Rise Together
Mentor. Inspire. Lead.®

Rise Together Program Evaluation

Applied Policy Project

Prepared by Bryan Christ for Rise Together

May 2020

Acknowledgements

I would like to thank Professor Lucy Bassett of the Frank Batten School of Leadership and Public Policy for her ongoing assistance and guidance in completing this project and managing difficult and evolving circumstances. I would also like to thank Mrs. Gina Christ, Rise Together Co-Director, for her immense support and permission to publish this report. I thank Mr. John Harris for providing me with background information on Rise Together's mentee survey and pointing me in the right direction for research supporting its subscales. Lastly, I want to thank my friends, family, and peers for supporting me and helping me proofread and improve this report.

Honor Statement

On my **honor** as a student, I have neither given nor received unauthorized aid on this assignment.

Disclaimers

The author conducted this study as part of the program of professional education at the Frank Batten School of Leadership and Public Policy, University of Virginia. This paper is submitted in partial fulfillment of the course requirements for the Master of Public Policy degree. The judgments and conclusions are solely those of the author, and are not necessarily endorsed by the Batten School, by the University of Virginia, or by any other agency.

Key Abbreviations

AVID- Advancement Via Individual Determination. Rise Together partners with Charlottesville/Albemarle high and middle schools to serve students in AVID, an elective class for first-generation and low-income college-bound students.

DID- Difference-in-difference, an approach that compares the difference in outcomes between treatment and control groups at the beginning and end of the treatment period.

ESL- English as a Second Language.

I AM- Increasing Access via Mentoring, an effective mentoring program for high school students.

MTP- MyTeachingPartner, an effective training program for secondary school teachers.

PYD- Positive Youth Development, a youth development framework that treats youth as assets rather than liabilities and guides many mentoring programs, including Rise Together.

RCT- Randomized control trial, a study that randomly assigns treatment and control groups.

RT- Rise Together, the nonprofit team mentoring program this program evaluation is designed for.

SES- Socioeconomic status.

UVA- University of Virginia.

VMP- Virginia Mentoring Partnership, a federally funded research organization that created Rise Together's mentee survey and is a subsidiary of MENTOR: The National Mentoring Partnership.

YWLP- Young Women Leaders Program, an effective mentoring program for middle school girls.

Key Terms

First-generation- Current or prospective college students whose parents have not completed four-year college.

Middle-range Student- Students who are not excelling enough academically to qualify for gifted programs but are also not low performing enough to qualify for disability or academic support.

Natural RCT- When treatment and control groups are assigned in a process that is as good as random, meaning there are no observable or unobservable differences between the groups.

Randomized ID Numbers- Random identification numbers that respondents use to fill out a survey in order to maintain anonymity.

Regression- A statistical approach that researchers use to determine if differences in average outcomes between treatment and control groups are statistically significant.

Reliability- The degree to which a survey instrument performs consistently across respondents.

Subscale- Related questions on a survey that collectively measure one construct such as self-esteem.

Validity- The degree to which a survey instrument effectively measures the intended information/constructs.

Table of Contents

Executive Summary **5**
Problem Statement **7**
Background on Rise Together **8**
Mission 8
AVID Students 8
Rise Together's Programming 9
Best Practices in Youth Mentoring **10**
Average Mentoring Program Effectiveness 10
Guiding Frameworks and Principles of Effective Mentoring Programs 10
Effective Mentoring Programming 11
Mentor Training 12
Mentor Recruitment 13
Research Design **15**
Primary Research Questions 15
Data Sources 15
Analytical Approach 17
Next Steps **21**
Get Surveys Approved Early 21
Shorten the Length of the Outcome Survey 21
Incentivize Students to Complete the Surveys 21
References **22**
Appendix A: Rise Together Survey Instruments **24**

Executive Summary

Though doubling the number of students served each year and showing initial success on staff-designed pre-post surveys, Rise Together (RT), a nonprofit team mentoring program, has never conducted a formal evaluation (Christ, 2019; G. Christ, personal communication, September 26, 2019). Because previous surveys were designed internally, they may contain biased wording or leading questions and are not based on validated measures of youth development, meaning they were not tested to ensure they accurately capture the outcomes RT intended to collect. ***Thus, without a program evaluation, the organization does not have an accurate sense of how it can improve before further scale and initial program outcomes are unlikely to be taken seriously by grant committees for larger sources of funding critical to program expansion.***

To combat this issue, RT staff worked with the Virginia Mentoring Partnership (VMP) to compile validated survey instruments for mentees (G. Christ, personal communication, September 26, 2019; Virginia Mentoring Partnership, n.d.). In this paper, I first summarize best practices and effectiveness of similar youth mentoring programs so that RT can effectively adapt their program to better serve mentees based on insights from their program evaluation. Then, I define the program evaluation RT will use in the 2020-21 school year to determine 1) whether RT has produced growth for mentees and 2) if RT's impact differs by student demographic characteristics (i.e. race, age, gender, socioeconomic status). The former explores if the program is effective, whereas the latter considers how RT could target its services to better meet the needs of a diverse range of students.

The program evaluation includes two separate analyses:

1. **Naïve comparison**: Uses regression to compare average RT mentee outcomes to those of similar students
2. **Difference-in-difference (DID) analysis**: Examines if the difference between mentee and comparison student outcomes changes from the beginning to the end of the year

The main benefit of the naïve comparison is that it allows RT to explore both overall program effectiveness using demographic controls for precision and RT's differential impact by student demographic characteristics. Additionally, because the way mentees are selected is as good as random, this analysis also serves as a natural randomized control trial (RCT), meaning any program effects could be causal in nature. However, the main downside of the naïve comparison is that RT has a relatively low sample size (80-100 students at each school), which could mean the evaluation is underpowered and fails to detect a meaningful effect (Angrist & Pischke, 2015; Bailey, 2016; G. Christ, personal communication, September 26, 2019; MENTOR, 2015).

The DID analysis' main virtue is providing another avenue to explore meaningful program effects in the event the naïve comparison is imprecise. Moreover, other mentoring evaluations have highlighted that effective mentoring programs may not increase mentee outcomes, but instead keep them from declining like their peers. While a naïve comparison may not be able to validate this change in trends because it only explores the difference in outcomes at the end of the year, a DID analysis would. However, the DID analysis' main drawback is that it would rely on limited pre-period data, which forms the backbone of the common trends assumption[1] this approach relies on.

