth outcomes inventory*. New York, NY: Author.
Patton, M. Q. (2002). *Qualitative research and evaluation methods*. (3rd ed.). Thousand Oaks, CA: Sage.
Pearsall, M. J., & Ellis, A. P. J. (2006). The effects of critical team member assertiveness on team performance and satisfaction. *Journal of Management, 32*(4), 577-594.
Phillips, T., Porticella, N., Bonney, R., & Grack Nelson, A. (2015). *Interest in Science Scale. Technical brief series*. Ithaca, NY: Cornell Lab of Ornithology.
Prichard, J. S., Bizo, L. A., & Stratford, R. J. (2006). The educational impact of team skills training: Preparing students to work in groups. *British Journal of Educational Psychology, 76*(1), 119-140.

Rosen, M. A., Schiebel, N., Salas, E., Wu, T. S., Silvestri, S., & King, H. B. (2013). How can team performance be measured, assessed, and diagnosed? In E. Salas & K. Frush (Eds.), *Improving patient safety through teamwork and team training*. (pp. 59-79). New York, NY: Oxford University Press.

Rosenfeld Halverson, E., & Sheridan, K. M. (2014). The maker movement in education. *Harvard Educational Review, 84*(4), 495-504.

Sacco, K. (2014). Measuring learning across ISE projects. [Blog post]. Retrieved from http://www.informalscience.org/news-views/measuring-learning-across-ise-projects

Salas, E. (2013). *The science of teamwork: What matters in practice*. Paper presented at the Society for Human Resource Management Conference, Chicago, IL.

Salas, E., Rosen, M. A., Burke, C. S., & Goodwin, G. F. (2009). The wisdom of collectives in organizations: An update of the teamwork competencies. In E. Salas, G. F. Goodwin & C. S. Burke (Eds.), *Team effectiveness in complex organizations. Cross-disciplinary perspectives and approaches*. (pp. 39-79). New York, NY: Taylor & Francis Group.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in teamwork? *Small Group Research, 36*(5), 555-599.

Schwarz, E., & Kay, K. (2006). The case for 21st century learning [Special issue]. *New Directions for Youth Development, 2006*(110).

Schwarz, E., & Stolow, D. (2006). Twenty first century learning in afterschool. *New Directions for Youth Development, 2006*(110), 81-99.

Science Learning Activation Lab. (2016). Measures technical brief: Scienfitic Sensemaking. Retrieved from http://www.activationlab.org/wp-content/uploads/2016/03/Sensemaking-Report-3.2-20160331.pdf

Science Museum of Minnesota. (2017). Kitty Andersen Youth Science Center 2015/2016 program year report. Retrieved from https://www.smm.org/sites/default/files/public/attachments/kaysc-2016report.pdf

Sibthorp, J., Bialschki, M. D., Morgan, C., & Browne, L. (2013). Validating, norming, and utility of a Youth Outcomes Battery for recreation programs and camps. *Journal of Leisure Research, 45*(4), 514 - 536.

Silva, E. (2008). *Measuring skills for the 21st century*. Washington, DC: Education Sector.

Sinharay, S., & Doran, N. J. (2010). Two simple approaches to overcome a problem with the Mantel-Haenszel statistic: Comments on Wang, Bradlow, Wainer, and Muller (2008). *Journal of Educational and Behavioral Statistics, 35*(4), 474-488.

Smith-Jentsch, K. A., Cannon-Bowers, J. A., Tannenbaum, S. I., & Salas, E. (2008). Guided team self-correction: Impacts on team mental models, processes, and effectiveness. *Small Group Research, 39*(3), 303-327.

Smith-Jentsch, K. A., Johnston, J. H., & Payne, S. C. (1998). Measuring team-related expertise in complex environments. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training*. (pp. 61-87). Washington, DC: American Psychological Association.

Smith-Jentsch, K. A., Zeisig, R. L., Acton, B., & McPherson, J. A. (1998). Team dimensional training: A strategy for guided team self-correction. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training*. (pp. 271-297). Washington, DC: American Psychological Association.

Stevens, M. J., & Campion, M. A. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of Management, 20*(2), 503-530.

Stevens, M. J., & Campion, M. A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management, 25*(2), 207-228.

The Forum for Youth Investment, & National Collaboration for Youth Research Group. (2012). *A shared vision for youth: Common outcomes and indicators*. Washington, DC: Authors.

The Secretary's Commission on Achieving Necessary Skills. (1999). *Skills and tasks for jobs: A SCANS report for America 2000*. Washington, DC: U.S. Department of Labor.

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York, NY: The Guilford Press.

Thorndike, R. M., & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education*. (8th ed.). Boston, MA: Pearson Education, Inc.

ThoughtSTEM. (2014). After-school coding programs - ThoughtSTEM. Retrieved from http://thoughtstem.com/teams

Wang, L., MacCann, C., Zhuang, X., Liu, O. L., & Roberts, R. D. (2009). Assessing teamwork and collaboration in high school students: A multimethod approach. *Canadian Journal of School Psychology, 24*(2), 108-124.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wilson-Ahlstrom, A., DuBois, D., Ji, P., & Hillaker, B. (2014). *From soft skills to hard data: Measuring youth program outcomes*. Washington, DC: The Forum for Youth Investment.

Wilson-Ahlstrom, A., Yohalem, N., DuBois, D., & Ji, P. (2011). *From soft skills to hard data: Measuring youth program outcomes*. Washington, DC: The Forum for Youth Investment.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036-1039.

Zhuang, X., MacCann, C., Wang, L., Liu, L., & Roberts, R. D. (2008). *Development and validity evidence supporting a teamwork and collaboration assessment for high school students*. Princeton, NJ: Educational Testing Service.

Zieky, M. (2003). *A DIF primer*. Princeton, NJ: Educational Testing Service.

Zwick, R. (2012). *A review of ETS differentical item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement*. Princeton, NJ: Educational Testing Service.

Appendix A: Protocol for Interviews with STEM OST Providers

Thank you for talking to me today about your STEM out-of-school time program. As a reminder, I'm talking to educators to understand the various ways programs use teamwork in their activities and how they define teamwork skills in their program. We are not judging programs in any way, just collecting information so we can understand how teamwork looks in a wide variety of STEM OST programs. These interviews will be useful for creating surveys that measure teamwork and collaboration skills in STEM OST programs.

As a reminder, your participation in this interview is completely voluntary. You can decide to skip any questions or stop the interview at any time. All of your responses will be stored in a secure location and our research team will only see your interview script.

Do you have any questions at all about the study?

On your consent form, you indicated that it was okay to audio record this conversation. Is that still okay? (If they have middle and high school programs, remind them that this interview is just going to focus on one grade range.)

Description of program

1. What is the focus of your programming?
2. What would you say are the overall goals of your program?
3. How would you describe the demographics of the youth you reach through your program?
4. What ages do you reach? How do you define your high school programs? What are the age cut offs?
5. How would you describe the length of your program? How many contact hours do you typically have with individual youth?
    a. Is there a fee for your program? Do youth get paid? Is there an incentive to participate?
6. How many youth are in your program? (Probe for how many in the program at one time)

Now I'm going to ask you about teamwork and collaboration. For this interview, I'm using the words interchangeably to both mean working together in a team.

7. How do you use teams in your program? What is the purpose of using teams?
    a. What role does the educator play in setting up the team experience? Helping during the team experience?
    b. How many youth do you typically have in a team?
    c. Do you have particular roles on the team? Do you assign roles?
8. What activities do you do in your program that require teamwork or collaboration among youth?
9. What skills are important for youth to successfully work together as a team during your program activities? What skills are necessary for them to collaborate? (Make sure to probe for them to define those skills)
10. What behaviors do you want to see happening while youth are working together as a team?
    a. How do you address those skills and behaviors in your program? Are there particular activities you do to help youth build those skills?

11. What are your program outcomes related to teamwork skills?
    a. What resources, if any, have you used to help define teamwork/collaboration outcomes and indicators?
12. What do you see in your program that suggests teamwork skill development? How do you know youth are developing these skills? What do you see, hear?
    a. Is there anything you expect to see that you haven't seen yet?
13. Have you noticed youth having any challenges working together as a team? If so, what are those challenges?
14. Do you do any type of team building or icebreaker activities? What are your goals for those activities? Purpose of those activities?

Evaluating or assessing teamwork skills.

15. Have you done any evaluation or assessment of teamwork skills in your program? If so, how do you evaluate or assess them?

*At the end of the interview - do they have program descriptions, logic model, or evaluation report they could share?

Appendix B: STEM OST Provider Interview and Review Consent Form

Collaboration in the 21st Century:
Measuring Essential Skills for the STEM Workforce

You are invited to participate in a research study to create surveys that ask questions about teamwork skills. The surveys will be used to study teamwork in science, technology, engineering or math (STEM) programs outside of school. You were selected as a possible participant because you are an educator in one of these programs. We ask that you read this form and ask any questions you may have before agreeing to participate in the study.

This study is being conducted by Dr. Frances Lawrenz in the Department of Educational Psychology at the University of Minnesota and Amy Grack Nelson in the Department of Evaluation & Research in Learning at the Science Museum of Minnesota.

Background Information
The purpose of this study is to create surveys that STEM youth programs across the country can use to assess the development of teamwork and collaboration skills in middle and high school youth.

Procedures
If you agree to participate in this study, we will ask you to participate in an interview and review survey questions and definitions.

The interview will be about teamwork and collaboration skills in your program. Interviews will be conducted in-person, by phone, or over Skype depending on your location and preference. The interview will take between 45 – 60 minutes. We would like to audio record the interview so that researchers can listen to it again in case we would like to remember something you said that might be important to the research. You can still participate in the interview if you would prefer to not be audio recorded.

A few months later, we will ask you to review draft survey questions and comment on our definitions of teamwork and collaboration skills. We will provide you with a set of questions to guide your review. The review will take 1 – 2 hours depending on the depth of your comments.

Risks and Benefits of Being in the Study
The only potential risk of participating in this research is that some of your answers to the interview questions and your review may not remain private. This is very unlikely and we will work to make sure this does not happen. We will store your interview responses, the audio recording, and your review responses on a computer that requires a password to login. Only the research staff will have access to this computer.

There is no direct benefit to subjects who participate in this study.

Compensation
You will receive $25 after completing the interview and another $25 after completing the review of the questions and definitions.

Confidentiality
Your confidentiality is important to us. The records of this study will be kept private and stored securely on a computer that requires a password to access. In any sort of report we might write, we will not include your name or any information that will make it possible to identify you.

Voluntary Nature of the Study
Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relationships with the University of Minnesota or the Science Museum of Minnesota. If you decide to participate, you are free to withdraw from the research at any time without affecting those relationships.

Contacts and Questions
The researchers conducting this study are Dr. Frances Lawrenz and Amy Grack Nelson. You may ask any questions you have now. If you have questions later, **you are encouraged** to contact Dr. Lawrenz at the University of Minnesota, [insert contact info] or Amy Grack Nelson at the Science Museum of Minnesota, [insert contact info].

If you have any questions or concerns about this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

You will be given a copy of this information to keep for your records.

Consent Statement

I have read the above information. I have asked questions and have received answers. I consent to participate in the study.

Your full name (please print): _____

Is it okay for the researcher to audio record your interview? Yes ☐ No ☐

Your signature:_____ Date: _____

Signature of researcher: _____ Date: _____

Appendix C: Advertising for STEM OST Program Recruitment by Data Collection Phase

Post	STEM OST interviews	Think-Alouds	Pilot test	Field test
4-H Youth Development Professionals LinkedIn group		x	x	x
A variety of National Science Teachers Association listservs		x	x	x
Afterschool Alliance blog post	x		x	x
Alliance for STEM Education LinkedIn group		x	x	x
American Alliance of Museums Committee on Audience Research and Evaluation Facebook page				x
American Alliance of Museums Committee on Audience Research and Evaluation LinkedIn group				x
American Alliance of Museums LinkedIn group		x		
American Educational Research Association OST group Facebook page		x		
American Evaluation Association LinkedIn group			x	
American Evaluation Association's Evaltalk listserv		x	x	x
Association of Science-Technology Centers Camp Community of Practice				x
Association of Science-Technology Centers Citizen Science Community of Practice			x	x
Association of Science-Technology Centers General forum	x	x	x	x
Association of Science-Technology Centers LinkedIn group				x
Association of Science-Technology Centers Research & Evaluation Community of Practice			x	x
Association of Science-Technology Centers Small Museums Community of Practice			x	x
Association of Science-Technology Centers STEM Afterschool Community of Practice	x	x	x	x
Association of Science-Technology Centers YOUMedia Learning Labs Network			x	x
Association of Science-Technology Centers Youth Development and Programming Community of Practice			x	x
Association of Science-Technology Centers Making and Tinkering Community of Practice				x
Associations of Zoos & Aquariums Education listserv		x	x	x
Center for the Advancement of Informal Science Education Facebook Page and Twitter	x	x		
Citizen Science listserv		x	x	x
Colorado Science Education Network			x	

Post	STEM OST interviews	Think-Alouds	Pilot test	Field test
Conservation and Environmental Education Resource Network LinkedIn group		x	x	x
EcoEd listserv		x	x	x
Environmental Education LinkedIn group		x	x	x
Florida Afterschool Network Facebook page		x		
Georgia State Afterschool Network Facebook page		x		
Gulf of Mexico Alliance Environmental Education Network listserv		x	x	x
Informal Science LinkedIn group				x
Iowa Afterschool Alliance	x			
Kansas Enrichment Network Facebook page		x		
Kentucky OST Alliance Facebook page		x		
Marine Educators' e-mail discussion list			x	x
Maryland Out of School Time Network newsletter	x			x
Missouri Afterschool Network Facebook page	x	x		
MN 21st Century Learning Centers shared with network				x
Museum Ed Facebook page				x
Museum Education Roundtable LinkedIn group		x	x	x
Museum-Ed LinkedIn group				x
Museum-Ed listserv	x	x	x	x
Nanoscale Informal Science Education Network Facebook page		x		
National Afterschool Association Facebook page		x		
National Afterschool Association LinkedIn group		x		x
National Afterschool Association blog post			x	
National Association for Environmental Education discussion board	x		x	x
National Association for Research on Science Teaching listserv	x	x	x	x
National Girls Collaborative Project Community	x	x		
National Girls Collaborative Project Facebook page		x		
National Science Teachers Association Informal Group LinkedIn group		x	x	x
National Science Teachers Association Informal Science Facebook page		x		
National U.S. Agricultural Education listserv			x	

Post	STEM OST interviews	Think-Alouds	Pilot test	Field test
Nebraska STEM Facebook page	x	x		
Network for Youth Success listserv	x		x	x
Nevada Afterschool Network shared it with their network				x
New Mexico Out of School Time Network Facebook page	x		x	
New York State Informal STEM listserv			x	x
North American Association for Environmental Education Facebook page		x		
North Carolina Environmental Education listserv		x	x	x
Oklahoma Afterschool Network Facebook page	x	x		
Oregon ASK Facebook page				x
Out of School Time Resource Network newsletter			x	
Partnership for 21st Century Learning Facebook page		x		
Schools Out Washington Facebook page	x	x		
Science Afterschool LinkedIn group		x	x	x
South Carolina Afterschool Facebook page		x		
Sprockets community			x	x
STEM Educators & Researchers LinkedIn group			x	x
STEM Learning and Research Center community	x	x		
STEM Learning and Research Center Facebook page		x		
Texas Partnership for Out of School Time	x			
Utah Afterschool Network	x			
Visitor Studies Association listserv	x	x	x	x
WATER-ED listserv			x	
Wyoming Afterschool Alliance email list	x			x

Appendix D: Teamwork Skills Coding Framework

Skills	Definitions from the literature and notes from the coding process
Leadership	***Definitions from the literature*** • "Leadership/team management – The ability to direct and coordinate the activities of other team members, assess team performance, assign tasks, motivate team members, plan and organize, and establish a positive atmosphere Subskills/Alternative labels - Task structuring (delegation and assignment, resource distribution, resource management, performance direction, establishment of priorities), mission analysis, motivation of others (leadership control, goal setting, drive to completion, goal orientation)" (Cannon-Bowers et al., 1995, p. 345). • "'Shared/distributed leadership: The transference of the leadership function among team members in order to take advantage of team members' strengths (e.g. knowledge, skills, attitudes, perspectives, contacts, and time available) as dictated by either environmental demands or the development stage of the team.' (Burke, Fiore, & Salas, 2004, p. 105). Team members…- accurately recognize and identify the member with the highest level of relevant knowledge and skill for a particular situation/problem. Team members shift leadership functions in response to changing task/environmental conditions" (Salas et al., 2009, p. 56). • "'Team leadership: Ability to direct and coordinate the activities of other team members, assess team performance, assign tasks, develop team knowledge, skills, and abilities, motivate team members to reach goals, plan and organize, and establish a positive atmosphere.' (Salas, Sims, & Burke, 2005, p.560). Team leaders… - instill shared affects and motivation and define team goals with debriefs. - Promote team learning through two-way interactions in debriefs to generate lessons learned from performance episodes. - Create team interdependencies. - Communicate a clear mission and vision for the team. - Gather and provide performance relevant information to team members. - Work to keep teams intact" (Salas et al., 2009, p. 60). • "Guide and lead others• Use interpersonal and problem-solving skills to influence and guide others toward a goal, • Leverage strengths of others to accomplish a common goal, • Inspire others to reach their very best via example and selflessness, • Demonstrate integrity and ethical behavior in using influence and power)" (Binkley et al., 2012, p. 47). • "Provide leadership and appropriate influence" (Johnson & Johnson, 2013, p. 44).
Communication	***Definitions from the literature*** 1) "The process by which information is clearly and accurately exchanged between two or more team members in the prescribed manner and with proper terminology, the ability to clarify or acknowledge the receipt of information Subskills/Alternative Labels Information exchange (closed-loop communication, information sharing, procedural talk, volunteering/requesting information), consulting with others (Effective influence, open exchange of relevant interpretations, evaluative interchange" (Cannon-Bowers et al., 1995, p. 345). 2) "Closed-loop communication/information exchange: C A pattern of communication characterized by 1) a message being initiated by the sender, 2) the message being received, interpreted, and acknowledged by the intended receiver, 3) a follow-up by the sender ensuring that the message was received and appropriately interpreted. Team members… - follow up to ensure that messages are received and understood. - Acknowledge messages when they are sent. - Cross check information with the sender to ensure that the message's meaning is understood. - Seek information from all available sources. - Provide big picture updates to one another as appropriate.