[1] This assumption means that because mentees follow similar trends in baseline outcomes as comparison students, comparison students' post outcome trends provide a good approximation of what would have happened to mentees in the absence of RT (Angrist & Pischke, 2015; Bailey, 2016).

Despite this limitation, RT can still feel relatively confident in this approach because it is unlikely that it will face other common confounding threats to DID analyses (Angrist & Pischke, 2015; Bailey, 2016; Henneberger et al., 2013; MENTOR, 2015).

I conclude by outlining three next steps RT should take when implementing this program evaluation in 2020-21:

1. Getting surveys approved as early as possible, preferably over the summer, so they have all the necessary permissions for students to fill them out by the beginning of the school year
2. Shortening the length of the outcome survey to reduce mentee frustration
3. Providing a small incentive for students when they finish the surveys

Each of these three recommendations seek to align the evaluation with the needs and challenges of the students and schools RT works with in order to make the process as effective and frictionless as possible (Angrist & Pischke, 2015; Bailey, 2016; G. Christ, personal communication, September 26, 2019; MENTOR, 2015).

Figure 1: RT Mentees (Rise Together, 2019b).

Problem Statement

Though doubling the number of students served each year and showing initial success on staff-designed pre-post surveys (see Rise Together (2019b)), Rise Together (RT), a nonprofit team mentoring program, has never conducted a formal evaluation (Christ, 2019; G. Christ, personal communication, September 26, 2019). Because previous surveys were designed internally, they have potential for biased wording or leading questions and are not based on validated measures of youth development, meaning they were not tested to ensure they accurately and reliably capture the outcomes RT intended to collect. ***Thus, without a program evaluation, the organization does not have an accurate sense of how it can improve before further scale and initial program outcomes are unlikely to be taken seriously by grant committees for larger sources of funding critical to program expansion.***

To combat this issue, RT staff worked with the Virginia Mentoring Partnership (VMP) to compile validated survey instruments for mentees (G. Christ, personal communication, September 26, 2019; Virginia Mentoring Partnership, n.d.). In this paper, I define a program evaluation that determines 1) whether RT has produced growth for mentees on these surveys and 2) if RT's impact differs by student demographic characteristics (i.e. race, age, gender, socioeconomic status, etc.). The former explores whether the program is effective, whereas the latter considers how RT could target its services to better meet the needs of a diverse range of students.

Figure 2: RT in Action (Rise Together, 2019b).

Background on Rise Together

Mission

RT strives to lessen barriers Charlottesville and Albemarle first-generation college-bound high and middle school Advancement Via Individual Determination (AVID) students face to postsecondary attainment by equipping them with the skills needed to thrive in college. RT provides targeted, culturally responsive, and adaptive mentoring and academic tutoring that promotes academic engagement, confidence, collaboration, effective communication, goal-orientation, and civic engagement. By training University of Virginia (UVA) students as mentors and tutors, RT facilitates measurable and lasting positive change for both mentors and mentees. RT's mission is to ensure that all young adults have access to personalized support that will enable them to determine their own path in life. RT prides itself on tailoring services to simultaneously meet student, educator, mentor, and mentee needs (Rise Together, 2019a).

AVID Students

The Department of Education, state governments, and non-profit organizations channel tremendous effort and resources into improving educational outcomes in the United States. A large portion of these efforts focus on providing support systems for the lowest performing regions, schools, and students. Another portion goes to providing enhanced development opportunities for gifted students. These are important and admirable efforts, but they allow thousands of middle range students to fall through the cracks, disabling them from having the personalized attention and support they need to reach their full academic potential (Rise Together, 2019a).[2]

AVID, an elective class offered nationally for high and middle school students, is designed to serve middle range learners who are predominately first-generation (have parents who have not completed college), low-income, and nonwhite by teaching study skills, providing college guidance, and furnishing academic support (AVID, n.d.).[3] Yet, of the AVID middle schoolers RT serves, only ⅓ feel prepared for 9th grade (Rise Together, n.d.). Of the 93% of AVID high schoolers interested in attending college, only 42% feel prepared to do so (Rise Together, n.d.). ***Of the students RT serves, 90% are first-generation, 63% are from a low socioeconomic status, and 93% are people of color, resulting in limited access to college guidance and mentorship*** (G. Christ, personal communication, September 26, 2019). In turn, nationally only ¼ first-generation students attend four-year colleges and of those, ⅓ drop out after 3 years (PNPI, 2018). AVID students survive academically but face significant barriers to college attainment. RT supports AVID teachers in empowering motivated and diverse middle-range students to break through these barriers and thrive.

[2] For the purposes of this paper, middle range students refer to those who are not excelling enough academically to qualify for gifted programs but are also not low performing enough to qualify for disability or academic support.
[3] Note: RT is not AVID. AVID is an independent nonprofit that creates the AVID curriculum and trains teachers across the nation to implement it, whereas RT partners with AVID teachers in the Charlottesville/Albemarle area to help them deliver the curriculum effectively (AVID, n.d.; Rise Together, 2019b).

Rise Together's Programming

RT implements a team mentoring program, whereby teams of 10-12 trained UVA student mentors conduct mentoring sessions for roughly 250 high school and middle school AVID students. The program has two key components: mentoring sessions for AVID students and training sessions for UVA mentors (Rise Together, 2019b).

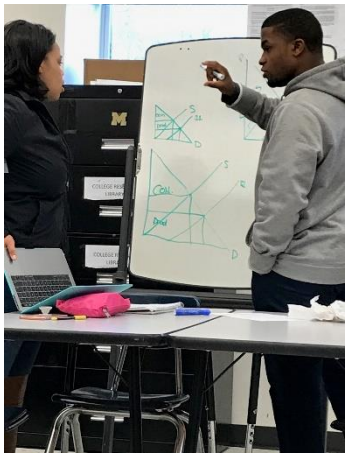
RT implements training sessions twice a month for UVA mentors with the goal of expanding their leadership and mentoring skills so that they become more effective mentors, leaders, and educators. During training sessions, mentors critically evaluate their own performance during past mentoring sessions, discuss strategies for improving group performance, collaboratively design mentoring session agendas, and participate in skill development training on mentoring competencies such as active listening and public speaking (Christ, 2019). Throughout the 2020-21 academic year, RT will provide nearly 150 UVA mentors with 45 hours of enriching community volunteering and 18 hours of personalized training/skill development (Rise Together, 2019b).

RT implements tutoring and mentoring sessions twice a month that are designed to teach AVID students the skills they need to be successful in college (Rise Together, 2019b). Throughout the 2020-21 academic year, RT will provide each of nearly 250 AVID students with 9 hours of tutoring and 18 hours of social-emotional learning (Rise Together, 2019b). Topics RT mentoring sessions address include (Rise Together, 2019b):

- Positive life choices, confidence, healthy boundaries, mental and physical wellness, examination of inherent biases
- Self-awareness, goal setting, time and stress management, project management, and resilience
- College, employment, and internship application & interview training
- Project-based civic engagement and service learning that extends the reach RT and builds soft skills such as public speaking, leadership, and communication

Figure 4: RT Mentors (Rise Together, 2019b).

Figure 3: RT Tutoring (Rise Together, 2019b).



Best Practices in Youth Mentoring

For RT to be able to effectively adapt their program to better serve mentees based on insights gained from their program evaluation, it is important that the organization is able to both contextualize it findings and build upon best practices in youth mentoring.

Average Mentoring Program Effectiveness

Building on previous studies, Raposa et al. (2019) conducted the most recent metanalysis of youth mentoring programs, relying on 70 quantitative mentoring outcome studies. The metanalysis is the most comprehensive and rigorous in the field, including all English-written outcome studies of intergenerational, one-on-one youth mentoring programs between 1975-2017 that are aligned with developmental theories of youth mentoring. With a sample size of 25, 286 youth, the metanalysis examined program outcomes in five broad categories: psychological, social, cognitive, health, or school. Compared to other effective youth-based interventions, the study revealed a moderate but statistically significant average effect size: mentees had a .21 standard deviation boost in outcomes, which was consistent across all five categories of outcomes. The effect size is equivalent to findings from previous meta-analyses, even those that contain less stringent inclusion criteria in including programs that use different models than traditional one-on-one mentoring relationships such as curriculum-based, peer, or group mentoring programs. Thus, despite its unique mentoring model, RT should interpret a similar moderate-sized effect as being on par with other effective mentoring programs (Raposa et al., 2019).

However, individual mentoring programs differ vastly in terms of target population, design, and effectiveness (Raposa et al., 2019; Tierney & Garcia, 2014). In turn, this review focuses on first generally defining best practices in youth mentoring to provide an overview of the field and then looking specifically at effective programs that are similar to RT because they will be the most informative for the organization when adapting its mentoring model.

Guiding Frameworks and Principles of Effective Mentoring Programs

At the organizational level, successful mentoring programs target disadvantaged youth (i.e. low-income or prospective first-generation college students), are flexible, focus on relationship building, involve parents and families, provide consistent oversight and support for mentors, and continuously evaluate and adapt their programming ("Successful," n.d.). Additionally, effective organizations seek to foster the Five C's of Positive Youth Development (PYD): Competence, Confidence, Connection, Character, and Caring, which all result in youth feeling like they can contribute to the program and society more broadly. Essentially, the best programs are adaptable, always place the needs of youth first, treat youth as assets rather than liabilities through empowering them to have a voice, and provide a high level of oversight, consistency, and structure (U.S. Department, 2007).

Effective Mentoring Programming

General Principles

The University of Southern California's Pullias Center for Higher Education (PCHE) designed The Increasing Access via Mentoring (I AM) program to explore best practices in youth mentoring programs targeted towards increasing postsecondary attainment because the literature is very thin and emerging in this area. I AM provides one-on-one support and mentorship to high school seniors throughout the college and financial aid application processes. I AM serves seniors who would be predominately first-generation and/or low-income college students, so insights gained from this program may be particularly relevant for RT. While only 30% of similar peers at nine participating Los Angeles schools attend a four-year institution, over 90% of I AM mentees do. From 10 years of implementing I AM and rigorously reviewing existing literature in the field, PCHE developed several mentoring principles that are relevant to RT (Tierney & Garcia, 2014):

1. Be clear- Mentoring programs should have clear objectives, identify mentors' roles, and provide mentor training that is targeted to specific program goals
2. Develop long-term relationships- Established and trusted relationships allow mentors to consistently support mentees throughout their ever-evolving needs
3. Tailor mentoring to meet students' needs- Mentoring programs should take the cultural, gender and socioeconomic backgrounds of mentees into account and develop individual mentoring plans that are aligned with each student's unique needs
4. Look to the community- Beyond providing consistent mentors, effective mentoring programs bring in outside support from business leaders, community groups, teachers and other leaders to provide valuable formal and informal mentorship

Mentoring programs should also engage mentees in community service as part of their regular programming. Community service has a host of benefits for disadvantaged youth, including creating a positive self-image, reducing risk behaviors, and increasing school performance, all of which complement the goals of mentoring programs (Wilson, 2011).

The Young Women Leaders Program (YWLP)

YWLP is a multi-site, efficacious mentoring program that is very well-studied and combines group and one-on-one mentoring. A quantitative analysis of the program found YWLP mentees had stable self-reported global self-esteem (i.e. one's overall feeling of satisfaction with self) relative to similar non-YWLP participants, who had declining self-reported global self-esteem (Henneberger et al., 2013). Additionally, ¾ YWLP participants reported the program helped them improve the way they talk with their friends, support their friends, deal with their problems, and listen to people with views different than their own ("Impact," n.d.).

A recent qualitative analysis further validated the program's effectiveness in finding mentees reported growth in four major domains as a result of YWLP: academics, relational development (developing trusting relationships), self-regulation (controlling one's actions), and self-understanding (feeling comfortable in one's own skin) (Deutsch et al., 2012). Both studies were rigorous and conducted by leading researchers in UVA's Curry School of Education and Human Development, further supporting their findings' validity. Taken together, these studies provide compelling evidence

YWLP's mentoring model is effective and based on best practices (Deutsch et al., 2012; Henneberger et al., 2013). In addition, though it's focused on serving middle school girls, insights garnered from YWLP are particularly relevant to RT because it relies on UVA student mentors and implements a similar ongoing training process and group mentoring model (Deutsch et al., 2012).

YWLP's mentoring model.

YWLP's qualitative analysis suggests that a combination of one-on-one and group mentoring may be a particularly effective model because each component has differentiated effects. YWLP found that while one-on-one relationships contribute to academic growth, group mentoring promotes relational and social development. The two elements also complement each other in that they are both attributed to increasing self-regulation and self-understanding. YWLP combines both mentoring types by conducting weekly hour-long group mentoring sessions with eight mentor pairs and requiring that mentors spend at least four hours of one-on-one time with their mentee each month. Furthermore, YWLP has found that its developmental and sequential curriculum, which addresses issues pertinent to youth and incorporates the 5 C's of PYD, provides group mentoring sessions with structure and scaffolding. Their curriculum and group mentoring sessions also give youth a voice to express their thoughts, feelings, and concerns, all of which are critical components of PYD (Deutsch et al., 2012).