	Proactively pass information without being asked" (Salas et al., 2009, p. 61). 3) "The ability to listen to ideas of others" (Organization for Economic Cooperation and Development, 2005, p.13). 4) Communicate openly and supportively, that is, to send messages, which are behavior or event-oriented, congruent, validating, conjunctive, and owned. Listen nonevaluatively and to appropriately use active listening techniques. Maximize consonance between nonverbal and verbal messages, and to recognize and interpret the nonverbal messages of others. Engage in ritual greetings and small talk and recognition of their importance (Stevens & Campion, 1994). 5) "Interact effectively with others: Speak with clarity and awareness of audience and purpose. Listen with care, patience, and honesty" (Binkley et al., 2012, p. 47). 6) "Ensure accurate and complete communication among members" (Johnson & Johnson, 2013, p. 44). ***Notes from the coding process*** 7) In the data, we also saw listening and not interrupting as communication skills. 8) OECD's definition had "present ideas." However, coded sharing ideas as "task-related assertiveness" so there is overlap between communication and task-related assertiveness codes.
Task-related assertiveness	***Definitions from the literature*** • "'The capacity to effectively communicate in interpersonal encounters by sharing ideas clearly and directly.' (Pearsall & Ellis, 2006, p.577) Team members…- communicate task-relevant information without hesitation. - Share their opinions with others in a persuasive manner" (Salas et al., 2009, p. 58). • "The ability to present ideas" (Organization for Economic Cooperation and Development, 2005, p.13). ***Notes from the coding process*** • Also comes up as contributing ideas in the data. • This skill is related to communication.
Goal setting	***Definitions from the literature*** • "Goal setting and performance management: The KSA to help establish specific, challenging, and accepted team goals. Monitor, evaluate, and provide feedback on both overall team performance and individual team performance" (Stevens & Campion, 1994, p.505). • "Ensure each other's commitment to clear mutual goals that highlight members' interdependence" (Johnson & Johnson, 2013, p. 44). ***Notes from the coding process*** • Also talked about as having a clear goal. • In most of the programs we coded as this, someone else defines the goal. The team is not identifying the goal. This could be a difference with adult vs. youth teams.
Mission analysis	***Definitions from the literature*** • "'The interpretation and evaluation of the team's mission, including identification of its main tasks as well as the operative environmental conditions and team resources available for mission execution.' (Marks, Mathieu, & Zaccaro, 2001, p. 365) Team members...explicitly articulate the team's objectives. - Discuss the purpose of the team in the context of the present performance environment. - Discuss how the available team resource can be applied to meeting team goals" (Salas et al., 2009, p. 57). ***Notes from the coding process*** • Also talked about as what needs to be done in order to reach the goal

Situational analysis	***Definitions from the literature*** • "Shared Situational awareness – the process by which team members develop compatible models to the team's internal and external environment. Includes skill in arriving at a common understanding of the situation and applying appropriate task strategies. Subskills/alternative labels- a. Situational awareness (orientation, team awareness, development of integrated model of environment, development of system awareness), shared problem-model development" (Cannon-Bowers et al., 1995, p. 344).
Shared responsibility	***Definitions from the literature*** • "Assume shared responsibility for collaborative work" (Partnership for 21st Century Learning, 2015, p. 4). ***Notes from the coding process*** • Also comes out in the data as taking turns, everyone contributing, delegating or dividing up tasks, people taking on roles, everyone doing something.
Coordination, Project management, Planning (Combined a number of similar skills)	***Definitions from the literature*** <u>Coordination</u> • "The process by which team resources, activities, and responses are organized to ensure that tasks are integrated, synchronized, and completed within established temporal constraints Subskills/Alternative labels - Task organization (coordination of task sequence, integration), task interaction (technical coordination, response coordination), timing and activity pacing" (Cannon-Bowers et al., 1995, p. 345). • "'Coordination: The process of orchestrating the sequence and timing of interdependent actions.' (Marks, Mathieu, & Zaccaro, 2001, pp. 367-368) Team task work behaviors are sequenced so that "down time" for team members is minimized (e.g. team members don't have to wait for other team members' input to do their task work). Team members communicate information about their status, needs, and objectives as often as necessary (and not more). Team members synchronize teamwork behaviors without overt communication in high-workload conditions. Team members pass information to one another relevant to the task in a timely and efficient manner" (Salas et al., 2009, p. 59). • "Implicit coordination strategies: 'Synchronization of member actions based on unspoken assumptions about what others in the group are likely to do.' (Wittenbaum & Strasser, 1996, p.23) Team members…- compensate for increasing workload conditions by reducing the "communication overhead" (i.e. explicit communication). - sequence interdependent task work without overt communication" (Salas et al., 2009, p. 56). <u>Project Management/Planning</u> • "Planning and Task Coordination: - The KSA to coordinate and synchronize activities, information, and task interdependencies between team members, -The KSA to help establish task and role expectations of individual team members, and to ensure proper balancing of workload in the team" (Stevens & Campion, 1994, p. 505). • "Planning: The generation of a proposed sequence of actions intended to accomplish a set goal. Team members...explicitly articulate expectations for how a proposed course of action should unfold. - Explicitly define desired outcomes. - Collectively visualize how a planned course of action will be carried out and where it can go wrong. - Seek out information and feed it to fellow team members. - Share unique information" (Salas et al., 2009, p. 59). • "Prioritize, plan and manage work to achieve the intended group result" (Binkley et al., 2012, p. 47).

	Notes from the coding process • Also comes up in the data as organization. • There is some overlap of coding between here and shared responsibility. When people said they delegated or divided tasks – it was coded under shared responsibility.
Problem detection	***Definitions from the literature*** • "An initial sensing that a problem requiring attention exists or will soon exist. Team members...rapidly detect problems or potential problems in their environment. Work to determine underlying causes in conflicting knowledge. Quickly recognize a need for action when it arises. Clearly communicate problem definitions" (Salas et al., 2009, p. 57).
Problem solving/ Decision making (Some talk about collaborative problem solving).	***Definitions from the literature*** • "The KSA to identify situations requiring participative group problem solving and to utilize the proper degree and type of participation. The KSA to recognize the obstacles to collaborative group problem solving and implement appropriate corrective actions" (Stevens & Campion, 1994, p.505). • "Decision making – the ability to gather and integrate information, use sound judgment, identify alternatives, select the best solution, and evaluate consequences (in team context, emphasizes skill in pooling information and resources in support of a response choice) Subskills/Alternative labels- Problem assessment, problem solving (emergence of solutions, probabilistic structure, hypothesis formulation, information processing, information evaluation), planning (plan development, use of information), metacognitive behavior, implementation" (Cannon-Bowers et al., 1995, p. 356). • "The process of 1) identifying and representing a discrepancy between the present and desired state of the environment and 2) discovering a means to close this gap. Team members…- rapidly knowledge information when needed. - Engage in contingency planning. - Accurately recognize the internal expertise in the team and weights input accordingly. - Accurately prioritize problem features. - Dynamically assess and adjust their problem solution" (Salas et al., 2009, p. 60). ***Notes from the coding process*** • Also talked about the design process and scientific method in the data.
Flexibility/ Adaptability	***Definitions from the literature*** <u>Flexibility</u> • "Flexibly use decision-making procedures that ensures all alternative course of action receive a fair and complete hearing and that each person's reasoning and conclusions are challenged and critically analyzed" (Johnson & Johnson, 2013). • "Exercise flexibility and willingness to be helpful in making necessary compromises to accomplish a common goal" (Partnership for 21st Century Learning, 2015, p. 4). <u>Adaptability</u> • "'Ability to adjust strategies based on information gathered from the environment through the use of backup behavior and reallocation of intrateam resources. Altering a course of action or team repertoire in response to changing conditions (internal or external).' (Salas, Sims, & Burke, 2005, p. 560). Team members… - modify or replace routine performance strategies when characteristics of the environment and task change. - Detect changes in the internal team and external environments. - Make accurate assessments about underlying causes of environmental changes" (Salas et al., 2009, p. 55). • "Adaptability: The process by which a team is able to use information gathered from the task environment to adjust strategies through the use of compensatory behavior and reallocation of intrateam resources. Subskills/alternative labels - Flexibility

	(capacity for closure, development of innovations, mutual adjustment), compensatory behavior (backing-up behavior, provide/ask for assistance, fail stop), dynamic reallocation of functions" (Cannon-Bowers et al., 1995, p. 344). ***Notes from the coding process*** • Compromise is really the most relevant piece here in terms of what was seen in the interviews.
Mutual performance monitoring, Intrateam feedback (There is overlap in terms of giving feedback so we put these skills together. In one case, the literature calls Intramember feedback a subskill of mutual performance monitoring)	***Definitions from the literature*** <u>Mutual Performance Monitoring</u> • "Performance management - The KSA to monitor, evaluate, and provide feedback on both overall team performance and individual team performance" (Stevens & Campion, 1994, p.505). • "Performance monitoring/feedback – The ability of team members to give, seek, and receive task-clarifying feedback; includes the ability to accurately monitor the performance of teammates, provide constructive feedback regarding errors, and offer advice for improving performance. Subskills/Alternative labels - Intramember feedback (performance feedback, planning review, feedback/reinforcement, acceptance of/giving suggestions, criticism), mutual performance monitoring (monitoring and cross-checking, systems monitoring, performance monitoring, error identification/correction, intrateam monitoring, strategy development, procedure maintenance)" (Cannon-Bowers et al., 1995, p. 343). • "'The ability of team members to keep track of fellow team members' work while carrying out their own, to ensure that everything is running smoothly.' (McIntyre & Salas, 1995, p.23) Team members...recognize errors in their teammates' performance. - recognize superior performance in their teammates. - Offer relevant information/resources before requested. - Have an accurate understanding of their teammates' workload. - Offer feedback to their fellow team members to facilitate self-correction" (Salas et al., 2009, p. 55). <u>Intrateam Feedback:</u> • "The provision of information about team or individual performance either before, during, or after performance episode. Team members…- engage in a cycle of prebrief, performance, debrief. - Provide pre-performance information (feed forward). - Develop and integrate lessons learned from past performance. - Provide information to correct deficient performance during a performance episode. - Provide constructive and specific comments to other team members" (Salas et al., 2009, p. 58). ***Notes from the coding process*** • This relates to communication.
Backup/supportive behavior	***Definitions from the literature*** • "Ability to anticipate other team members needs through accurate knowledge about their responsibilities. This includes the ability to shift workload among members to achieve balance during high periods of workload or pressure" (Salas et al., 2005, p. 560). • "Team members… - proactively step in to assist fellow team members when needed. - Communicate the need for assistance. - can identify unbalanced workload distributions. - Redistribute workload to underutilized team members" (Salas et al., 2009, p. 55). ***Notes from the coding process*** • Programs mainly talk about asking for help and giving help.

Motivation of others	***Definitions from the literature*** • "Generating and maintaining goal directed effort toward completion of the team's mission. Team members…- encourage each other to perform better or to continue performing well. - Provide feedback regarding team successes. - Communicate beliefs of the teams' ability to succeed" (Salas et al., 2009, p. 58). ***Notes from the coding process*** • Programs mostly talk about this as encouragement. • This is only motivating others, it does not relate to self-motivation.
Value individual contributions	***Definitions from the literature*** • "The capacity to make decisions that allow for different shades of opinion" (Organization for Economic Cooperation and Development, 2005, p.13). • "Value the individual contributions made by each team member" (Partnership for 21st Century Learning, 2015, p. 4). ***Notes from the coding process*** • Also coded when they talked about respecting other people's ideas. • This skill is grouped with shared responsibility in the P21 literature. • There are some similarities between the codes Motivation of Others and Value Individual Contributions.
Conflict resolution /negotiation	***Definitions from the literature*** • "'Conflict resolution/management: Preemptive conflict management involves establishing conditions to prevent, control, or guide team conflict before it occurs. Reactive conflict management involves working through task and interpersonal disagreements among team members.' (Marks, Mathieu, & Zaccaro, 2001, p. 363) Team members...seek solutions that have mutual gains for all interests. Openly discuss task related conflict. Find it acceptable to change their minds and express their doubts" (Salas et al., 2009, p. 57) . • "The ability to manage and resolve conflicts. The ability to - analyze the issues and interests at stake (e.g. power, recognition of merit, division or work, equity), the origins of the conflict and the reasoning of all sides, recognizing that there are different possible positions. - identify areas of agreement and disagreement - reframe the problem - prioritize needs and goals, deciding what they are willing to give up and under what circumstances." (Organization for Economic Cooperation and Development, 2005, p. 13). • "Ability to negotiate" (Organization for Economic Cooperation and Development, 2005, p. 13). • "Resolve their conflicts constructively" (Johnson & Johnson, 2013, p. 44). • "Recognize and encourage desirable, but discourage undesirable, team conflict. Recognize the type and source of conflict confronting the team and to implement an appropriate conflict resolution strategy. Employ an integrative (win-win) negotiation strategy rather than the traditional (win-lose) strategy" (Stevens & Campion, 1994, p.505). ***Notes from the coding process*** • There is some overlap with communication here as communication is tied to conflict resolution skills.
Work with diverse teams	***Definitions from the literature*** • "Demonstrate ability to work effectively and respectfully with diverse teams" (Partnership for 21st Century Learning, 2015, p. 4). • "Leverage social and cultural differences to create new ideas and increase both

	innovation and quality of work" (Binkley et al., 2012, p. 47). ***Notes from the coding process*** • Examples include: Different ages - older students with younger students. Boys and girls working together. People of different racial/ethnic backgrounds. SES differences. Homeschool, public, private.
Empathy/ perspective taking	***Definitions from the literature*** • "Empathy or taking the role of the other person and imaging the situation from his or her perspective. This leads to self-reflection, when, upon considering a wide range of opinions and beliefs, individuals recognize that they take for granted in a situation is not necessarily shared by others." (Organization for Economic Cooperation and Development, 2005, p.12).
Professionalism	***Definitions from the literature*** • "Conduct themselves in a professional manner" (Binkley et al., 2012, p. 47).
Respect others (in general)	***Notes from the coding process*** • This code emerged from the data. • It was different from respecting ideas, which are coded under Value Individual Contributions. More of a general idea of being respectful to people.
Self-efficacy	***Notes from the coding process*** • This code emerged from the data. • This related to youth realizing their own strengths and what they can contribute to a team.
Other	Under the "other" code were items such as work ethic, self-worth, self-motivation, persistence, and knowing how to follow a leader.