Mentor Training

Pre-program Training

MENTOR: The National Mentoring Partnership is a federally funded organization that disseminates best practices for youth mentoring and distributes funding to state mentoring partnerships (i.e. VMP) to provide ongoing support and technical assistance to individual mentoring programs like RT. As such, MENTOR is seen as the industry expert on youth mentoring and its recommendations are highly credible, as they are based on rigorous mentoring research and approved by leading practitioners and researchers in the field. Though MENTOR focuses more specifically on one-on-one assigned mentoring matches, which is not a component of RT, it still highlights several best practices for training mentors before they initiate their mentoring relationship that are relevant for RT. Specifically, pre-program training should be at least two hours and highlight (MENTOR, 2015):

- Program logistics (number and length of visits, attendance policy, meeting dates and times)
- Goals and expectations of the program and mentors
- Sources of support and assistance available to mentors
- Appropriate behavior for mentors, including safety considerations, confidentiality, digital and social media use, and healthy and acceptable forms of contact
- Basic information on how youth development and demographic characteristics of mentees affect mentoring relationships

Ongoing Training

YWLP also highlights the role of ongoing mentor training in increasing the effectiveness of mentors and, therefore, driving more positive outcomes for mentees. Mentors in YWLP participate in a year-long service-learning course at UVA. Throughout the first semester of the course, mentors participate in 2 hours of training per week. The first hour each week is focused on teaching developmental issues facing adolescent girls, cultural competency (i.e. serving mentees while taking their individual backgrounds into account), and specific mentoring skills. The second hour each week is focused on supporting individual mentor-mentee matches by highlighting areas for improvement and strategies to address them. For the second semester, training is one hour a week and mostly focused on supporting mentor-mentee matches (Henneberger et al., 2013).

The literature on how to train effective mentors is relatively thin (Rhodes & Lowe, 2008; Tierney & Garcia, 2014); however, the literature on how to train effective teachers is vast and many of the same educational strategies and approaches are generalizable to mentoring. While research has shown primary teacher training programs are effective, secondary teacher training programs overall have limited or no effectiveness (Allen et al., 2011). In contrast to these traditional approaches, MyTeachingPartner (MTP) is an effective and proven method used to train both primary and secondary teachers. For example, a rigorous and well-powered randomized control trial (RCT) found that secondary teachers who participated in MTP produced learning gains for their students in the following year equivalent to moving the average student from the 50th to 59th percentile in achievement test scores (Allen et al., 2011).

Because it is geared towards secondary teachers, MTP's model is relevant to RT and provides insight into how technology can aid in the process of improving mentor effectiveness. Though MTP is focused on providing professional development and support for teachers, several of its components can be easily adapted to mentoring programs. In adapting its model, mentoring programs should (Allen et al., 2011; "MyTeachingPartner," n.d.):

1. Include ongoing training on effective methods of mentoring
2. Video record both good and bad mentor-mentee interactions during mentoring sessions
3. Conduct training sessions led by an education professional to assess filmed interactions and identify strategies to improve mentoring performance

Mentor Recruitment

College Students as Mentors

YWLP and other similar programs rely on college students as mentors because they are closer in age to mentees than adult mentors who participate in traditional community-based mentoring programs. A mentor who is closer in age to their mentee may be beneficial and more impactful because social learning theory holds that students are more likely to learn from similar others. In addition, college students typically have a prosocial academic orientation, which is important because many mentees in YWLP, RT, and other mentoring programs struggle academically. Indeed, as mentees begin to look up to mentors, they may also emulate their positive attitude towards school (Henneberger et al., 2013).

YWLP has also identified that college mentors who have strong academic self-worth and relationships with their parents have more positive and satisfactory mentoring relationships. These

mentors tend to be more confident that they can have a positive impact on mentees, so their mentoring relationships naturally end up being more fruitful. Additionally, YWLP found that mentors with high levels of autonomy typically have less effective mentoring relationships because they adopt an eager and proactive approach to mentoring, which may be met with resistance by their mentees (i.e. viewed as too pushy). Taken together, these findings suggest that mentor recruitment efforts and ongoing training should be focused on providing confident college mentors who leave space for their mentees to develop at their own pace and on their own terms (Leyton-Armakan et al., 2012).

Figure 5: RT Service-Learning Project (Rise Together, 2019b).

Figure 7: RT Mentoring Session (Rise Together, 2019b).

Figure 6: RT Mentor and Mentee (Rise Together, 2019b).

Research Design

Primary Research Questions

My primary research questions are:

1. Do RT mentees have better growth in social-emotional development and academic engagement than similar students not in the program?
2. Is there a differential impact of RT on social-emotional development and academic engagement by demographic background, specifically race, first-generation status, English as a Second Language (ESL) status, socioeconomic status (SES), and gender?

The first question asks, "Does RT achieve its stated goals?" The second asks, "Does RT achieve its goals for some types of students more than others?" Both questions are critical, but the second will allow RT to determine if it should target or adapt its programming to better serve different demographics of students. For example, if RT produces improvement for low-income black students, but has no effect for low-income white students, then they may want to consider serving more low-income black students to maximize their effect. Additionally, they could adapt their programming to better serve the low-income white students who are not well supported in their current model.

Data Sources

RT has two main data sources (See Figure 8):

1. An outcome survey that assesses student performance on the key social-emotional and college readiness competencies RT targets to examine research question 1.
2. A demographics survey that both allows RT to characterize the students it serves and explore research question 2 (i.e. determine any differential effects by demographic categories).

Demographics Survey

As shown in Figure 8, RT collects information on basic demographic characteristics of mentees once per year. These characteristics include race, age, gender, grade level, length of participation in RT, ESL status, free and reduced lunch status (a proxy for SES), and first-generation college student status.

Outcome Survey

As shown in Figure 8, RT's outcome survey includes subscales that assess students on academic engagement, confidence, grit and persistence, and communication and collaboration at the beginning and end of the year. All included subscales map onto RT's goals and curriculum for mentees. VMP designed the survey for RT based existing validated[4] and reliable[5] tools used by other

[4] i.e. the instrument has been previously tested to ensure it effectively measures the intended information/constructs (MENTOR, 2015).
[5] i.e. the tool performs accurately and consistently for respondents (MENTOR, 2015).

mentoring programs (see Appendix A for more information on each subscale). The survey generates scores from 0-100 on each subscale and averages those scores into an overall outcome score for each student (J. Harris personal communication, November 2, 2019).

Figure 8: Summary of Data Sources.

Survey Name	**Information Collected**	**Date(s) of Collection**
Demographics Survey	Demographic characteristics: • Race • Age • Gender • Grade level • Length of participation in RT • ESL status • Free and reduced lunch status • First-generation status	January-March
Outcome Survey	Student scores on the following subscales: • Communication apprehension[6] • General self-esteem • Cooperation/collaboration • Academic self-esteem • Persistence • Grit • Connectedness to future • Self-regulation • Educational attainment • Growth mindset	Beginning and end of year (September and June)

[6] Apprehension refers to a student's discomfort talking to others and communicating their ideas in small and large groups (J. Harris personal communication, November 2, 2019)

Analytical Approach

To provide the most robust analysis possible, I recommend RT conduct the two following analyses in the 2020-21 school year:

Naïve Comparison

This analysis uses a simple regression to compare mentees to AVID students who are not in RT at one of its high school and middle school sites. This analysis explores the impact of RT for high and middle school mentees on both their overall outcome score and each specific subscale shown in Figure 8. The naïve comparison for RT's impact on mentee outcome scores uses a basic regression:

$$\text{Naïve Equation 1: } Y_i = \alpha + \beta_1 + \varepsilon_i$$

In this regression, Y_i refers to mentees' average outcome score, α refers to comparison students' average outcome score, and β_1 refers to the difference in average outcome scores between mentees and comparison students, which equates to RT's impact.

Additionally, I recommend RT run two other regression analyses for mentees. The first utilizes the same basic framework as the regression above, but adds in control variables:

$$\text{Naïve Equation 2: } Y_i = \alpha + \beta_1 + X_i + \varepsilon_i$$

This regression includes all coefficients in Equation 1 and includes X_i – a vector of demographic controls—in order to improve the precision of RT's estimates by adding additional explanatory power into the model.

The second alternative regression for mentees explores the impact of RT by demographic category:

$$\text{Naïve Equation 3: } Y_i = \alpha + \beta_1 + \beta_2 (X_i \times RT) + \varepsilon_i$$

This regression follows the same format as Equation 1, but adds in β_2, an interaction term that captures the additional impact of RT on a specific demographic group of students. A positive β_2 would indicate RT impacts a specific subgroup of mentees more than the average mentee, whereas a negative β_2 would indicate that RT impacts a specific subgroup of mentees less than the average mentee.

Each of these naïve comparisons could be argued as a natural RCT because the selection process for mentees is quasi-random in that the decision on which classes RT serves depends solely on what times staff members are available to lead sessions and common times university students would be available to serve as mentors. As a result, comparison students would be in AVID at the same school and happen to be in a class RT does not serve, a process that is as good as random. A preliminary analysis from piloting this study this year validated this approach: Comparison students were statistically equivalent to mentees in pre outcome survey scores and demographic characteristics. In turn, RT could make a compelling argument that any differences they detect could be casual in nature (Angrist & Pischke, 2015; Bailey, 2016; G. Christ, personal communication, September 26, 2019; MENTOR, 2015).

Difference-in-Difference (DID) Analysis

Assumptions and Validity.

While the naïve comparison will be unbiased, it could be underpowered, as there are only 80-100 AVID students at each school site RT serves, which could make the estimates imprecise. In turn, there is a chance RT could fail to show a meaningful effect even if there is one. Thus, a DID analysis provides another avenue for exploring meaningful program effects by examining if the difference in mentee and comparison student outcomes substantively changes from the beginning to the end of the year. Additionally, YWLP and other mentoring evaluations have highlighted that effective mentoring programs may not increase mentee outcomes, but instead keep them from declining like their peers. While a naïve comparison may not be able to validate this change in trends because it only explores the difference in outcomes at the end of the year, a DID analysis would (Angrist & Pischke, 2015; Bailey, 2016; Henneberger et al., 2013; MENTOR, 2015).

This DID analysis rests on the common trends assumption, which means that because mentees follow similar trends in baseline outcomes as comparison students, comparison students' post outcome trends provide a good approximation of what would have happened to mentees in the absence of RT (see Figure 9 for a visual representation of this assumption). If conditions hold from piloting this approach this year, while comparison students and mentees would have statistically equivalent outcomes at the beginning of the year, RT will not have any earlier pre-treatment data. Without more pre-treatment data, the common trends assumption is weaker than a typical DID because RT cannot be completely sure that the similar trends observed at the beginning of the year were persistent, so it could have been a coincidental overlap in diverging trends (Angrist & Pischke, 2015; Bailey, 2016).

While this is cause for concern, RT can still feel relatively confident in this approach because it is unlikely that it will face other common confounding threats to DID analyses. One potential threat is that perhaps RT's high and middle school partners would encourage RT to work with certain AVID classes over others in the 2020-21 school year. For example, perhaps a school partner notices that one class of their 9th grade AVID students is struggling a lot more than their other freshman class, so they encourage RT to work with that class. In turn, mentees would have had declining outcomes relative to comparison students anyway, so the groups are no longer comparable, biasing RT's estimates. However, this would be unlikely because, as discussed above, the AVID classes RT works with depend on staff availability rather than school preference. While there is still a slight chance school administrators and teachers could encourage RT to work with some classes over others, RT staff have reported that they have never been overtly pushed to serve specific classes (Angrist & Pischke, 2015; Bailey, 2016; G. Christ, personal communication, September 26, 2019).