Appendix E: STEM OST Provider Expert Review Instructions

Thank you for agreeing to review this first draft of survey questions for the Collaboration in the 21st Century (C2C) project's *Team Communication Skills* instrument. As STEM OST providers, you provide valuable first-hand insight into teamwork experiences youth have in your programs. This insight is useful as we work to make sure the survey is relevant to the teamwork communication skills addressed by your programs and your evaluation needs. You are also our first step in making sure the questions are appropriate for middle and high school youth.

What skills does this survey measure?
This survey measures team communication skills. These skills were defined based on the literature and interviews with all of you about teamwork in your programs. Team communication skills encompass a number of skill areas including task-related assertiveness or speaking up, closed-loop communication, and listening. Below is list of skills that fall under each of these skill areas. Items on the survey were created to attempt to cover this range of skills. On the survey, you will see acronyms (TRA, CLC, or L) that identify which of the three skill areas the item is meant to address.

Task-related assertiveness/Speaking up (TRA)
- Openly and clearly sharing ideas, opinions, and knowledge with teammates
- Confidently speaking up

Closed-loop communication (CLC)
- Communicating clearly
- Acknowledging receipt of information
- Asking for clarification
- Negotiating meaning
- Ensuring information is understood

Listening (L)
- Active listening
- Balancing listening and speaking
- Not interrupting

What are we asking you to do?
On the following pages, you'll find 13 STEM OST expert review questions embedded within the draft of the *Team Communication Skills* survey. The actual *Team Communication Skills* survey will be in a different font (the same font as this sentence) to help differentiate between the actual survey and the review questions for you. Answer the 13 expert review questions within this document. When you've answered the questions, email this completed document back to Amy Grack Nelson. If you have any questions about what to do for the review, please don't hesitate to send Amy an email or give her a call. We are asking you to complete the review by Friday, January 22.

We recognize the youth survey that you are reviewing has 60 statements and is really long. Don't worry; the final youth survey won't look like this! It is just part of the survey development process. We need to start with a large pool of statements and then over time revise and delete the statements that don't work well for youth. Our goal is to end up with a survey that only takes about 5-10 minutes for youth to complete and is 1-2 pages long.

STEM OST Expert Review

Your name: _____

As you review this survey, think about the [**middle or high school] program(s)** you work with that use a teamwork structure.

Please indicate the name(s) of the program(s) you were thinking about as you did this review:

_____

_____

The review starts with the survey introduction followed by expert review questions…

Team Communication Skills Survey

For the following questions, you'll be asked to imagine yourself working with a team in your afterschool program and then respond to some questions. You aren't graded at all on this and there isn't a right or wrong answer to the questions. We just want you to answer openly and honestly about how you might respond to the situation. The staff in your program won't see your individual responses; they will only get a report of everyone's information together so they won't know how you personally answered the survey questions.

Introductory Scenario

Imagine you are in your afterschool program [Insert name of program] and have just been placed in a team to work on a new project. There are four other youth on the team, so there are five of you total. There is a mix of boys and girls and everyone is in your grade. You know two of the people because you worked with them on a team before in your afterschool program, but you have just met the other two people. You will be working together over the next few weeks. Your team is given a challenge that you have to work together to solve. First, your team decides to do some research on the computer to learn more about the topic of your challenge. Then, you each share what you've found and start to brainstorm ideas of how you might solve the challenge. People share their ideas and opinions with each other and together you come up with a plan to solve the challenge. The team works together to gather the information and materials you need to put your plan into action and solve the challenge.

STEM OST expert review questions about the introductory scenario

*Questions 1 – 3 are related to the introductory scenario paragraph above.

1. Does the scenario above seem like something that might happen in your program? If not, what about it doesn't align with your program?

2. Do you feel like there might be an important part of the team experience that is missing in the scenario that might be useful to add? If so, what is missing?

3. Do you have any other comments about the scenario?

STEM OST Expert Review questions about the individual statements 1 - 60

*Questions 4 – 7 below refer to all of the likelihood, comfort, and agreement statements that you will see on the following pages. As you look at the 60 statements, you'll notice that there is a corresponding acronym beside each statement to indicate which skill area it addresses – task-related assertiveness (TRA), closed-loop communication (CLC), or listening (L). These won't be part of the final survey; they are just there for your reference so you can see how the individual statements cover the different skill areas of teamwork communication skills.

4. Do any of the statements discuss something you would <u>not</u> expect youth to be able to do in your program? If so, why isn't it an expectation of youth? Write the statement number(s) below with your comments or make comments directly to the statements(s) using track changes or by inserting comments.

5. Are there any statements that may be difficult for youth in your program to understand? If so, describe why it might be difficult. If you have an idea of how to rephrase the statement you can include that too. Write the statement number(s) below or make comments directly in the scale using track changes or by inserting comments.

6. Are there any statements that relate to a skill that you do <u>not</u> work on in your program so you wouldn't expect to see a change from a pre-to post-survey? If so, list the statement numbers below or make comments directly in the scale using track changes or by inserting comments.

7. Are there any teamwork communication skills that are missing from these statements? If so, what do you feel is missing?

How likely or unlikely are you to do the following things as part of this imagined team? Please answer openly and honestly. There isn't a right or wrong answer and you won't be graded on this.

[*Note to STEM OST reviewers*: This scale will have six options - Very Unlikely, Unlikely, Somewhat Unlikely, Somewhat Likely, Likely, Very Likely.]

How likely is it that you would…	**Skill area**
1. Explain an idea you have to the team.	TRA
2. Keep your ideas to yourself and only share them with the team after a teammate asks you if you have an idea. (Adapted from Johnson & Johnson, 2013)	TRA
3. Speak up and share an idea without first being asked by a team member.	TRA
4. Share an idea even if you think one of your teammates might dislike it.	TRA
5. Ask your teammates questions about what they think of your idea.	CLC
6. Ask your teammates if they understand your idea.	CLC
7. Ask your teammate what might be confusing about what you said if they are confused about your idea	CLC
8. Repeat your idea in a different way to help clarify what you said if a teammate is confused about your idea.	CLC
9. Listen to teammates' opinions before evaluating their positions as good or bad. (Adapted from Aguado et al., 2014)	L
10. Express your opinion about what you think the group should do to solve the challenge.	TRA

11. Provide an opinion that is different from a teammate's opinion.	TRA
12. Share an opinion even if you know others on your team will disagree with you.	TRA
13. Provide your opinion to the team without being asked.	TRA
14. Speak up and tell the team about a skill you have that would help solve the challenge.	TRA
15. Share with your team information you found about the topic of the challenge.	TRA
16. Waiting until someone in the group is done talking to share your idea, even if you have something important to say.	L
17. Listen closely to someone share their idea, instead of focusing on what you are going to say to the team about your own idea.	L
18. Speak up without hesitation to share your idea after a teammate is done talking.	TRA
19. Keep your opinions about your teammates' ideas to yourself. (Adapted from Johnson & Johnson, 2013)	TRA
20. Avoid a discussion with a teammate about their idea if it might lead to a disagreement.	TRA
21. Participate in a discussion about a teammate's idea even if it leads to disagreements.	TRA
22. Tell your teammate if something about their idea is confusing.	CLC
23. Ask your teammate to repeat what they said because something was confusing about their idea. (Adapted from Johnson & Johnson, 2013)	CLC
24. Ask your teammates questions about something that is confusing about their idea. (Adapted from Aguado et al., 2014; Center for Health Science Interprofessional Education Research and Practice, 2012)	CLC
25. Ask your teammate to explain their idea in a different way so you can understand it.	CLC
26. Repeat back an idea your teammate had to make sure you heard them correctly. (Adapted from Center for Health Science Interprofessional Education Research and Practice, 2012)	CLC
27. Keep quiet if you are unsure what someone means instead of asking them to clarify. (Adapted from Johnson & Johnson, 2013)	CLC
28. Ask a teammate for their opinion if they are quiet and not speaking up.	TRA

How comfortable or uncomfortable would you be doing the following things as a part of this imagined team? Please answer openly and honestly. There isn't a right or wrong answer and you won't be graded on this.

[Note to STEM OST reviewers: This scale will have six options - Very Uncomfortable, Uncomfortable, Somewhat Uncomfortable, Somewhat comfortable, Comfortable, Very Comfortable.]

How comfortable or uncomfortable would you be…	**Skill area**
29. Explaining an idea you have to the team.	TRA
30. Speaking up and sharing an idea without first being asked by a team member.	TRA
31. Sharing an idea even if you think one of your teammates may dislike it.	TRA
32. Asking your teammates questions about what they think of your idea.	TRA
33. Asking your teammates if they understand your idea.	CLC
34. If a teammate is confused about your idea, asking questions to understand what might be confusing about what you said.	CLC
35. If a teammate is confused about your idea, repeating your idea in a different way to help clarify what you said.	CLC
36. Expressing your opinion about what you think the group should do to solve the challenge.	TRA
37. Providing an opinion that is different from a teammate's opinion.	TRA
38. Sharing an opinion even if you know others on your team will disagree with you.	TRA
39. Providing your opinion to the team without being asked.	TRA
40. Speaking up and telling the team about a skill you have that would help solve the challenge.	TRA
41. Sharing with your team information you found about the topic of the challenge.	TRA
42. Waiting until someone in the group is done talking to share your idea, even if you have something important to say.	L
43. Listening closely to someone share their idea, instead of focusing on what you are going to say to the team about your own idea.	L
44. When someone is done talking, speaking up without hesitation to share your idea.	TRA
45. Participating in a discussion about a teammate's idea even if it leads to disagreements.	TRA
46. Telling your teammate that something about their idea is confusing.	CLC
47. Asking your teammate to repeat what they said because something was confusing about their idea. (Adapted from Johnson & Johnson, 2013)	CLC
48. Asking your teammates questions about something that is confusing about their idea. (Adapted from Aguado et al., 2014; Center for Health Science Interprofessional Education Research and Practice, 2012)	CLC
49. Asking your teammate to explain their idea in a different way so you can understand it.	CLC
50. Repeating back an idea your teammate had to make sure you heard them correctly. (Adapted from Center for Health Science Interprofessional Education Research and Practice, 2012)	CLC
51. Asking a teammate for their opinion if they are quiet and not speaking up.	TRA

Thinking about the imagined team again, how much would you agree or disagree with the following statements? Please answer openly and honestly. There isn't a right or wrong answer and you won't be graded on this.

[Note to STEM OST reviewers: This scale will have six options - Strongly Agree, Agree, Somewhat Agree, Somewhat Disagree, Disagree, Strongly Disagree.]

Item	Skill area
52. It is okay if one person does most of the talking during the team meetings.	TRA
53. If one person wants to keep quiet during team discussions that is okay.	TRA
54. If someone on the team is talking to the group, it is okay if team members are half listening while they work on something else.	L
55. If a teammate has something important to say about an idea someone is talking about, they should interrupt the person speaking.	L
56. If someone on the team is sharing an idea with the group, it is okay if team members are thinking about their own idea instead of fully listening.	L
57. If someone is worried that other team members will ignore what they have to say, they should probably keep their idea to themselves instead of sharing it.	L
58. If a team member has an opinion that is different from everyone else on the team, they should share it.	TRA
59. A teammate should avoid bringing up their opinion if they think it will make others upset.	TRA
60. If a teammate is confused by someone's idea, they should wait to see if anyone else on the team is confused before they ask a clarifying question.	CLC

STEM OST Expert Review questions about the overall survey

8. Do you think you would use this survey to evaluate your STEM OST program(s) that use a team structure?
   - ☐ Yes → Continue on to Question 9
   - ☐ No → Skip to Question 11
   - ☐ Maybe → Skip to Question 11

9. (If Yes to Question 8) What about this survey would make it useful for your program?
10. (If Yes to Question 8) How might you use this survey to evaluate your program? Check all that apply.
    - ☐ As a pre-survey that youth fill out when they start the program. You would use the information as a baseline for your youth and/or to help inform the teamwork activities in your program.
    - ☐ As a mid-survey after youth have been in the program for a period of time (possibly comparing to pre-survey data). The information would be used for formative evaluation purposes to find out where youth still need to improve their teamwork communication skills and to help determine how you might want to change your programming to further develop their skills.
    - ☐ As a post-survey to compare to pre-survey results. This would be used for a summative evaluation to look at your program's impact on youth's teamwork communication skills.
    - ☐ Other use (please explain):
11. (If No or Maybe to Question 8) Why might this survey not be useful to evaluate your program(s)?

12. Is there anything that would make this survey more useful for you?

13. Do you have any additional comments or suggestions about this survey as we continue to develop it?

Appendix F: Grant Advisor Expert Review Instrument

Thank you for your time and expertise to help us gather content validity evidence for the C2C Team Communication Skills survey! This document will ask you to comment on the survey's teamwork scenario and individual items. There is room at the end of the document to provide overall comments on the survey itself. Feel free to use track changes if you have specific suggested edits to the scenario or items. If you have any questions about this review or what you are asked to do, please don't hesitate to contact Amy Grack Nelson. It would be great if you can complete this review and send it back to Amy by the end of the day **Monday, February 29**. That will give us time to look over your review before the advisors meeting on March 2nd and help focus our discussion during that meeting. Thanks again for your help!

Feedback About the Teamwork Scenario

The survey's teamwork scenario was developed based on interviews with 34 STEM out-of-school time programs about teamwork in their programs. A sample of 11 of these programs then reviewed the scenario and provided suggestions to make it more in line with their program activities. The scenario is purposefully meant to be vague in order to cover the wide range of STEM topics and experiences in these programs. The scenario is meant to provide grounding for all of the survey questions and a similar frame of reference for all youth completing the survey. Review the scenario and then answer the questions below.

Teamwork scenario

Imagine you are at [Insert name of out-of-school time program] and have just been placed in a team to [solve a challenge/work on a project] together. There are two other youth on the team, one girl and one boy, so there are three of you total. All three of you are in [middle or high] school. You met your teammates earlier when you all participated in a get to know you activity where everyone in the program shared their name and five interesting facts about themselves.

Your team reviews the details of the [challenge/project] and you all make sure everyone on the team understands what you need to do. Then team members share with each other what they already know about the [challenge's/project's] topic. As a team, you decide you still need to learn more about the topic. The team members split up to find information by searching online, reading books or magazines, or looking at information provided by your program. After everyone has done some research, the team comes back together and each member shares what he or she learned. Your team then starts to share and discuss ideas about what they might need to do to complete the [challenge/project]. Your team comes to agreement about the steps needed to [solve the challenge/complete the project]. You split up to gather materials and start working together to complete the [challenge/project]. Throughout the process, you check in with each other to make sure you are all working toward the same goal.

1. Do you think the scenario provides effective grounding for the survey items? If not, what might need to be changed? (You'll probably want to come back and answer this question after you've completed the review of the items.)
2. Do you have any other comments or suggestions about the scenario?

Feedback About Individual Items, Skill Area Coverage, and Skill Area Definitions

The following pages are broken down by three skill areas related to the construct of team communication skills*:* 1) task-related assertiveness, 2) closed-loop communication, and 3) listening. For each skill area, there is a table that includes the items that were created to cover that aspect of the skill area. We'd like

you to look at each item to see if the item aligns with that particular skill area of team communication skills. Indicate if it does, doesn't, or kind of aligns. There is also space for you to comment or provide suggestions on the alignment or the wording of the individual item. We are interested in if you think the way the item is phrased is adequate to get at the particular skill area that is being measured and what changes could be made to an item to better measure that area of the skill. At the end of each table are some overall questions about that skill area.