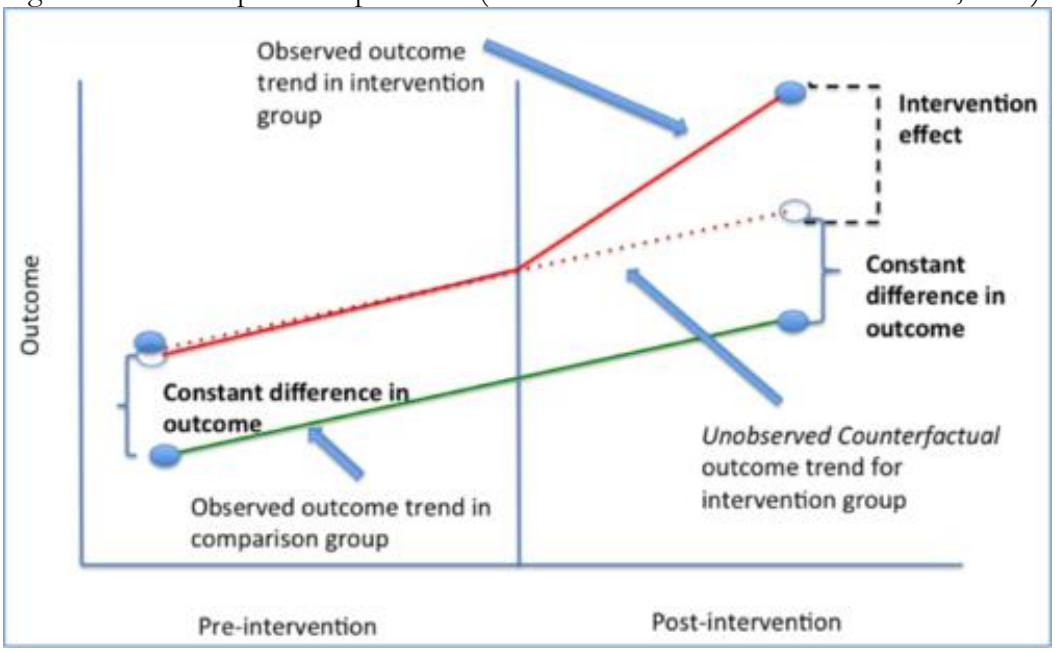
Another potential threat to RT's DID analysis is that high and middle school comparison students may decide to switch to AVID classes RT serves to take advantage of the program, fundamentally altering the composition of the underlying treatment and control groups. For example, if more motivated comparison students wanted to switch to AVID classes RT serves, the average outcomes of mentees would increase relative to comparison students regardless of RT's effectiveness. However, this also seems unlikely because RT's high and middle school partners only authorize changing student schedules in extreme circumstances (i.e. bullying or needing to drop from a more

advanced academic course to a lower-level course) (Angrist & Pischke, 2015; Bailey, 2016; G. Christ, personal communication, September 26, 2019).

A third potential threat to RT's DID analysis is that high and middle school mentees or teachers could teach comparison students the lessons they learn in RT. In this way, comparison students would benefit from the treatment, so they no longer provide a good approximation for what would have happened to mentees in the absence of RT. While possible, this threat is still unlikely because RT implements structured social-emotional learning lessons that rely on trained mentors guiding mentees, which would be difficult for mentees or teachers to reproduce on their own (Angrist & Pischke, 2015; Bailey, 2016; G. Christ, personal communication, September 26, 2019).

The last potential threat to RT's DID analysis is that high and middle school mentees might have felt empowered and worked harder than comparison students during the 2019-20 academic year because they knew they were going to get additional support from RT in the 2020-21 school year. In this way, perhaps mentees' outcomes were on an upward trend relative to comparison students anyway, so any effects RT observes are not actually due to the program. This threat is very unlikely because there is no way mentees could know they would be participating in RT prior to the start of the 2020-21 school year, as RT works with school partners over the summer to determine the AVID classes they will serve (Angrist & Pischke, 2015; Bailey, 2016; G. Christ, personal communication, September 26, 2019).

Figure 9: DID Graphical Explanation ("Difference-in-Difference Estimation," n.d.).



Regression Specification.

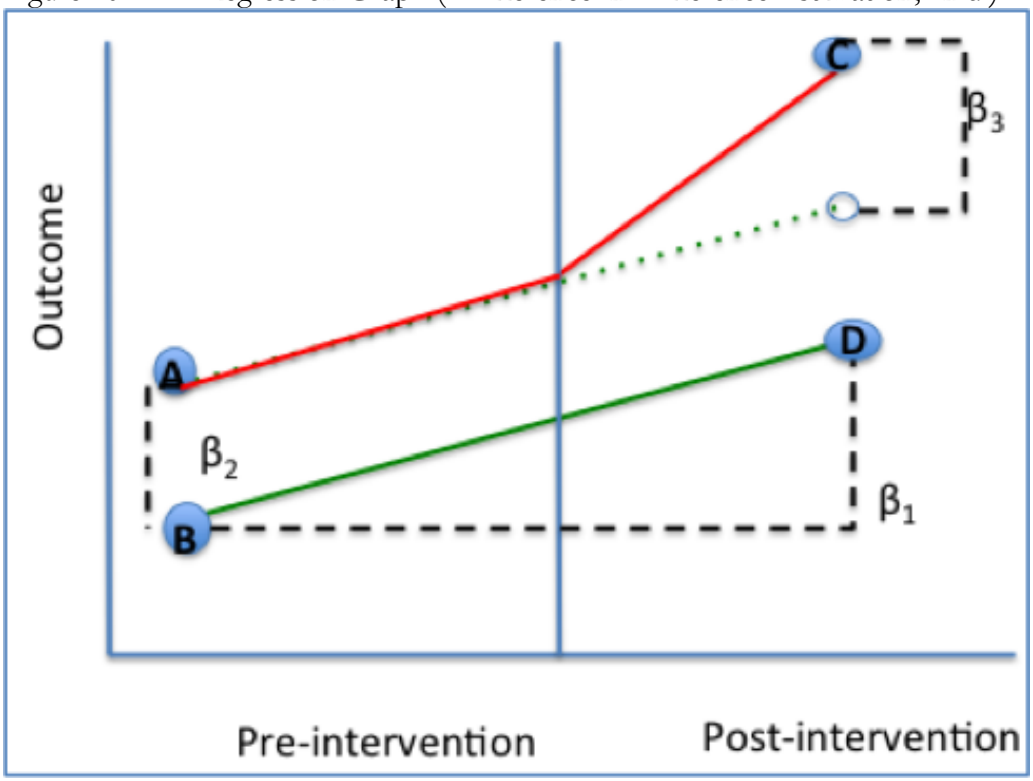
This DID analysis uses a simple regression to identify any changes in the difference in outcomes between RT mentees and comparison students at the beginning and end of the year (see Figure 10 for a graphical representation of what this approach would look like in practice):

$$Y_{it} = \alpha\,(\text{pre}) + \beta_1\,(\text{post}) + \beta_2\,(\text{RT}) + \beta_3\,(\text{RT} \times \text{post}) + \varepsilon_i$$

The components of this regression are:

- Y_{it}: mentees' average post-period outcome score
- α: comparison students' average pre-period outcome score
- β_1: effect of the post period on comparison students' average outcome score
- β_2: difference in pre-period outcome score between mentees and comparison students
- β_3: RT treatment effect (difference in predicted versus actual mentee post-period average outcome score)

Figure 10: DID Regression Graph ("Difference-in-Difference Estimation," n.d.).