Skill Area 1 of Team Communication Skills: Task-related assertiveness

Question stem	Items	Does this item align with skill area 1: task-related assertiveness?	Comments or questions about the item
How likely or unlikely are you to do the following things as part of this imaginary team?	a. Share with your team information you found about the topic of the [challenge/project].	☐ Yes ☐ No ☐ Kind Of	
	b. Explain an idea you have to the team.	☐ Yes ☐ No ☐ Kind Of	
	c. Keep your ideas to yourself and only share them with the team after a teammate asks you if you have an idea. (Adapted from Johnson & Johnson, 2013)	☐ Yes ☐ No ☐ Kind Of	
	d. Speak up and share an idea without first being asked by a team member.	☐ Yes ☐ No ☐ Kind Of	
	e. Share an idea that is different from a teammate's.	☐ Yes ☐ No ☐ Kind Of	
	f. Share an idea even if you think one of your teammates might dislike it.	☐ Yes ☐ No ☐ Kind Of	
	g. Tell your team what you think the team should do next even if you know others on your team will disagree with you.	☐ Yes ☐ No ☐ Kind Of	
	h. Speak up and tell the team about a skill you have that would help complete the [challenge/project].	☐ Yes ☐ No ☐ Kind Of	
	i. Tell your team what you think should be done to complete the [challenge/project].	☐ Yes ☐ No ☐ Kind Of	
	j. Speak up right away to share your idea after a teammate is done talking.	☐ Yes ☐ No ☐ Kind Of	
	k. Keep what you think about your teammates' ideas to yourself. (Adapted from Johnson & Johnson, 2013)	☐ Yes ☐ No ☐ Kind Of	
	l. Avoid a discussion with a teammate about their idea if it might lead to a disagreement.	☐ Yes ☐ No ☐ Kind Of	
	m. Participate in a discussion about a teammate's idea even if it leads to disagreements.	☐ Yes ☐ No ☐ Kind Of	
	n. Ask a teammate to share their ideas if they are quiet and not speaking up.	☐ Yes ☐ No ☐ Kind Of	
	o. Provide advice to a teammate if it appears they are having problems with what they are working on.	☐ Yes ☐ No ☐ Kind Of	
	p. Ask a teammate for help when you run into a problem.	☐ Yes ☐ No ☐ Kind Of	

How comfortable or uncomfortable would you be doing the following things as a part of this imaginary team?	q. Sharing with your team information you found about the topic of the [challenge/project].	☐ Yes ☐ No ☐ Kind Of	
	r. Explaining an idea you have to the team.	☐ Yes ☐ No ☐ Kind Of	
	s. Speaking up and sharing an idea without first being asked by a team member.	☐ Yes ☐ No ☐ Kind Of	
	t. Sharing an idea that is different from a teammate's.	☐ Yes ☐ No ☐ Kind Of	
	u. Sharing an idea even if you think one of your teammates may dislike it.	☐ Yes ☐ No ☐ Kind Of	
	v. Telling your team what you think the team should do next even if you know others on your team will disagree with you.	☐ Yes ☐ No ☐ Kind Of	
	w. Speaking up and telling the team about a skill you have that would help complete the [challenge/project].	☐ Yes ☐ No ☐ Kind Of	
	x. Telling your team what you think should be done to complete the [challenge/project].	☐ Yes ☐ No ☐ Kind Of	
	y. Speaking up right away to share your idea after a teammate is done talking.	☐ Yes ☐ No ☐ Kind Of	
	z. Participating in a discussion about a teammate's idea even if it leads to disagreements.	☐ Yes ☐ No ☐ Kind Of	
	aa. Asking a teammate to share their ideas if they are quiet and not speaking up.	☐ Yes ☐ No ☐ Kind Of	
	bb. Providing advice to a teammate if it appears they are having problems with what they are working on.	☐ Yes ☐ No ☐ Kind Of	
	cc. Asking a teammate for help when you run into a problem.	☐ Yes ☐ No ☐ Kind Of	
Thinking about the imaginary team again, how much would you agree or disagree with the following statements?	dd. It is okay if one person does most of the talking during the team meetings.	☐ Yes ☐ No ☐ Kind Of	
	ee. It is okay if one person wants to keep quiet during team discussions.	☐ Yes ☐ No ☐ Kind Of	
	ff. If a team member has an idea that is different from everyone else on the team, they should share it.	☐ Yes ☐ No ☐ Kind Of	
	gg. A teammate should avoid bringing up their opinion if they think it will make others upset.	☐ Yes ☐ No ☐ Kind Of	

1. Based on the definition of task-related assertiveness skills as described in the Test Specifications document, do these items seem to fully represent the skill area of task-related assertiveness on teams? If not, what aspect(s) of task-related assertiveness is missing from the items above?
2. Are there aspects of task-related assertiveness missing from our definition of the skill area in the Test Specifications document? If so, what is missing?
3. Do you have any additional comments or questions about items in this section or the way we are defining the skill area of task-related assertiveness?

Skill Area 2 of Team Communication Skills: Closed-loop communication

Question stem	Items	Does this item align with skill area 2: Closed-loop communication?	Comments or questions about the item
How likely or unlikely are you to do the following things as part of this imaginary team?	a. Ask your teammates questions about what they think of your idea.	☐ Yes ☐ No ☐ Kind Of	
	b. Ask your teammates if they understand your idea.	☐ Yes ☐ No ☐ Kind Of	
	c. Ask your teammates what is confusing about your idea if they have trouble understanding it.	☐ Yes ☐ No ☐ Kind Of	
	d. Repeat your idea in a different way to help explain what you said if a teammate is confused about your idea.	☐ Yes ☐ No ☐ Kind Of	
	e. Correct a teammate if it is clear that they misinterpreted something you said.	☐ Yes ☐ No ☐ Kind Of	
	f. Tell your teammate if something about their idea is confusing.	☐ Yes ☐ No ☐ Kind Of	
	g. Ask your teammate to repeat what they said because something was confusing about their idea. (Adapted from Johnson & Johnson, 2013)	☐ Yes ☐ No ☐ Kind Of	
	h. Ask your teammates questions about something that is confusing about their idea. (Adapted from Aguado et al., 2014; Center for Health Science Interprofessional Education Research and Practice, 2012)	☐ Yes ☐ No ☐ Kind Of	
	i. Ask your teammate to explain their idea in a different way so you can understand it.	☐ Yes ☐ No ☐ Kind Of	
	j. Repeat back an idea your teammate had to make sure you heard them correctly. (Adapted from Center for Health Science Interprofessional Education Research and Practice, 2012)	☐ Yes ☐ No ☐ Kind Of	
	k. Keep quiet if you have problems understanding what someone is saying instead of asking them to explain. (Adapted from Johnson & Johnson, 2013)	☐ Yes ☐ No ☐ Kind Of	
	l. Asking your teammates questions about what they think of your idea.	☐ Yes ☐ No ☐ Kind Of	
	m. Asking your teammates if they understand your idea.	☐ Yes ☐ No ☐ Kind Of	
	n. Asking your teammates what is confusing about your idea if they have trouble understanding it.	☐ Yes ☐ No ☐ Kind Of	
	o. Repeating your idea in a different way to help explain what you said if a teammate is confused about your idea.	☐ Yes ☐ No ☐ Kind Of	
	p. Correcting a teammate if it is clear that they misinterpreted something you said.	☐ Yes ☐ No ☐ Kind Of	

How comfortable or un-comfortable would you be doing the following things as a part of this imaginary team?	q. Telling your teammate that something about their idea is confusing.	☐ Yes ☐ No ☐ Kind Of	
	r. Asking your teammate to repeat what they said because something was confusing about their idea. (Adapted from Johnson & Johnson, 2013)	☐ Yes ☐ No ☐ Kind Of	
	s. Asking your teammates questions about something that is confusing about their idea. (Adapted from Aguado et al., 2014; Center for Health Science Interprofessional Education Research and Practice, 2012)	☐ Yes ☐ No ☐ Kind Of	
	t. Asking your teammate to explain their idea in a different way so you can understand it.	☐ Yes ☐ No ☐ Kind Of	
	u. Repeating back an idea your teammate had to make sure you heard them correctly. (Adapted from Center for Health Science Interprofessional Education Research and Practice, 2012)	☐ Yes ☐ No ☐ Kind Of	
Thinking about the imaginary team again, how much would you agree or disagree with the following statements?	v. If a teammate is confused by someone's idea, they should wait to see if anyone else on the team is confused before they ask a clarifying question.	☐ Yes ☐ No ☐ Kind Of	

1. Based on the definition of closed-loop communication skills as described in the Test Specifications document, do these items seem to fully represent the skill area of closed-loop communication on teams? If not, what aspect(s) of closed-loop communication is missing from the items above?
2. Are there aspects of the closed-loop communication missing from our definition of the skill area in the Test Specifications document? If so, what is missing?
3. Do you have any additional comments or questions about items in this section or the way we are defining the skill area of closed-loop communication?

Skill Area 3 of Team Communication Skills: Listening

Question stem	Items	Does this item align with skill area 3: listening?	Comments or questions about the item
How likely or unlikely are you to do the following things as part of this imaginary team?	a. Wait until someone in the group is done talking to share your idea, even if you have something important to say.	☐ Yes ☐ No ☐ Kind Of	
	b. Listen to a teammate fully explain their idea before deciding if the idea is strong or weak. (adapted from Aguado et al., 2014).	☐ Yes ☐ No ☐ Kind Of	
	c. Listen closely to someone share their idea, instead of focusing on what you are going to say to the team about your own idea.	☐ Yes ☐ No ☐ Kind Of	
	d. Interrupt a teammate if you have something important to add on to what they are saying.	☐ Yes ☐ No ☐ Kind Of	
	e. Remind a teammate to listen if they are off task and not paying attention to the person who is talking.	☐ Yes ☐ No ☐ Kind Of	
How comfortable or uncomfortable would you be doing the following things as a part of this imaginary team?	f. Waiting until someone in the group is done talking to share your idea, even if you have something important to say.	☐ Yes ☐ No ☐ Kind Of	
	g. Listening to a teammate fully explain their idea before deciding if the idea is strong or weak. (adapted from Aguado et al., 2014)	☐ Yes ☐ No ☐ Kind Of	
	h. Listening closely to someone share their idea, instead of focusing on what you are going to say to the team about your own idea.	☐ Yes ☐ No ☐ Kind Of	
	i. Reminding a teammate to listen if they are off task and not paying attention to the person who is talking.	☐ Yes ☐ No ☐ Kind Of	
Thinking about the imaginary team again, how much would you agree or disagree with the following statements?	j. If someone on the team is talking to the group, it is okay if team members are kind of listening while they work on something else.	☐ Yes ☐ No ☐ Kind Of	
	k. If a teammate has something important to say about an idea someone is talking about, they should interrupt the person speaking.	☐ Yes ☐ No ☐ Kind Of	
	l. If someone on the team is sharing an idea with the group, it is okay if team members are thinking about their own idea instead of fully listening.	☐ Yes ☐ No ☐ Kind Of	
	m. If someone is worried that other team members will ignore what they have to say, they should probably keep their idea to themselves instead of sharing it.	☐ Yes ☐ No ☐ Kind Of	

1. Based on the definition of listening skills as described in the Test Specifications document, do these items seem to fully represent the skill area of listening on teams? If not, what aspect(s) of listening is missing from the items above?
2. Are there aspects of the listening missing from our definition of the skill area in the Test Specifications document? If so, what is missing?

3. Do you have any additional comments or questions about items in this section or the way we are defining the skill area of listening?

Feedback About the Overall Construct and Survey

1. For the C2C project we have decided to define the construct of team communication skills based on the three skill areas of task-related assertiveness, closed-loop communication, and listening. We recognize this is not the only way that people may define this construct. How would you define the construct of team communication skills?

2. Is our definition missing an area of team communication skills? If yes, what would make this assessment more representative of the construct of team communication skills?

3. You each bring a certain area of expertise as an advisor (Measurement, Teamwork research, Youth development evaluation). Do you have any comments, questions, concerns, or suggestions about the instrument related specifically to your area of expertise that you haven't already mentioned? This can also include comments you might have in relation to the information included in the Test Specifications document.

Appendix G: Team Communication Skills Think-Aloud Interview Protocol

Note: This interview protocol is from the beginning of Phase 3. The scenario and items were revised over the course of the think-aloud interview process and revisions to items.

ID Number: _____ Program Name:_____ Interviewer: _____

Thank you so much for agreeing to help us out today. My name is [tell them your name] and I work at the Science Museum of Minnesota.

I'm part of a team of researchers who are creating a new survey about teamwork and are talking to you today to make sure the questions make sense for people your age. Your feedback will help us improve the questions before programs across the country use this survey. As a reminder, we'll send you a $10 VISA gift card as a thank you for all of your help today. The interview will last up to an hour. If for any reason you want to stop the interview before the hour is done, that is completely fine, just let me know. You can also feel free to skip any questions you may not feel comfortable answering. Did you have any questions about the interview?

[Verify again that we have assent for youth]. Before we start, I want to make sure it is okay to audio record what you say to make sure I don't miss anything. It will also help the interview go faster since I won't have to write down what you say. Just so you know, myself and the researchers I'm working with to create this survey will be the only people who listen to the recording. No one else will hear it. Are you OK with me recording this conversation?

[If no] No problem. I'll just jot down notes while you talk. I might ask you to slow down occasionally to make sure I'm able to write down everything you say.

[If yes] Thank you! I will start the recorder and then we can start.

1. Before we talk about the survey questions, I want to hear a little bit about your experience working in teams. What kind of team experiences have you been involved with in [insert program name]? (Probe: How have you been involved in teams in [insert program name]? How does your program use teams?) [If they haven't used teams before in their program, that is okay – you can skip this question. We also don't need a long explanation, just want to get a sense of their experience.]

2. How long have you been involved in the [insert program name]?

Next, we are going to do what is called a think-aloud interview to try out the survey questions. Instead of taking the survey quietly like you typically would, you are going to be talking about the questions and explaining why you choose certain answers. So a think-aloud interview is just like it sounds – you read the questions out loud and then think out loud as you answer the questions. We want to know what you are thinking as you answer the questions and the reasons why you chose a particular answer. This may seem a little awkward at first since we usually just answer a question and don't talk about what we are thinking as we answer it. I'll help you through it and remind you to think out loud if you forget. It usually takes people a little while to get used to it and then they catch on.

I'll give you an example of what I mean when I say think out loud. So if one of the questions was, "How many windows are there in the kitchen where you live?" I would say…. **[answer the question out loud]**

So instead of thinking quietly to myself and then answering the question on the survey, I would talk about what I'm thinking so you know how I came up with that number. This helps find out if there are any problems with the survey question. Does that make sense?

How about if you try it? [Have the youth answer about their kitchen. If they don't have windows in their kitchen, have them chose a room that has windows]

Do you have any questions about what we'll be doing before we start? Does it make sense what I mean when I say think out loud?

Now you can click next on the survey. The survey starts with a scenario before we get to the survey questions. Read everything on this page out loud and then I'll ask you a few questions.

Teamwork Skills Survey Instrument

As you complete this survey, you'll be asked to imagine yourself working with a team that is described in the teamwork scenario below. Please answer the survey questions openly and honestly about how you might respond to the team situation. You aren't graded on this and there are no right or wrong answers to the questions. The staff in your program won't see your responses.

Teamwork Scenario

Imagine you are in your program (the one you are at today) and have just been placed in a team to complete a [challenge/project] together. There are two other youth on the team, one girl and one boy, so there are three of you total. All three of you are in middle school. You met your teammates earlier when you all participated in an activity where everyone in the program shared their name and five interesting facts about themselves.

Your team reviews the details of the [challenge/project] and makes sure everyone on the team understands what they need to do. Then team members share with each other what they already know about the [challenge/project] topic. As a team, you decide that you all still need to learn more about the topic. The team members split up to find information by searching online, reading books or magazines, or looking at information provided by your program. After everyone has done some research, the team comes back together and team members explain what they learned. Your team then starts to share and discuss ideas about what they might need to do to complete the [challenge/project]. The team decides what tasks need to be done, who will work on which tasks, and then gets to work. Team members work on tasks both together and alone. Completing the [challenge/project] is dependent on everyone's contributions so team members are constantly checking in with each other to make sure the team is on track to reach their goal.

Keep this imaginary team in mind as you answer the following questions.

Interview Questions

1. Was there anything confusing or hard to understand about this scenario?

2. Can you imagine yourself in this situation? If no, what about this scenario makes it hard for you to imagine yourself in it?

Now what I'd like you to do is read each of the statements out loud and then think out loud as you answer them - just tell me everything you are thinking about as you answer the question. You won't be judged on your responses or ability to understand a question. Let me know if a question doesn't make sense to you or is hard to answer. There might be times when I stop and ask you additional questions. Okay, you can click to the next page and start. Remember to read the instructions and questions out loud too.

[Below is the text for the survey. As they read the items aloud, make sure they are explaining why they answered the way they did. You may need to remind them to think aloud or explain how they came up with their answer.]

For each of the following statements, think about the imaginary team. First, we want to know how good you feel you would be at doing the following things as a part of this imaginary team. Second, we want to know how comfortable or uncomfortable you would be doing it. Third, we want to know how likely you think you would be to actually do it. Please answer openly and honestly. There isn't a right or wrong answer and you won't be graded on this.