Next Steps

Based on my experience helping pilot this evaluation this year as well as best practices for program evaluations, I have three main recommendations that outline how RT could best implement the full study design in the 2020-21 school year. These three recommendations also seek to align the evaluation with the needs and challenges of the students and schools RT works with in order to make the process as effective and frictionless as possible (Angrist & Pischke, 2015; Bailey, 2016; G. Christ, personal communication, September 26, 2019; MENTOR, 2015).

Get Surveys Approved Early

RT should get its outcome and demographic surveys approved as early as possible, preferably over the summer, so they have all the necessary permissions for students to fill them out. When RT piloted this analytical approach this year, they sent the surveys to their high school partner for approval at the beginning of the school year. Even though they requested that students fill the surveys out with randomized and anonymous identification (ID) numbers, because the demographic survey contained personally identifying information, it had to be sent to the district administration for approval. While it would have been very difficult or impossible to link the randomized ID numbers back to individual students, the school district wanted to be sure the process was internally vetted. The approval process was taking much longer than anticipated because district administrators had other pressing priorities at the start of the school year, which meant RT had to have high school students fill the surveys out entirely anonymously (i.e. without names or randomized ID numbers that would allow RT to link demographic and outcome surveys). Thus, if I had been able to conduct the full analysis this year, I would have been unable to explore RT's differential impact by demographic category. By starting the approval process earlier, RT can hopefully have all students fill the surveys out with randomized ID numbers to enable them to fully answer all my research questions (G. Christ, personal communication, September 26, 2019).

Shorten the Length of the Outcome Survey

When piloting the outcome survey, many students complained that it was too long (43 questions) and got frustrated, making the experience very negative for them. At best, the length of the survey added unneeded stress to the students' lives, and, at worst, it could have led them to quickly click through answers or stop filling out the survey, both of which would bias the results. While all the survey questions are important, to prioritize getting unbiased results and reducing cognitive strain, RT should shorten the survey to include only the most important subscales. To this end, I recommend RT eliminate the grit subscale because it asks very similar questions to the persistence subscale but is a lot longer, which would cut 8 questions from the survey (G. Christ, personal communication, September 26, 2019; J. Harris personal communication, November 2, 2019).

Incentivize Students to Complete the Surveys

RT should provide a small incentive (i.e. a lollipop or piece of candy, depending on what mentees prefer) if students complete the surveys. While this seems trivial, RT staff have stated that providing this incentive has made their students more willing to fill out surveys in the past, particularly the outcome survey, which, as mentioned, is fairly long and asks very personal questions (G. Christ, personal communication, September 26, 2019).

References

Advancement Via Individual Determination (AVID). (n.d.). What AVID is. Retrieved September 27, 2019, from https://www.avid.org/what-avid-is

Allen, J., Pianta, R., Gregory, A., Yee Mikami, A., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, *333*(6045). https://doi.org/10.1126/science.1207998

Angrist, J. D., & Pischke, J.-T. (2015). *Mastering `metrics: The path from cause to effect*. Princeton (N.J.): Princeton University Press.

Bailey, M. A. (2016). *Real stats: Using econometrics for political science and public policy*. New York: Oxford University Press.

Bandy, T., & Moore, K. (2010, October). *Child Trends Research-to-Results Brief: Assessing self-regulation: A guide for out-of-school time program practitioners* (Issue Brief No. 2010-23).

Christ, B. (2019). Rise Together. Retrieved September 27, 2019, from https://mgiep.unesco.org/article/rise-together

Deutsch, N., Williams, J., Henneberger, A., Reitz-Krueger, C., Futch, V., & Lawrence, E. (2012, March). Young Women Leaders: Outcomes of a Group and One-to-One Mentoring Intervention for Girls.

Difference-in-Difference Estimation. (n.d.). Retrieved April 15, 2020, from https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation

Growth Mindset Scale. (n.d.). Retrieved April 15, 2020, from http://sparqtools.org/mobility-measure/growth-mindset-scale/

Henneberger, A., Deutsch, N., Lawrence, E., & Sovik-Johnston, A. (2013). The Young Women Leaders Program: a mentoring program targeted toward adolescent girls. *School Mental Health*, *5*(3), 132-143. Retrieved from https://link.springer.com/article/10.1007/s12310-012-9093-x

Impact. (n.d.). Retrieved October 28, 2019, from https://womenscenter.virginia.edu/young-women-leaders-program/impact

Karcher, M. J. (2012). *The Hemingway Measure of Adolescent Connectedness*. Retrieved April 15, 2020, from http://adolescentconnectedness.com/media/HemingwayManual2012.pdf

Leavy, P. (2017) *Research Design Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches*. New York: The Guilford Press.

Leyton-Armakan, J., Lawrence, E., Deutsch, N., Lee Williams, J., & Lee Williams, A. (2012). Effective Youth Mentors: The Relationship between Initial Characteristics of College Women Mentors and Mentee Satisfaction and Outcome. *Journal of Community Psychology*, *40*(8). Retrieved from EBSCO database. (Accession No. 82335109)

McCroskey, J. C., Beatty, M. J., Kearney, P., & Plax, T. G. (1985). The content validity of the PRCA-24 as a measure of communication apprehension across communication contexts. *Communication Quarterly*, *33*(3), 165-173.

MENTOR. (2015). *Elements of effective practice for mentoring*.

MyTeachingPartner. (n.d.). Retrieved October 28, 2019, from https://curry.virginia.edu/myteachingpartner

Nakkula, M. J. & Harris, J. T. (2000). Self-Reported Assessment of Behavior and Social Skills.

Postsecondary National Policy Institute (PNPI). (2018, September). First-generation students. Retrieved September 27, 2019, from https://pnpi.org/first-generation-students/

Raposa, E.B., Rhodes, J., Stams, G.J.J.M. et al. (2019, March). The Effects of Youth Mentoring Programs: A Meta-analysis of Outcome Studies. *J Youth Adolescence* **48,** 423–443. Retrieved from https://doi-org.proxy01.its.virginia.edu/10.1007/s10964-019-00982-8

Research. (n.d.). Retrieved April 15, 2020, from https://angeladuckworth.com/research/
Rhodes, J., & Lowe, S. R. (2008, December). Youth Mentoring and Resilience: Implications for Practice. *Child Care in Practice, 14(1)*, 9-17. Retrieved from https://doi-org.proxy01.its.virginia.edu/10.1080/13575270701733666
Rise Together. (2019, July). Rise Together one pager. Retrieved September 27, 2019, from https://docs.google.com/document/d/1CAhpqMdYDJy0qm0nLIGdN1YLKJxkEC1QweqpMwnZrK0/edit?usp=sharing
Rise Together. (2019, July). Rise Together 2019-20 program snapshot. Retrieved September 27, 2019, from https://docs.google.com/document/d/15bjj2zAOwpvfrF67kEl-IQ7a5rYkiUm1F-jDpBk07SQ/edit?usp=sharing
Self-Esteem. (n.d.). Retrieved April 15, 2020, from https://nationalmentoringresourcecenter.org/index.php/toolkit/item/397-self-esteem.html
Successful Relationships & Programs. Retrieved March 14, 2017, from http://youth.gov/youth-topics/mentoring/best-practices-mentoring-relationships-and-programs
Tierney, W., & Garcia, L. (2014, October). *Getting in: increasing access to college via mentoring.*
U.S. Department of Education Office of Safe and Drug-Free Schools. (2007, February). *Putting Youth Development Principles to Work in Mentoring Programs* (Mentoring Fact Sheet No. 14)
Virginia Mentoring Partnership. (n.d.). Core functions. Retrieved September 27, 2019, from https://vamentoring.org/about/core-functions/
Wilson, T. (2011). *Redirect: The Surprising New Science of Psychological Change*. New York, NY: Little, Brown and Company.

Appendix A: Rise Together Survey Instruments

RT worked with John Harris, a leader in the field of mentoring evaluation and technical service provider for VMP, to develop its mentee survey instruments. The summary below discusses general validity and reliability evidence for each of the mentee survey subscales and references additional research for more information (J. Harris personal communication, November 2, 2019).

General Caveat

According to John Harris, he and RT did make some practical decisions to modify some of the subscales included in the survey and use some subscales that have not yet been tested for validity and reliability, but he does not think these decisions significantly threaten the validity of the survey instrument. In his words,

> "As a general note on the development of this survey, we ended up valuing content over validity in a number of cases. For instance, we used incomplete scales in two cases. That's a clear validity threat, though not necessarily a practical problem. I also gave you the unpublished scale from a tool Mike Nakkula and I developed, and which has no published validity evidence. Finally, I adjusted the rating options for many of the scales. That's probably not a big validity threat, but it does reduce the validity" (J. Harris personal communication, November 2, 2019).

Well-Researched and Validated Subscales

Personal Report of Communication Apprehension

The Personal Report of Communication Apprehension (PRCA) is a six-question survey and the dominant tool researchers and practitioners have used in measuring communication apprehension[7] for over three decades. The instrument has strong predictive validity (accurately predicts the respondent's communication apprehension), reliability (performs accurately across respondents), and content validity (scores are independent of the context-based content of the questions and can accurately predict communication apprehension in different contexts not mentioned in the questions) (see McCroskey et al. (1985) for more information).

Connectedness to Future

This six-question subscale, which measures the degree to which youth feel in control over their future, is part of the 78-item Hemingway Measure of Adolescent Connectedness, a comprehensive survey for scoring adolescents' sense of overall belonging and relatedness. The survey is rigorously constructed, translated into multiple languages, and has been shown to have high validity in over four quantitative studies (see Karcher (2012) for more information).

[7] Apprehension refers to a student's discomfort talking to others and communicating their ideas in small and large groups (J. Harris personal communication, November 2, 2019)

Dweck Growth Mindset

Carol Dweck's three-item Growth Mindset Scale measures how much a person believes they can become smarter through hard work. The survey is one of the most widely used and well-studied instruments in the field of adolescent development and is primarily shown to be valid and reliable for college and high school students (see "Growth Mindset Scale" (n.d.) for more information).

Questionnaire on Self-Regulation

RT's mentee survey includes the three-item cognitive self-regulation subscale, which measures the degree to which students can plan ahead to reach goals and is a component of the 13-item Questionnaire on Self-Regulation. Both the subscale and full questionnaire are well-researched and recommended by Child Trends, a leading nonprofit research center that provides research, data, and analysis to help practitioners better serve youth (see Bandy & Moore (2010) for more information).

Less Researched or Validated Subscales

Duckworth Grit Scale

Angela Duckworth designed this eight-item scale to measure grit, or the ability to stay interested in and put effort towards long-term goals. The Grit Scale has been shown to have high reliability and validity in assessing differences in individual students' grit but has not been specifically tested to show within-student changes in grit over time. This means it is uncertain whether it is a valid indicator for pre-post change as a result of RT, which is another reason why I recommended removing it from RT's mentee survey (see "Research" (n.d.) for more information).

Self-Esteem Questionnaire

RT's mentee survey includes five items from the Global Self-Worth Scale, a subscale of the Self-Esteem Questionnaire, which is intended to measure a youth's global, or overall, self-esteem (see "Self-Esteem" (n.d.) for more information). The subscale itself has strong evidence of reliability and validity, but John Harris and RT cut it down from eight items to five. While this is a concern, John Harris reported that a large study he conducted showed that the shorter scale is both valid on its own and predictive of a large portion of the variation in scores generated by the full eight-item scale (J. Harris personal communication, November 2, 2019).

RT's mentee survey also includes four items from the Academic Self-Esteem subscale of the Self-Esteem Questionnaire. While this subscale is also well-validated, John Harris and RT cut it down from its full length, weakening its validity (see "Self-Esteem" (n.d.) for more information). Unlike the Global Self-Worth Scale, John Harris has not conducted a previous study verifying the validity of this shortened version of the Academic Self-Esteem subscale (J. Harris personal communication, November 2, 2019).

Expectations about Educational Attainment

Unfortunately, there is no real validation evidence for this scale, which measures the degree to which students believe they can attain their postsecondary aspirations. However, John Harris selected

prompts that have been used widely, meaning other researchers believe this scale performs consistently (J. Harris personal communication, November 2, 2019).

Self-Assessment of Behavior & Social Skills

This survey was designed by John Harris and Michael Nakkula to measure the degree to which students feel comfortable collaborating with their peers (see Nakkula & Harris (2000) for more information). The survey has no published validity evidence, but Harris and Nakkula use it often in their mentoring evaluations and report that it performs consistently (J. Harris personal communication, November 2, 2019).