[Youth would be asked these three questions after each statement]

- How good or bad do you think you would be at doing this on the team? (Bad at this, Kind of bad at this, Kind of good at this, Good at this)
- How comfortable or uncomfortable would you be doing this on the team? (Uncomfortable doing this, Kind of uncomfortable doing this, Kind of comfortable doing this, Comfortable doing this)
- How likely or unlikely would you be to actually do this with the team? (Unlikely to do this, Kind of unlikely to do this, Kind of likely to do this, Likely to do this)

Item	**Notes**
1a. Repeating back an idea your teammate has to make sure you heard them correctly. (Adapted by Center for Health Science Interprofessional Education Research and Practice, 2012)	
2a. Letting your teammate know if you have trouble understanding something about their idea.	
3a. Asking your teammate to explain their idea in a different way so you can understand it better.	
4a. Asking your teammate a question about something that is confusing about their idea to make sure you understand what they are saying. (Adapted from Aguado et al., 2014; Center for Health Science Interprofessional Education Research and Practice, 2012)	
5a. Asking your teammate to repeat what they said because you are unsure if you heard them correctly. (Adapted from Johnson & Johnson, 2013)	
6a. Asking questions to better understand the information your teammate found about the topic of the [challenge/project]. (Adapted from Aguado et al.,	

2014; Center for Health Science Interprofessional Education Research and Practice, 2012)	
7a. Checking to make sure you understand what a teammate means before you agree or disagree with what they said. (Adapted from Johnson & Johnson, 2013)	
8a. Asking your teammates if they understand your idea.	
9a. Encouraging your teammates to ask you questions about your idea to make sure they understand it correctly. (Adapted from Johnson & Johnson, 2013)	
10a. Explaining your idea in more than one way if a teammate is confused about your idea.	
1b. Explaining an idea you have to the team.	
2b. Sharing an idea that is different from a teammate's.	
3b. Sharing a new idea even if you think one of your teammates might dislike it.	
4b. Bringing up a new idea that might be different from what the rest of team has been discussing.	
5b. Sharing with your team information you found about the topic of the [challenge/project].	
6b. Sharing information that none of your teammates have mentioned yet.	
7b. Speaking up and sharing information without first being asked to do so by a team member.	
8b. Sharing information with the team even if you are unsure if it will be useful to project.	
9b. Providing the team with a summary of the information everyone has provided in relation to the [challenge/project].	
10b. Updating the team with new information as you find it.	
11b. Speaking up if a teammate is dominating the conversation to make sure you can share new information you found about the [challenge/project].	

Interview questions

1. The questions asked how good you were at doing something, how comfortable you were, and how likely you were to do it. How would you explain the difference between being good at something, comfort with it, and likelihood of doing it?

2. Did you understand the differences between the four options for each of the three questions? (If not) What is confusing about these options?

3. As you read through the questions, were you able to keep the imaginary team in mind or did you start thinking about other teamwork experiences? (If they say they thought about other experiences) What other teamwork experiences were you thinking about?

For the next set of questions, continue thinking about the scenario. You can go back and read it again if you'd like. We'd like to know how easy or hard you think it would be to do the following with this imaginary team. Please answer openly and honestly. There isn't a right or wrong answer and you won't be graded on this.

How easy or hard would it be for you to do this with the imaginary team?
[Options] Hard to do this, Kind of hard to do this, Kind of easy to do this, Easy to do this

Item	Notes
1c. Stay focused on what a teammate is saying when you would rather be working on your part of the team project.	
2c. Listen closely to a teammate share their idea, instead of focusing on what you are going to say to the team about your own idea.	
3c. Fully focus on what a teammate is saying instead of thinking about what you are going to say next to the team.	
4c. Wait until a teammate is done explaining their idea before you decide if the idea is strong or weak. (Adapted from Aguado et al, 2014)	
5c. Listen to a teammate that has an opinion of what the team should do that is different than your opinion. (Adapted from Loughry, Ohland, & Moore, 2007)	
6c. Keep listening to what a teammate is saying even if you disagree with their idea.	
7c. Wait until the team is done discussing someone's idea before you share yours, even if you are really excited about sharing it.	
8c. Keep listening to a teammate if it is hard to understand what they are talking about.	
9c. Stay focused on the conversation your team is having instead of letting your mind wander.	
10c. Stop working on the team's project while a teammate is updating the group with new information.	

Interview Question

28. Did you understand what the four different options meant for the questions? (If not) What was confusing about these options?

[After they fill out the demographics and you remind them to submit the survey]

Thank you so much for helping us out today. The information from you is going to be really important in helping to improve the survey so it makes sense to youth your age. Did you have any final comments or questions for me?

I will be sending your $10 VISA pre-paid gift card to your program and they will give that to you. You should get it within 1-2 weeks. Thanks again.

Appendix H: Think-Aloud Consent Letter and Forms

August 4, 2016

Dear parent or guardian,

Your child is invited to be in a research study to create surveys that out-of-school time and afterschool programs across the country can use to assess teamwork skills in middle and high school youth. Teamwork skills are important for youth to succeed in school and future careers. We are helping programs understand how well they help youth, such as your child, develop these skills and how they can change their programs to help youth further improve these skills. Our study is funded by the National Science Foundation and is being led by researchers at the University of Minnesota and the Science Museum of Minnesota.

Your child was selected because he or she participates in an out-of-school time program. If you agree to include your child in this study, we will ask him or her to participate in a phone interview with one of our researchers. The interview will last around 30 – 45 minutes and take place over the phone at your child's youth program site. Your child will receive a $10 VISA gift card as a thank you for their time. The attached consent form provides more information about the study. If you are interested in having your child participate, please sign the form and return it to the educator at their youth program. We also ask that your child sign the youth interview assent form. Thank you in advance for your time and consideration

If you have any questions about the research, please contact Amy Grack Nelson at the Science Museum of Minnesota, [insert contact info].

Sincerely,

Dr. Frances Lawrenz
Associate Vice President for Research
University of Minnesota

Amy Grack Nelson
Evaluation & Research Manager
Science Museum of Minnesota

Youth Interview Consent Form

Your child is invited to be in a research study to create surveys that ask questions about teamwork skills. The surveys will be used to study teamwork skills in science, technology, engineering or math programs outside of school. Your child was selected as a possible participant because he or she participates in one of these programs. We ask that you read this form and ask any questions you may have before agreeing to let your child be in the study.

This study is being conducted by Dr. Frances Lawrenz in the Department of Educational Psychology at the University of Minnesota and Amy Grack Nelson in the Department of Evaluation & Research in Learning at the Science Museum of Minnesota.

Background Information
The purpose of this study is to create surveys that youth programs across the country can use to assess the development of teamwork skills in middle and high school youth.

Procedures
If you agree to let your child participate in this study, we will ask them to participate in a phone interview with a researcher. The interview will take between 45-60 minutes and will take place at your child's program site. We would like to audio record the interview with your child so that the researchers can listen to it again in case we would like to remember something your child said that might be important to the research. Your child can still participate in the interview if you would prefer they are not audio recorded.

Risks of Being in the Study
The only potential risk of participating in this research is that some of your child's interview responses may not remain private. This is very unlikely and we work to make sure this does not happen. We will store your child's interview responses and the audio recording of the interview on a computer that requires a password to login. Only the research staff will have access to this computer.

Compensation
Your child will receive a $10 VISA gift card for participating in the interview.

Confidentiality
Your child's confidentiality is important to us. The records of this study will be kept private and stored securely on a computer that requires a password to access. In any sort of report we might write, we will not include your child's name or any information that will make it possible to identify your child.

Voluntary Nature of the Study
Participation in this study is voluntary. Your decision whether or not to allow your child to participate will not affect your or your child's current or future relationship with the University of Minnesota or the Science Museum of Minnesota. If you decide to let your child participate, you are free to withdraw your child from the research at any time without affecting those relationships. You also have the choice to allow your child's interview to be audio recorded or not.

Contacts and Questions
The researchers conducting this study are Dr. Frances Lawrenz and Amy Grack Nelson. You may ask any questions you have now. If you have questions later, **you are encouraged** to contact Dr. Lawrenz at the University of Minnesota [insert contact info] or Amy Grack Nelson at the Science Museum of Minnesota, [insert contact info].

If you have any questions or concerns about this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

You will be given a copy of this information to keep for your records.

Parent or Guardian Permission
I have read the above information. I have asked questions and have received answers. I consent to allow my child to participate in the study.

Your child's full name (please print): _____

Is it okay for the researcher to audio record your child's interview? ____ Yes ___ No

Signature of parent or guardian:_____ Date: _____

Signature of researcher: _____ Date: _____

Youth Interview Assent Form for Youth Under 18 Years Old

What is our study about? You are invited to be part of a research study to create surveys that ask questions about teamwork skills. The surveys will be used to study teamwork in youth programs. You were selected as a possible participant because you are in a youth program. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

What will you be asked to do? If you agree to be in this study, we will ask you to participate in an interview about teamwork. A researcher will interview you over the phone. The interview will take between 45 - 60 minutes. We would like to audio record the interview so that the researchers can listen to it again in case we would like to remember something you said that might be important to our research.

What are the risks? The only potential risk of participating in this research is that some of your answers in the interview may not remain private. This is very unlikely and we will work to make sure this does not happen. We will store your interview responses and the audio recording of your interview on a computer that requires a password to login. Only the research staff will have access to this computer.

What are the benefits of participating? You will receive a $10 VISA gift card for participating in the interview.

Your confidentiality is important to us. We will not use your name in any reports we write about this study. We might include quotes from you, but we will not include any information that will make it possible to identify who you are.

Participation in this study is completely up to you. You can decide to participate, but chose not to have your interview audio recorded. That is completely fine and your decision. If you decide to participate, you are free to skip any interview questions. You can also decide later that you don't want to participate in the research. You can ask any questions that you have about this study. If you have a question that you didn't think of now, you can always ask later. Signing here means you have read this paper and are willing to be in this study. Being in this study is up to you, and no one will be upset with you if you don't want to participate or even if you change your mind later. If you don't want to be in this study, you don't have to sign.

Your First and Last Name (Please print): _____ Your Age: _____

Is it okay for the researcher to audio record your interview? ___ Yes ___ No

Your Signature: _____ Date: _____

Signature of Researcher: _____ Date: _____

Appendix I: Changes to Survey Items During Think-Aloud Process

Note: Some changes were made to items to remove plural pronouns (their, they) when the item was meant to be about one youth. Those changes are not noted because they are so minor.

Changes to Closed-Loop Communication Items

Item 1a

Date change made	Item wording that was tested	Changes to item	Reason for change
5/24/16 (after 16 think-alouds)	1a. Repeating back an idea your teammate has to make sure you heard them correctly. (Adapted from Center for Health Science Interprofessional Education Research and Practice, 2012)	1a. Repeating back your teammate's idea to make sure you understood it correctly.	A number of youth were misinterpreting this as repeating their own idea to their teammate instead of repeating back the teammate's idea. The item was reworded to try to make it clear it was the teammate's idea they were repeating back. Also took out "heard" like in Q5a so the item was more about clarifying understanding of the idea, not the youth's listening skills.

Item 2a

Date change made	Item wording that was tested	Changes to item	Reason for change
9/26/16 (after 18 think-alouds)	2a. Letting your teammate know if you have trouble understanding something about their idea.	2a. Letting a teammate know that you are having trouble understanding his or her idea.	Some youth interpreted this as they usually understand. Changed "if" to "that" so it is clear the statement is saying they do not understand the idea. Also changed wording "you have trouble" to "you are having trouble" and took out "something about" their idea because some youth were having problems interpreting the statement.

Item 5a

Date change made	Item wording that was tested	Changes to item	Reason for change
5/24/16 (after 15 think-alouds)	5a. Asking your teammate to repeat what they said because you are unsure if you heard them correctly. (Adapted from Johnson & Johnson, 2013)	5a. Asking your teammate to repeat his or her idea because you are unsure if you understood it correctly.	Three youth were worried people would be mad and think they were not paying attention or listening. One of these youth said when they ask someone to repeat what they said, the other youth might say they need a hearing aid and then they would feel sad. This item is related to the construct of closed-loop communication, which is about clarifying understanding. The wording was changed so the item was not focused on hearing but understanding. Also changed language to "idea" to align it with the other items.

Item 6a

Date change made	Item wording that was tested	Changes to item	Reason for change
9/26/16 (after 22 think-alouds)	6a. Asking questions to better understand the information your teammate found about the topic of the project/ challenge. (Adapted from Aguado et al., 2014; Center for Health Science Interprofessional Education Research and Practice, 2012)	6a. Asking a question to better understand the information your teammate found about the topic of the project/challenge.	Two youth talked about asking more than one question. They said they might feel differently about asking one question versus coming up with multiple questions. The idea with this item was supposed to be just about asking a question at all, not coming up with multiple questions, so changed the item to "asking a question"

Item 7a

Date change made	Item wording that was tested	Changes to item	Reason for change
5/24/16 (after 15 think-alouds)	7a. Checking to make sure you understand what a teammate means before you agree or disagree with what they said. (Adapted from Johnson & Johnson, 2013)	7a. Checking to make sure you understand what a teammate means before you agree with their idea.	The disagreement part of the item is sometimes tripping youth up because they do not want to disagree with a team member. Some youth also focus on that part of the item instead of the checking understanding part. One youth was not sure what part of the statement they were rating comfort with - comfort checking understanding or comfort disagreeing/agreeing. The construct is not so much about agree/disagree but checking understanding. For this reason, disagree was removed from the item so it just focused on agreement.
9/26/16 (after 18 think-alouds)	7a. Checking to make sure you understand what a teammate means before you agree with their idea.	7a. Checking to make sure you understand your teammate's idea before you decide if you like it.	Changed to “like” since it aligns better with how someone might view an idea – like versus agree. Agreement seemed related more to someone’s opinion.
11/20/16 (after 26 think-alouds)	7a. Checking to make sure you understand your teammate's idea before you decide if you like it.	REMOVED ITEM	Some youth were unclear what was meant by "checking" and interpreted the word differently than what was intended. A few youth thought “checking” was repeating the idea back to themselves quietly. “Checking” is supposed to be about making sure they understand their teammate’s idea. After reviewing the closed-loop communication items on the survey, other items were more specific around checking for understanding so the content is covered in other items. Plus, when looking closely at the construct, judging an idea is not part of the construct. For this reason, the item was removed.

Changes to Information Exchange Items

Item 2b

Date change made	Item wording that was tested	Changes to item	Reason for change
9/26/16 (after 21 think-alouds)	2b. Sharing an idea that is different from a teammate's.	2b. Sharing an idea about the [project/challenge] that is different from a teammate's.	This item had problems similar to Item 4b. There were a number of youth interpreting "different" as oppositional or off topic. There were even some youth that said they would answer one way if the different idea was on topic versus off topic - so they would answer differently depending on how they interpreted the item. Added [project/challenge] to stress that the idea is on topic.

Item 3b

Date change made	Item wording that was tested	Changes to item	Reason for change
5/19/16 (after 15 think-alouds)	3b. Sharing a new idea even if you think one of your teammates might dislike it.	3b. Sharing an idea even if you think one of your teammates might dislike it.	Took out "new" because it did not matter if the idea was new or not. It was the idea of sharing something even if people might dislike it that was important.
9/26/16 (after 21 think-alouds)	3b. Sharing an idea even if you think one of your teammates might dislike it.	3b. Sharing an idea even if you think your team might dislike it.	Youth are interpreting "one of your teammates" as one person or two people. A few youth said they would respond differently if it was one versus two people. Changed to "your team" to avoid multiple interpretations.

Item 4b

Date change made	Item wording that was tested	Changes to item	Reason for change
5/24/16 (after 14 think-alouds)	4b. Bringing up a new idea that might be different from what the rest of team has been discussing.	4b. Bringing up an idea that might be different from what the team has been discussing.	Took out "rest of" because one youth said it made it sound like the team was bigger than three people. Took out "new" because it did not matter if the idea was new or not.
9/26/16 (after 20 think-alouds)	4b. Bringing up an idea that might be different from what the team has been discussing.	4b. Bringing up an idea for the [project/challenge] that is different from the idea the team just finished discussing.	Some youth were interpreting this as interrupting and they did not want to do that. That is not the intention of this item, it is to bring up a new idea. There are also youth interpreting "different" as being on topic or off topic. Added "[project/challenge]" to help stress that it is not off topic. Also changed the end of the sentence to make it sound less like youth are interrupting.

Item 6b

Date change made	Item wording that was tested	Changes to item	Reason for change
4/14/16 (after 3 think-alouds)	6b. Sharing information that none of your teammates have mentioned yet.	6b. Sharing information about the project/challenge that none of your teammates have mentioned yet.	Added "about the project/challenge" because a youth asked if the information was about the project.
9/26/16 (after 21 think-alouds)	6b. Sharing information about the project/challenge that none of your teammates have mentioned yet.	6b. Sharing information you found about the topic of the [project/challenge] that none of your teammates have mentioned yet.	Two youth seemed to be interpreting it as trying to find something to share that people haven't already covered and they may not be able to find something. Changed it to "sharing information you found about..." so it is clear they already have the information.

Item 7b

Date change made	Item wording that was tested	Changes to item	Reason for change
5/24/16 (after 16 think-alouds)	7b. Speaking up and sharing information without first being asked to do so by a team member.	7b. Sharing information that is important to the [challenge/project] without having to be asked for it by a teammate.	Some youth were tripped up by “first” saying that they would not want to be the first one to speak up or that they are waiting to be called on first before talking. Removed “first” from the item to address that. Two youth interpreted the item as interrupting someone to talk. Some youth said it depended on the information. Language was added to make sure it was clear that it was focused on the challenge/project.
9/26/16 (after 21 think-alouds)	7b. Sharing information that is important to the [challenge/project] without having to be asked for it by a teammate.	7b. Sharing information you found about the topic of the [challenge/project] when you notice it might be useful to the team, instead of waiting for a teammate to ask you for the information.	The previous changes did not seem to help. Youth do not seem to be interpreting it correctly as speaking without having to be asked to. One was interpreting it as having to get teammates' attention. Reworded the item quite a bit to see if it helped.

Item 9b

Date change made	Item wording that was tested	Changes to item	Reason for change
4/14/16 (after 3 think-alouds)	9b. Providing the team with a summary of the information everyone has provided in relation to the [challenge/project].	9b. Providing the team with a verbal summary of the information everyone has provided in relation to the [challenge/project].	Added “verbal” because some people talking about a written summary.
4/26/16 (after 7 think alouds)	9b. Providing the team with a verbal summary of the information everyone has provided in relation to the [challenge/project].	9b. Providing your team with a verbal summary of the ideas each teammate has shared.	Some youth talked about public speaking or speaking to large groups so tried to make the item more team focused. Changed from “everyone” to “teammate” to clarify that they are just summarizing information about their small team and to help focus this item on their team experience.
9/26/16 (after 17 think-alouds)	9b. Providing your team with a verbal summary of the ideas each teammate has shared.	9b. Summarizing out loud the ideas your team discussed about the [project/challenge].	One youth did not understand the word "verbal summary." Three youth seem to be interpreting this as individual ideas and summarizing each individual idea, not everything together.

Item 10b

Date change made	Item wording that was tested	Changes to item	Reason for change
5/24/16 (after 16 think-alouds)	10b. Updating the team with new information as you find it.	10b. Updating the team with new information you found related to the [project/challenge].	A few youth said they do not want to share every time they find new information. A few said they may want to wait until they find a little more info and then share it instead. The construct is really about sharing new information, not that they have to do it as they find it so the wording was changed to reflect that.

Item 11b

Date change made	Item wording that was tested	Changes to item	Reason for change
5/9/16 (after 8 think-alouds)	11b. Speaking up if a teammate is dominating the conversation to make sure you can share new information you found about the [challenge/topic].	REMOVED	Kids said they did not understand what dominating means. Looking back at the construct, speaking up is not part of those skills, the sharing new information piece is. We already had that in the other items so item was removed.

Changes to Listening Items

Item 4c

Date change made	Item wording that was tested	Changes to item	Reason for change
9/26/16 (after 15 think-alouds)	4c. Wait until a teammate is done explaining their idea before you decide if the idea is strong or weak. (Adapted from Aguado et al., 2014)	REMOVED	Similar feedback here about waiting versus active listening (see 7c comments). Less about listening and more about reasoning - deciding if the idea is strong or weak.

Item 5c

Date change made	Item wording that was tested	Changes to item	Reason for change
9/26/16 (after 17 think-alouds)	5c. Listen to a teammate that has an opinion of what the team should do that is different than yours. (Adapted from Loughry, Ohland, & Moore, 2007).	5c. Listen to a teammate that has an idea of what the team should do that is different from your idea.	This says opinion but some youth are answering talking about idea. Changed to "idea" so consistent with the wording of the rest of the items.

Item 7c

Date change made	Item wording that was tested	Changes to item	Reason for change
9/26/16 (after 15 think-alouds)	7c. Wait until the team is done discussing someone's idea before you share yours, even if you are really excited about sharing it.	REMOVED	Wait until the team is done discussing – this is more about waiting and patience, not as much about listening. This item is supposed to be about ears open/mouth shut listening (as one youth educator described their program) not patience and waiting. However, some youth are not talking about listening but more about waiting until people are done. Giving people a chance to share their idea. They could be waiting and not actively listening.

Item 8c

Date change made	Item wording that was tested	Changes to item	Reason for change
5/24/16 (after 11 think-alouds)	8c. Keep listening to a teammate if it is hard to understand what they are talking about.	8c. Pay attention to what a teammate is saying, even if it is hard to understand they are talking about.	Youth are responding to this item in two different ways. Some say it is hard to keep listening because they do not understand, while others say it is hard because they want to interrupt their teammate to ask questions and clarify. Changed wording to "pay attention….even if…" to see if that helps kids avoid interpreting the item as interrupting.
9/26/16 (after 15 think-alouds)	8c. Pay attention to what a teammate is saying, even if it is hard to understand they are talking about.	8c. Pay full attention to what a teammate is saying, even if it is hard to understand what the teammate is talking about.	Added "full" because one of the youth mentioned that they would rate it differently if it said "pay full attention."
11/20/16 (after 23 think-alouds)	8c. Pay full attention to what a teammate is saying, even if it is hard to understand what the teammate is talking about.	REMOVED	Some youth are still saying it is hard or kind of hard because they would want to interrupt someone to ask for clarification, while other youth are interpreting it as not zoning out. Because of the two different interpretations the item was removed.

Appendix J: Parent/Guardian Passive Consent Letter for the Pilot and Field Tests

Dear Parent or Guardian,

My name is Amy Grack Nelson and I am Evaluation & Research Manager at the Science Museum of Minnesota. I'm conducting a study about teamwork skills with my colleague Dr. Frances Lawrenz at the University of Minnesota. We are doing research to help make sure programs that take place outside of the school day are successful in teaching teamwork skills to middle and high school youth.

Because your child participates in the [program], he/she is invited to help us test a survey that youth programs can use to assess teamwork skills. If you agree to let your child participate in this study, we will ask him/her to complete a short survey. The survey will take around [10-20/5-10] minutes to complete and will be filled out at the program site.

There are no known risks or benefits to your child for participating in this study. However, the program your child participates in will receive a [$25 or $50] VISA pre-paid card for being part of the research study.

Your child's privacy is important to us. Your child's survey will be kept private and stored either in a locked cabinet or on a computer that requires a password to access. In any report we might write, we will not include any information that will make it possible to identify your child.

If you are okay with your child being a part of this study, you are not required to do anything. However, if you would **not** like your child to participate in this study, please let [name] from [program] know by [date]. [She/He] can be reached by email [email] or phone [phone]. If you would rather, you can contact me at [insert contact info].

Participation in this study is voluntary. Your decision whether or not to allow your child to participate will not affect you or your child's relationship with the [program], University of Minnesota, or Science Museum of Minnesota. If you decide to let your child participate, you are free to withdraw your child from the research at any time without affecting those relationships. If you have questions about this study, please contact me. If you have questions or concerns and would like to talk to someone other than a researcher, you are encouraged to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

Thank you for your time and consideration.

Sincerely,
Amy Grack Nelson
Evaluation & Research Manager
Science Museum of Minnesota

Appendix K: Pilot Test Survey

Hello Youth in Programs Across the Country!

We're researchers from the Science Museum of Minnesota and University of Minnesota. We are inviting you to participate in a study to help us learn about teamwork in youth programs across the country. This survey asks various questions about your participation on teams, and will take around 15 – 20 minutes to complete.

As a thank you for participating, we will provide your program with a gift certificate and share with them information from this survey to help improve the experiences you have on teams in your program.

Your answers to the survey questions will be anonymous. This means that you don't have to give your name and no one will know how you personally answer the survey questions. After everyone has completed the survey, we will share the survey results with your program with everyone's answers combined so they won't know who said what.

This survey is optional, which means you can skip any questions you don't feel comfortable answering or are unsure how to answer. You can also stop filling out the survey at any time, and you won't get in trouble with your program.

If you have any questions about the survey, please ask an adult in your program before you start. Once you are ready to begin, go to the next page.

Thank you in advance for your time! We greatly appreciate it!

Amy Grack Nelson, Science Museum of Minnesota
Dr. Frances Lawrenz, University of Minnesota

Teamwork Survey

First read the Teamwork Scenario below. You will want to keep this imaginary team in mind as you answer the questions in the survey.

Teamwork Scenario

Imagine you are in the [Insert Program Name] and have just been placed in a team to complete a [project/challenge] together. There are two other youth on the team, one girl and one boy, so there are three of you total. All three of you are in [middle/high] school. You met your teammates for the first time today. Before working on your [project/challenge] together, you all participated in an activity where everyone in the program shared their name and five interesting facts about themselves.

Your team reviews the details of the [project/challenge] and makes sure everyone on the team understands what they need to do. Then team members share with each other what they already know about the [project/challenge] topic. As a team, you decide that you all still need to learn more about the topic. The team members split up to find information by searching online, reading books or magazines, or looking at information provided by your program. After everyone has done some research, the team comes back together, and team members explain what they learned. Your team then starts to share and discuss ideas about what they might need to do to complete the [project/challenge]. The team decides what tasks need to be done, who will work on which tasks, and then gets to work. Team members work on tasks both together and alone. Completing the final [project/challenge] is dependent on everyone's contributions so team members are constantly checking in with each other to make sure the team is on track to reach their goal.

Throughout the survey, you'll be asked to imagine yourself doing lots of different things as a member of the imaginary team described in the Teamwork Scenario on the previous page. Please answer the survey questions openly and honestly about what you might do as part of this team. There are no right or wrong answers to the questions and you aren't graded on them.

On the following pages you'll see a statement and then three questions. First, read each statement carefully. Imagine yourself doing what the statement says as a member of the imaginary team. You will then be asked how good or bad you think you might be at doing what the statement says as part of the imaginary team, how comfortable or uncomfortable you might be doing it, and how likely or unlikely it would be that you would actually do it on the imaginary team. Remember, the imaginary team is only you and two other people in your program.

For the following statements, imagine that your team is researching information about the topic of the [project/challenge].

The first thing for you to think about doing on the imaginary team is:

Sharing information you found about the topic of the [project/challenge] with your team.

1) How good or bad do you think you would be at doing this on the imaginary team?
   - ☐ Bad at this
   - ☐ Kind of bad at this
   - ☐ Kind of good at this
   - ☐ Good at this

2) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
   - ☐ Uncomfortable doing this
   - ☐ Kind of uncomfortable doing this
   - ☐ Kind of comfortable doing this
   - ☐ Comfortable doing this

3) **Share information you found about the topic of the [project/challenge] with your team.**
   How likely or unlikely would you be to actually do this with the imaginary team?
   - ☐ Unlikely to do this
   - ☐ Kind of unlikely to do this
   - ☐ Kind of likely to do this
   - ☐ Likely to do this

Sharing information you found about the topic of the [project/challenge] that none of your teammates have mentioned yet.

4) How good or bad do you think you would be at doing this on the imaginary team?
   - ☐ Bad at this
   - ☐ Kind of bad at this
   - ☐ Kind of good at this
   - ☐ Good at this

5) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
   - ☐ Uncomfortable doing this
   - ☐ Kind of uncomfortable doing this
   - ☐ Kind of comfortable doing this
   - ☐ Comfortable doing this

6) **Share information you found about the topic of the [project/challenge] that none of your teammates have mentioned yet.**
   How likely or unlikely would you be to actually do this with the imaginary team?
   - ☐ Unlikely to do this
   - ☐ Kind of unlikely to do this
   - ☐ Kind of likely to do this
   - ☐ Likely to do this

Sharing information you found about the topic of the [project/challenge] when you notice it might be useful to the team, instead of waiting for a teammate to ask you for the information.

7) How good or bad do you think you would be at doing this on the imaginary team?
   - ☐ Bad at this
   - ☐ Kind of bad at this
   - ☐ Kind of good at this
   - ☐ Good at this

8) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
   - ☐ Uncomfortable doing this
   - ☐ Kind of uncomfortable doing this
   - ☐ Kind of comfortable doing this
   - ☐ Comfortable doing this

9) **Share information you found about the topic of the [project/challenge] when you notice it might be useful to the team, instead of waiting for a teammate to ask you for the information.**
   How likely or unlikely would you be to actually do this with the imaginary team?
   - ☐ Unlikely to do this
   - ☐ Kind of unlikely to do this
   - ☐ Kind of likely to do this
   - ☐ Likely to do this

Sharing information with the team even if you are unsure if it will be useful to the [project/challenge].

10) How good or bad do you think you would be at doing this on the imaginary team?
   - ☐ Bad at this
   - ☐ Kind of bad at this
   - ☐ Kind of good at this
   - ☐ Good at this

11) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
   - ☐ Uncomfortable doing this
   - ☐ Kind of uncomfortable doing this
   - ☐ Kind of comfortable doing this
   - ☐ Comfortable doing this

12) **Share information with the team even if you are unsure if it will be useful to the [project/challenge].**
   How likely or unlikely would you be to actually do this with the imaginary team?
   - ☐ Unlikely to do this
   - ☐ Kind of unlikely to do this
   - ☐ Kind of likely to do this
   - ☐ Likely to do this

Updating the team with new information you found related to the [project/challenge].

13) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

14) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

15) **Update the team with new information you found related to the [project/challenge].**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do

Asking a question to better understand the information your teammate found about the topic of the [project/challenge].[6]

16) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

17) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

18) **Ask a question to better understand the information your teammate found about the topic of the [project/challenge].**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

[6] Adapted from Aguado et al. (2014); Center for Health Science Interprofessional Education Research and Practice (2012).

For the next set of statements, imagine you are sharing ideas about the [project/challenge] with your two teammates in the imaginary team.

The next thing for you to think about doing on the imaginary team is:

Explaining an idea you have to the team.

19) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

20) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

21) **Explain an idea you have to the team.**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

Asking your teammates if they understand your idea.

22) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

23) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

24) **Ask your teammates if they understand your idea.**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

Encouraging your teammates to ask you questions about your idea to make sure they understand it correctly.[7]

25) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

26) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

27) **Encourage your teammates to ask you questions about your idea to make sure they understand it correctly.**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

Explaining your idea in more than one way if a teammate is confused about your idea.

28) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

29) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

30) **Explain your idea in more than one way if a teammate is confused about your idea.**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

---

[7] Item adapted from D.W. Johnson & Johnson (2013).

Sharing an idea about the [project/challenge] that is different from a teammate's.

31) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

32) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

33) **Share an idea about the [project/challenge] that is different from a teammate's.**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

Sharing an idea even if you think your team might dislike it.

34) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

35) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

36) **Share an idea even if you think your team might dislike it.**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

Bringing up an idea for the [project/challenge] that is different from the idea the team just finished discussing.

37) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

38) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

39) **Bring up an idea for the [project/challenge] that is different from the idea the team just finished discussing.**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

Now you are going to imagine that <u>your teammates</u> are sharing ideas with the imaginary team about the [project/challenge]. Remember, the imaginary team is you and two other people.

The next thing for you to think about doing on the imaginary team is:

Repeating back your teammate's idea to make sure you understood it correctly.[8]

40) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

41) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

42) **Repeat back your teammate's idea to make sure you understood it correctly.**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

Letting a teammate know that you are having trouble understanding his or her idea.

43) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

44) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

45) **Let a teammate know that you are having trouble understanding his or her idea.**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

---

[8] Item adapted from Center for Health Science Interprofessional Education Research and Practice (2012).

Asking your teammate to explain his or her idea in a different way so you can understand it better.

46) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

47) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

48) **Ask your teammate to explain his or her idea in a different way so you can understand it better.**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

Asking a question if there is something confusing about your teammate's idea to make sure you understand it.[9]

49) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

50) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

51) **Ask a question if there is something confusing about your teammate's idea to make sure you understand it.**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

---

[9] Item adapted from Aguado et al. (2014); Center for Health Science Interprofessional Education Research and Practice (2012).

Asking your teammate to repeat his or her idea because you are unsure if you understood it correctly.[10]

52) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

53) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

54) **Ask your teammate to repeat his or her idea because you are unsure if you understood it correctly.**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

Summarizing out loud for your teammates the ideas the team discussed about the [project/challenge].

55) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

56) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

57) **Summarize out loud for your teammates the ideas the team discussed about the [project/challenge].**
How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

---

[10] Item adapted from D.W. Johnson & Johnson (2013).

You are almost done! This next set of statements are about a variety of things that might happen while your imaginary team is working together on the [project/challenge]. Think about how easy or hard it would be for you to do what each statement says as part of this team.

58) How easy or hard would it be for you to do each of these things with the imaginary team?

	Hard to do this	Kind of hard to do this	Kind of easy to do this	Easy to do this
Keep listening to what a teammate is saying even if you disagree with the teammate's idea.	☐	☐	☐	☐
Listen closely to a teammate share an idea instead of focusing on what you are going to say to the team about your own idea.	☐	☐	☐	☐
Listen to a teammate that has an idea of what the team should do that is different from your idea.[11]	☐	☐	☐	☐
Fully focus on what a teammate is saying instead of thinking about what you are going to say next to the team.	☐	☐	☐	☐
Stay focused on the conversation your team is having instead of letting your mind wander.	☐	☐	☐	☐
Stay focused on what a teammate is saying when you would rather be working on your part of the team [project/challenge].	☐	☐	☐	☐
Stop working on the team's [project/challenge] while a teammate is updating the group with new information.	☐	☐	☐	☐

[11] Item adapted from Loughry et al. (2007).

This is the last page! The following questions help us make sure our survey works well for a wide variety of youth.

59) How long have you been in the [INSERT PROGRAM NAME]?
- ☐ Less than a month
- ☐ 1 - 6 months
- ☐ 7 - 12 months (a year)
- ☐ Over a year

60) What grade are you in?
- ☐ 5th grade
- ☐ 6th grade
- ☐ 7th grade
- ☐ 8th grade
- ☐ 9th grade
- ☐ 10th grade
- ☐ 11th grade
- ☐ 12th grade

61) How far in school did your mother go?
- ☐ Did not finish high school
- ☐ Graduated high school
- ☐ Some education after high school
- ☐ Graduated college
- ☐ I don't know

62) What is your gender?
- ☐ Male
- ☐ Female
- ☐ Transgender
- ☐ Prefer not to answer

63) What is your race or ethnicity? (Check all that apply)
- ☐ Asian
- ☐ Black/African-American
- ☐ Hispanic/Latino
- ☐ Native American/American Indian/Alaskan Native
- ☐ White or Caucasian
- ☐ Other (please describe): _____
- ☐ Prefer not to answer

Thank you for taking this survey today! We really appreciate it!

Researchers at the Science Museum of Minnesota and University of Minnesota

Appendix L: Field Test Survey

Hello Youth in Programs Across the Country!

We're researchers from the Science Museum of Minnesota and University of Minnesota. We are inviting you to participate in a study to help us learn about teamwork in youth programs across the country. This survey asks various questions about your participation on teams, and will take around 7 - 10 minutes to complete.

As a thank you for participating, we will provide your program with a gift certificate and share with them information from this survey to help improve the experiences you have on teams in your program.

Your answers to the survey questions will be anonymous. This means that you don't have to give your name and no one will know how you personally answer the survey questions. After everyone has completed the survey, we will share the survey results with your program with everyone's answers combined so they won't know who said what.

This survey is optional, which means you can skip any questions you don't feel comfortable answering or are unsure how to answer. You can also stop filling out the survey at any time, and you won't get in trouble with your program.

If you have any questions about the survey, please ask an adult in your program before you start. Once you are ready to begin, go to the next page.

Thank you in advance for your time! We greatly appreciate it!

Amy Grack Nelson, Science Museum of Minnesota
Dr. Frances Lawrenz, University of Minnesota

Teamwork Survey

First read the Teamwork Scenario below. You will want to keep this imaginary team in mind as you answer the questions in the survey.

Teamwork Scenario

Imagine you are in the [PROGRAM] and have just been placed in a team to complete a [project/challenge] together. There are two other youth on the team, one girl and one boy, so there are three of you total. All three of you are in [Middle/high] school. You met your teammates for the first time today. Before working on your [project/challenge] together, you all participated in an activity where everyone in the program shared their name and five interesting facts about themselves.

Your team reviews the details of the [project/challenge] and makes sure everyone on the team understands what they need to do. Then team members share with each other what they already know about the [project/challenge] topic. As a team, you decide that you all still need to learn more about the topic. The team members split up to find information by searching online, reading books or magazines, or looking at information provided by your program. After everyone has done some research, the team comes back together, and team members explain what they learned. Your team then starts to share and discuss ideas about what they might need to do to complete the [project/challenge]. The team decides what tasks need to be done, who will work on which tasks, and then gets to work. Team members work on tasks both together and alone. Completing the final [project/challenge] is dependent on everyone's contributions so team members are constantly checking in with each other to make sure the team is on track to reach their goal.

Throughout the survey, you'll be asked to imagine yourself doing lots of different things as a member of the imaginary team described in the Teamwork Scenario on the previous page. Please answer the survey questions openly and honestly about what you might do as part of this team. There are no right or wrong answers to the questions and you aren't graded on them.

On the following pages you'll see a statement followed by three questions. Imagine yourself doing what the statement says as a member of the imaginary team. You will be asked how good or bad you think you might be at doing what the statement says, how comfortable or uncomfortable you might be doing it, and how likely or unlikely it would be that you would actually do it on the imaginary team. Remember, the imaginary team is only you and two other people in your program.
The first thing for you to think about doing on the imaginary team is:

Sharing information you found about the topic of the [project/challenge] **that none of your teammates have mentioned yet.**

1) How good or bad do you think you would be at doing this on the imaginary team?
   - ☐ Bad at this
   - ☐ Kind of bad at this
   - ☐ Kind of good at this
   - ☐ Good at this

2) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
   - ☐ Uncomfortable doing this
   - ☐ Kind of uncomfortable doing this
   - ☐ Kind of comfortable doing this
   - ☐ Comfortable doing this

3) How likely or unlikely would you be to actually do this with the imaginary team?
   - ☐ Unlikely to do this
   - ☐ Kind of unlikely to do this
   - ☐ Kind of likely to do this
   - ☐ Likely to do this

For the following statements, imagine you are sharing ideas about the [project/challenge] with your two teammates in the imaginary team.

The next thing for you to think about doing on the imaginary team is:

Explaining an idea you have to the team.

4) How good or bad do you think you would be at doing this on the imaginary team?
   - ☐ Bad at this
   - ☐ Kind of bad at this
   - ☐ Kind of good at this
   - ☐ Good at this

5) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
   - ☐ Uncomfortable doing this
   - ☐ Kind of uncomfortable doing this
   - ☐ Kind of comfortable doing this
   - ☐ Comfortable doing this

6) How likely or unlikely would you be to actually do this with the imaginary team?
   - ☐ Unlikely to do this
   - ☐ Kind of unlikely to do this
   - ☐ Kind of likely to do this
   - ☐ Likely to do this

Asking your teammates if they understand your idea.

7) How good or bad do you think you would be at doing this on the imaginary team?
   - ☐ Bad at this
   - ☐ Kind of bad at this
   - ☐ Kind of good at this
   - ☐ Good at this

8) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
   - ☐ Uncomfortable doing this
   - ☐ Kind of uncomfortable doing this
   - ☐ Kind of comfortable doing this
   - ☐ Comfortable doing this

9) How likely or unlikely would you be to actually do this with the imaginary team?
   - ☐ Unlikely to do this
   - ☐ Kind of unlikely to do this
   - ☐ Kind of likely to do this
   - ☐ Likely to do this

Encouraging your teammates to ask you questions about your idea to make sure they understand it correctly.[12]

10) How good or bad do you think you would be at doing this on the imaginary team?
   - ☐ Bad at this
   - ☐ Kind of bad at this
   - ☐ Kind of good at this
   - ☐ Good at this

11) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
   - ☐ Uncomfortable doing this
   - ☐ Kind of uncomfortable doing this
   - ☐ Kind of comfortable doing this
   - ☐ Comfortable doing this

12) How likely or unlikely would you be to actually do this with the imaginary team?
   - ☐ Unlikely to do this
   - ☐ Kind of unlikely to do this
   - ☐ Kind of likely to do this
   - ☐ Likely to do this

[12] Item adapted from D.W. Johnson & Johnson (2013).

Sharing an idea even if you think your team might dislike it.

13) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

14) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

15) How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

Bringing up an idea for the [project/challenge] that is different from the idea the team just finished discussing.

16) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

17) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

18) How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

Now you are going to imagine that <u>your teammates</u> are sharing ideas with the imaginary team about the [project/challenge]. Remember, the imaginary team is you and two other people.
The next thing for you to think about doing on the imaginary team is:

Asking your teammate to explain his or her idea in a different way so you can understand it better.

19) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

20) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

21) How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

Asking your teammate to repeat his or her idea because you are unsure if you understood it correctly.[13]

22) How good or bad do you think you would be at doing this on the imaginary team?
- ☐ Bad at this
- ☐ Kind of bad at this
- ☐ Kind of good at this
- ☐ Good at this

23) How comfortable or uncomfortable do you think you would be doing this on the imaginary team?
- ☐ Uncomfortable doing this
- ☐ Kind of uncomfortable doing this
- ☐ Kind of comfortable doing this
- ☐ Comfortable doing this

24) How likely or unlikely would you be to actually do this with the imaginary team?
- ☐ Unlikely to do this
- ☐ Kind of unlikely to do this
- ☐ Kind of likely to do this
- ☐ Likely to do this

[13] Item adapted from D.W. Johnson & Johnson (2013).

You are almost done! This next set of statements are about a variety of things that might happen while your imaginary team is working together on the [project/challenge]. Think about how easy or hard it would be for you to do what each statement says as part of this team.

25) How easy or hard would it be for you to do each of these things with the imaginary team?

	Hard to do this	Kind of hard to do this	Kind of easy to do this	Easy to do this
Listen closely to a teammate share an idea instead of focusing on what you are going to say to the team about your own idea.	☐	☐	☐	☐
Stay focused on the conversation your team is having instead of letting your mind wander.	☐	☐	☐	☐
Fully focus on what a teammate is saying instead of thinking about what you are going to say next to the team.	☐	☐	☐	☐
Stay focused on what a teammate is saying when you would rather be working on your part of the team [project/challenge].	☐	☐	☐	☐

These are the last questions! The following questions help us make sure our survey works well for a wide variety of youth.

26) How long have you been in the [INSERT PROGRAM NAME]?
- ☐ Less than a month
- ☐ 1 - 6 months
- ☐ 7 - 12 months (a year)
- ☐ Over a year

27) What grade are you in?
- ☐ 5th grade
- ☐ 6th grade
- ☐ 7th grade
- ☐ 8th grade
- ☐ 9th grade
- ☐ 10th grade
- ☐ 11th grade
- ☐ 12th grade
- ☐ Other (please explain): _____

28) How far in school did your mother go?
- ☐ Did not finish high school
- ☐ Graduated high school
- ☐ Some education after high school
- ☐ Graduated college
- ☐ I don't know

29) What is your gender?
- ☐ Male
- ☐ Female
- ☐ Transgender (Later changed to: A gender identity other than male or female)
- ☐ Prefer not to answer

30) What is your race or ethnicity? (Check all that apply)
- ☐ Asian
- ☐ Black/African-American
- ☐ Hispanic/Latino
- ☐ Native American/American Indian/Alaskan Native
- ☐ White or Caucasian
- ☐ Other (please describe): _____
- ☐ Prefer not to answer

Thank you for taking this survey today! We really appreciate it!

Researchers at the Science Museum of Minnesota and University of Minnesota

Appendix M: Path Diagrams for Five-Factor Model

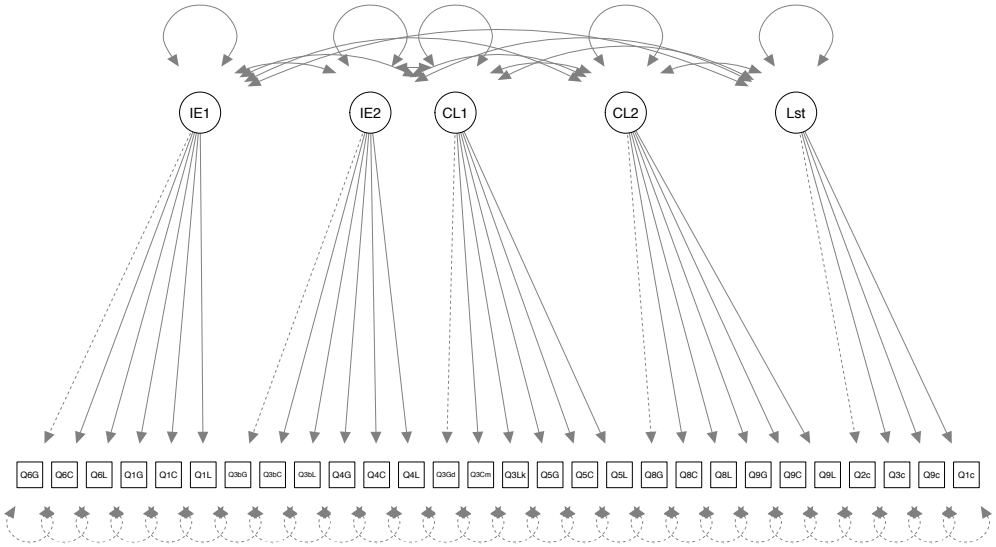

Figure M1. Five-factor model.

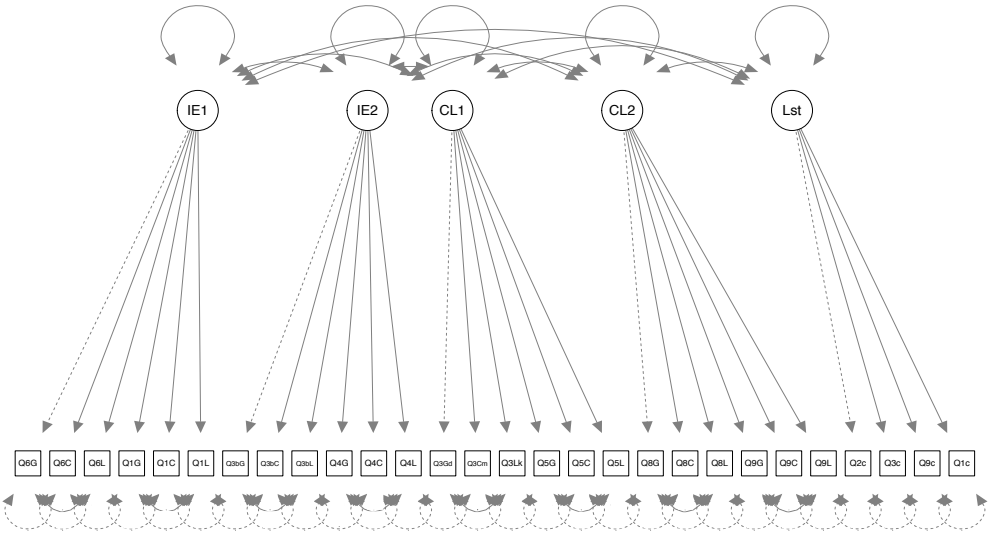

Figure M2. Five-factor model with correlated errors.

Appendix N: Modification Indices for the Five-Factor Model

			Modification index	EPC
Q9aGood	~~	Q9aLikely	144.38	0.23
Q3bGood	~~	Q4bComfort	112.70	-0.31
Q3bLikely	~~	Q4bComfort	105.48	-0.31
Q3bGood	~~	Q3bLikely	94.13	0.18
Q3bComfort	~~	Q3bLikely	93.33	0.18
Q3bLikely	~~	Q4bGood	87.55	-0.29
Q3aGood	~~	Q3aLikely	74.65	0.18
Q3bComfort	~~	Q4bLikely	72.26	-0.26
Q3bGood	~~	Q3bComfort	72.17	0.16
Q3bComfort	~~	Q4bGood	71.90	-0.25
Q4bGood	~~	Q4bComfort	71.30	0.16
Q4bGood	~~	Q4bLikely	70.02	0.16
Q8aComfort	~~	Q9aLikely	68.49	-0.28
Q3bComfort	~~	Q4bComfort	67.10	-0.22
Q3bGood	~~	Q4bLikely	62.94	-0.23
Q8aGood	~~	Q9aLikely	62.76	-0.26
Q8aComfort	~~	Q9aGood	57.25	-0.25
Q3bGood	~~	Q4bGood	54.66	-0.21
Q3bLikely	~~	Q4bLikely	46.95	-0.20
Q3aGood	~~	Q5aComfort	46.71	-0.21
Q8aGood	~~	Q9aComfort	45.51	-0.21
Q8aGood	~~	Q8aComfort	41.19	0.14
FactorIE2	=~	Q3aLikely	39.78	-0.17
Q3aComfort	~~	Q5aLikely	38.22	-0.20
Q8aGood	~~	Q8aLikely	38.02	0.14
Q8aLikely	~~	Q9aGood	37.37	-0.20
Q4bComfort	~~	Q4bLikely	37.25	0.12
Q3aComfort	~~	Q5aGood	36.44	-0.18
Q6bGood	~~	Q6bLikely	36.30	0.18
Q3aLikely	~~	Q5aComfort	35.95	-0.19
Q5aGood	~~	Q5aLikely	34.69	0.12
FactorIE1	=~	Q9aLikely	31.56	-0.27
Q2c	~~	Q3c	31.41	0.16
FactorIE1	=~	Q3aLikely	29.97	-0.22
Q3aComfort	~~	Q3aLikely	29.31	0.12
Q3aGood	~~	Q5aGood	25.16	-0.13

			Modification index	EPC
FactorIE1	=~	Q4bComfort	24.61	0.18
Q3aLikely	~~	Q5aGood	24.49	-0.15
FactorIE1	=~	Q3c	23.22	-0.20
Q6bLikely	~~	Q1bComfort	22.86	-0.17
FactorCL2	=~	Q4bComfort	22.67	0.13
Q3aGood	~~	Q5aLikely	22.55	-0.14
FactorIE1	=~	Q9c	22.55	0.19
Q9aComfort	~~	Q9aLikely	21.96	0.10
Q1bComfort	~~	Q4bComfort	21.90	0.13
Q8aGood	~~	Q9aGood	21.34	-0.14
Q5aGood	~~	Q5aComfort	20.95	0.10
Q8aLikely	~~	Q9aComfort	19.15	-0.13
Q6bGood	~~	Q1bComfort	18.82	-0.16
Q3aGood	~~	Q3aComfort	18.71	0.09
FactorIE2	=~	Q5aComfort	18.47	0.12
Q6bComfort	~~	Q1bLikely	17.80	-0.15
FactorCL1	=~	Q9aLikely	17.26	-0.14
FactorListen	=~	Q3aLikely	16.88	-0.17
FactorIE1	=~	Q8aComfort	16.77	0.20
Q6bGood	~~	Q6bComfort	16.03	0.12
FactorListen	=~	Q6bComfort	15.68	0.17
Q8aComfort	~~	Q9aComfort	15.65	-0.11
FactorCL2	=~	Q3c	15.09	-0.14
FactorCL1	=~	Q9aGood	14.51	-0.13
Q6bComfort	~~	Q1bGood	14.22	-0.14
Q6bLikely	~~	Q1bGood	14.03	-0.13
FactorIE2	=~	Q9aComfort	13.74	0.11
FactorIE1	=~	Q5aComfort	13.67	0.15
FactorCL2	=~	Q3aLikely	13.15	-0.12
Q6bComfort	~~	Q6bLikely	13.07	0.11
FactorIE2	=~	Q1bComfort	12.28	0.12
Q3bComfort	~~	Q5aComfort	12.08	0.10
Q4bComfort	~~	Q9aComfort	11.58	0.10
FactorCL2	=~	Q9c	11.33	0.12
Q4bComfort	~~	Q8aComfort	10.97	0.10
FactorCL1	=~	Q8aComfort	10.81	0.11
FactorIE1	=~	Q9aGood	10.71	-0.15
FactorListen	=~	Q4bComfort	10.60	0.11

			Modification index	EPC
Q6bLikely	~~	Q8aLikely	10.24	0.11
FactorCL1	=~	Q4bComfort	10.24	0.08
Q3aComfort	~~	Q8aComfort	10.03	0.09
FactorIE2	=~	Q6bLikely	9.98	-0.12
FactorIE1	=~	Q9aComfort	9.55	0.15
FactorIE2	=~	Q1c	9.34	0.09
Q3aComfort	~~	Q5aComfort	9.24	-0.08
Q4bGood	~~	Q3aLikely	9.09	-0.11
Q9aGood	~~	Q9aComfort	9.02	0.07
FactorIE1	=~	Q3bLikely	8.99	-0.11
Q6bGood	~~	Q3bLikely	8.96	-0.11
Q1bGood	~~	Q1bLikely	8.67	0.08
Q8aComfort	~~	Q8aLikely	8.65	0.07
Q4bComfort	~~	Q5aComfort	8.61	0.09
Q1bComfort	~~	Q1bLikely	8.44	0.07
FactorIE2	=~	Q9aLikely	8.22	-0.08
Q6bGood	~~	Q1bLikely	8.14	-0.10
Q1bGood	~~	Q1bComfort	8.00	0.07
FactorIE2	=~	Q8aComfort	7.62	0.08
FactorListen	=~	Q3bLikely	7.62	-0.09
FactorCL1	=~	Q8aGood	7.49	0.09
Q3aGood	~~	Q8aGood	7.42	0.08
Q1bComfort	~~	Q9aLikely	7.31	-0.10
Q6bComfort	~~	Q5aComfort	7.29	0.09
Q1bComfort	~~	Q3aLikely	7.28	-0.10
FactorCL1	=~	Q3c	7.24	-0.09
FactorIE2	=~	Q3c	7.22	-0.08
Q1bComfort	~~	Q3c	7.12	-0.10
FactorCL1	=~	Q6bComfort	7.06	0.10
Q6bLikely	~~	Q3bComfort	6.92	-0.10
Q6bLikely	~~	Q9c	6.86	0.10
Q6bComfort	~~	Q9c	6.78	0.09
Q5aComfort	~~	Q5aLikely	6.77	0.06
Q4bComfort	~~	Q3aLikely	6.67	-0.09
Q1bLikely	~~	Q8aLikely	6.38	0.08
Q5aGood	~~	Q8aGood	6.11	0.08
Q8aLikely	~~	Q9aLikely	6.04	-0.07
Q1bLikely	~~	Q3c	6.04	-0.10

			Modification index	EPC
Q1bGood	~~	Q3c	5.99	-0.10
Q4bLikely	~~	Q5aLikely	5.94	0.08
Q2c	~~	Q9c	5.94	-0.09
Q6bComfort	~~	Q9aGood	5.62	-0.08
FactorListen	=~	Q3aComfort	5.61	0.09
Q3aLikely	~~	Q9aComfort	5.41	-0.08
FactorListen	=~	Q5aComfort	5.33	0.09
Q1bComfort	~~	Q8aLikely	5.28	-0.08
Q6bComfort	~~	Q8aComfort	5.23	0.08
Q3bComfort	~~	Q3aLikely	5.14	-0.08
Q3bComfort	~~	Q8aLikely	5.08	-0.08
Q5aGood	~~	Q9aLikely	5.06	-0.08
Q6bLikely	~~	Q5aLikely	5.04	0.08
Q6bComfort	~~	Q1bComfort	5.04	-0.07
Q3aLikely	~~	Q8aLikely	4.81	0.07
Q2c	~~	Q1c	4.81	-0.07
Q5aLikely	~~	Q9aGood	4.76	-0.08
FactorIE1	=~	Q3bComfort	4.72	-0.08
FactorCL2	=~	Q3bLikely	4.65	-0.06
Q1bGood	~~	Q9aLikely	4.60	-0.08
Q4bComfort	~~	Q3aComfort	4.56	0.07
FactorCL2	=~	Q1c	4.51	0.07
FactorCL1	=~	Q1c	4.50	0.07
Q6bLikely	~~	Q3aComfort	4.50	-0.08
FactorCL1	=~	Q8aLikely	4.32	0.07
Q3bComfort	~~	Q8aGood	4.32	-0.07
Q5aLikely	~~	Q8aLikely	4.31	0.07
Q5aComfort	~~	Q9aLikely	4.26	-0.07
FactorIE2	=~	Q2c	4.23	-0.06
Q6bGood	~~	Q1bGood	4.23	-0.07
Q3aComfort	~~	Q9aGood	4.20	-0.07
Q1bGood	~~	Q4bComfort	4.19	0.07
Q3bGood	~~	Q9aLikely	4.01	-0.07
Q1bGood	~~	Q2c	3.98	-0.08
Factor2	=~	Q3bLikely	3.85	-0.05

Note. ~~ are the modification indices between items, =~ are the modification indices for the factor loadings between a factor and an item.

Appendix O: Error Variance-Covariance Matrix for Five-Factor Model with Correlated Errors

	3aG	3aC	3aL	5aG	5aC	5aL	8aG	8aC	8aL	9aG	9aC	9aL	6bG	6bC	6bL	1bG	1bC	1bL	3bG	3bC	3bL	4bG	4bC	4bL
3aG	.40																							
3aC	.42	.40																						
3aL	.52	.44	.54																					
5aG	0	0	0	.33																				
5aC	0	0	0	.42	.32																			
5aL	0	0	0	.44	.32	.44																		
8aG	0	0	0	0	0	0	.47																	
8aC	0	0	0	0	0	0	.45	.45																
8aL	0	0	0	0	0	0	.43	.32	.52															
9aG	0	0	0	0	0	0	0	0	0	.51														
9aC	0	0	0	0	0	0	0	0	0	.44	.42													
9aL	0	0	0	0	0	0	0	0	0	.61	.49	.56												
6bG	0	0	0	0	0	0	0	0	0	0	0	0	.58											
6bC	0	0	0	0	0	0	0	0	0	0	0	0	.32	.51										
6bL	0	0	0	0	0	0	0	0	0	0	0	0	.65	.27	.66									
1bG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.47								
1bC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.35	.38							
1bL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.33	.35	.45						
3bG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.49					
3bC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.63	.49				
3bL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.63	.63	.56			
4bG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.48		
4bC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.60	.38	
4bL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.55	.52	.52

Appendix P: Field Test Item Analysis and Frequency for Each Response Option

Factors and items	Discrimination	Difficulty (*M*)	*SD*	Response options 1	2	3	4
Closed-loop Communication 1							
3aGood	.71	3.29	0.78	3%	11%	40%	46%
3aComfort	.69	3.20	0.88	5%	17%	32%	46%
3aLikely	.67	3.25	0.86	5%	13%	35%	48%
5aGood	.72	3.38	0.78	3%	10%	33%	54%
5aComfort	.70	3.24	0.88	5%	16%	31%	49%
5aLikely	.67	3.33	0.85	5%	11%	32%	53%
Closed-loop Communication 2							
8aGood	.64	3.27	0.78	3%	13%	40%	45%
8aComfort	.61	3.25	0.82	3%	15%	36%	46%
8aLikely	.63	3.20	0.85	5%	14%	37%	44%
9aGood	.69	2.98	0.90	7%	20%	41%	32%
9aComfort	.71	3.04	0.92	7%	19%	37%	37%
9aLikely	.67	2.90	1.00	11%	22%	33%	34%
Information Exchange 1							
1bGood	.63	3.32	0.75	2%	10%	41%	46%
1bComfort	.66	3.25	0.80	3%	15%	38%	45%
1bLikely	.66	3.37	0.76	3%	9%	36%	52%
6bGood	.58	3.41	0.65	1%	6%	44%	49%
6bComfort	.59	3.32	0.74	1%	13%	39%	47%
6bLikely	.54	3.35	0.75	2%	10%	39%	50%
Information Exchange 2							
3bGood	.71	2.90	0.94	9%	22%	39%	30%
3bComfort	.71	2.69	0.97	12%	32%	32%	25%
3bLikely	.68	2.84	0.95	11%	23%	39%	28%
4bGood	.69	3.04	0.86	6%	18%	44%	33%
4bComfort	.70	2.88	0.94	9%	24%	37%	30%
4bLikely	.65	2.94	0.93	9%	21%	39%	32%
Listening							
1c	.57	3.12	0.86	5%	18%	39%	39%
2c	.61	3.29	0.77	2%	13%	39%	46%
3c	.64	3.12	0.82	4%	17%	42%	37%
9c	.57	3.25	0.84	4%	15%	35%	47%

Note. Response options for Good items: (1) Bad at this, (2) Kind of bad at this, (3) Kind of good at this, (4) Good at this. Response options for Comfort items: (1) Uncomfortable doing this, (2) Kind of uncomfortable doing this, (3) Kind of comfortable doing this, (4) Comfortable doing this. Response options for Likely items: (1) Unlikely to do this, (2) Kind of unlikely to do this, (3) Kind of likely to do this, (4) Likely to do this. Response options for Listening items: (1) Hard to do this, (2) Kind of hard to do this, (3) Kind of easy to do this, (4) Easy to do this.

Appendix Q: DIF Analysis Results

Table Q1

DIF Analysis Results Comparing Girls (Focal Group) and Boys (Reference Group)

Factors and items	χ^2	*p*	Δ	ETS category
Closed-loop Communication 1				
3aGood	1.40	.24		A
3aComfort	0.89	.34		A
3aLikely	2.21	.14		A
5aGood	0.39	.53		A
5aComfort	0.18	.68		A
5aLikely	0.17	.68		A
Closed-loop Communication 2				
8aGood	4.78	.19		A
8aComfort	0.53	.47		A
8aLikely	0.004	.95		A
9aGood	0.36	.55		A
9aComfort	1.36	.24		A
9aLikely	0.02	.90		A
Information Exchange 1				
1bGood	3.54	.06		A
1bComfort	2.52	.64		A
1bLikely	0.11	.74		A
6bGood	0.25	.62		A
6bComfort	0.001	.98		A
6bLikely	0.53	.47		A
Information Exchange 2				
3bGood	0.001	.98		A
3bComfort	2.80	.09		A
3bLikely	0.07	.79		A
4bGood	2.56	.11		A
4bComfort	0.02	.89		A
4bLikely	2.06	.15		A
Listening				
1c	0.02	.89		A
2c	0.01	.91		A
3c	0.46	.50		A
9c	4.81	.03	-1.29	B-

Table Q2

DIF Analysis Results Comparing African-American/Black Youth (Focal Group) and White Youth (Reference Group)

Factors and items	χ^2	*p*	Δ	ETS category
Closed-loop Communication 1				
3aGood	1.44	.23		A
3aComfort	0.31	.58		A
3aLikely	0.002	.94		A
5aGood	3.42	.06		A
5aComfort	0.76	.38		A
5aLikely	1.18	.28		A
Closed-loop Communication 2				
8aGood	0.46	.50		A
8aComfort	3.29	.07		A
8aLikely	1.64	.20		A
9aGood	1.47	.23		A
9aComfort	5.34	.02	-2.25	C-
9aLikely	0.96	.33		A
Information Exchange 1				
1bGood	0.29	.59		A
1bComfort	2.91	.09		A
1bLikely	0.78	.38		A
6bGood	1.08	.30		A
6bComfort	0.03	.86		A
6bLikely	1.89	.17		A
Information Exchange 2				
3bGood	2.56	.11		A
3bComfort	2.05	.15		A
3bLikely	0.64	.43		A
4bGood	0.02	.89		A
4bComfort	0.08	.79		A
4bLikely	2.09	.15		A
Listening				
1c	0.39	.53		A
2c	0.19	.66		A
3c	0.03	.85		A
9c	2.07	.15		A

Table Q3

DIF Analysis Results Comparing Asian Youth (Focal Group) and White Youth (Reference Group)

Factors and items	χ^2	*p*	Δ	ETS category
Closed-loop Communication 1				
3aGood	0.03	.87		A
3aComfort	0.65	.42		A
3aLikely	8.08	.004	2.46	C-
5aGood	7.46	.01	-2.57	C+
5aComfort	0.01	.91		A
5aLikely	0.57	.45		A
Closed-loop Communication 2				
8aGood	0.07	.80		A
8aComfort	0.13	.72		A
8aLikely	1.40	.24		A
9aGood	0.00	1.00		A
9aComfort	0.00	.99		A
9aLikely	0.01	.92		A
Information Exchange 1				
1bGood	0.17	.68		A
1bComfort	1.36	.24		A
1bLikely	0.54	.46		A
6bGood	0.02	.88		A
6bComfort	2.77	.10		A
6bLikely	3.44	.06		A
Information Exchange 2				
3bGood	0.01	.91		A
3bComfort	0.82	.37		A
3bLikely	0.01	.92		A
4bGood	3.07	.08		A
4bComfort	0.05	.83		A
4bLikely	3.16	.08		A
Listening				
1c	0.07	.79		A
2c	0.93	.34		A
3c	3.41	.07		A
9c	0.66	.42		A