Promoting Student Outcomes by Increasing Afterschool Site Quality

Annie Wright, Dylan Famer, Patrick Lan
Southern Methodist University

Rachel Johns
Dallas Afterschool

Author Note
Annie Wright is the Director of Evaluation at the Center on Research and Evaluation (CORE), Annette Caldwell Simmons School of Education, Southern Methodist University.

Correspondence concerning this article should be directed to Annie Wright, Center on Research and Evaluation, Southern Methodist University, P.O. Box 75051, Dallas, TX 75275.

Contact: anniew@smu.edu

Key words: afterschool, quality, school engagement, academic efficacy, achievement

Abstract

This study examines what aspects of afterschool result in positive outcomes for students, and how these effective elements can best be promoted in diverse afterschool sites. Southern Methodist University's Center on Research and Evaluation (CORE) is conducting an ongoing evaluation of Dallas Afterschool's Program Quality Improvement initiative, where targeted training and coaching (Program Quality Improvement activities) are provided to a network of 100+ afterschool sites with the aim of increasing overall afterschool quality and positive outcomes for youth. We first address how afterschool quality is being operationalized and measured, including comparing DAS's quality assessment tool to core concepts of quality being utilized nationwide, as well as results of factor analyses. We then examine whether the amount of training and/or coaching provided to sites by DAS staff could explain overall differences in quality across sites. Last, we look at whether site-level quality can predict academic outcomes for students, as measured by self-reported school engagement and efficacy as well as by extant school records. Analyses indicated that provision of PQI positively impacted one element of afterschool quality, Activities and Programming, and that higher overall quality was related to positive gains for literacy in lower elementary aged students. These findings partially support the theory of change being tested here, that training and technical assistance can increase site quality and that site quality in turn, drives positive student outcomes. However, the findings are not robust overall and additional analyses with a stronger research design are warranted. Nevertheless these analyses represent how an ongoing, real-world evaluation partnership is leveraging annual data collection to continuously improve services as well as to demonstrate long term impacts over time.

Promoting Student Outcomes by Increasing Afterschool Site Quality

Introduction

Six and a half million American school children in kinder through 12th grade are estimated to participate in some kind of after or summer school programming, largely referred to as Out of School Time (OST) programs (Harvard Family Research Project, 2008). OST and closely related extended learning models are gaining traction as a model for addressing stubborn academic disparities by exposing underprivileged students to enrichment activities that seem to be part and parcel of the middle class experience (cite). Indeed, one estimate shows that by the time American students reach the 6th grade, middle class youth have likely spent 6,000 more hours *learning* in a variety of in and out of school settings than youth born into poverty. The OST field has emerged as an ameliorative counter to this disparity, aiming to “close the gap” in pure number of hours of exposure to learning environments. Beyond being a childcare provider, OST programs are experiencing a professionalization of sorts and aim to augment and extend learning opportunities through strategic programming guided by best practices. Further, as OST programs seek to grow in number, to provide after-school access to more students, and provide more effective programs overall, questions are emerging about how best to ensure quality implementation and scaling of best practices.

This study addresses site-level programmatic impacts on student outcomes as well as a training and coaching model for ensuring quality programming at scale and addresses recent calls in the literature to pursue exactly these two, parallel strands of inquiry (Granger, 2010). This study leverages an ongoing evaluation partnership between Dallas Afterschool, who supports quality afterschool practices across a network of 133 after and summer school sites, and the Center on Research and Evaluation (CORE) at Southern Methodist University. Through this partnership we are able to examine how afterschool quality is being monitored and promoted across sites. Second, we are able to look at the impacts of quality afterschool on student outcomes.

What is quality afterschool?

Operationalizing quality afterschool has been a central focus of OST research and practice (Tracy, Charmaraman, Ceder, Richer & Surr, 2016; Vandell, 2013). With some meaningful exceptions (see, for example, The Forum on Youth Investment impact findings from the Youth Program Quality Intervention study), the bulk of the work done on quality has been largely descriptive (HFRP, 2008). It is now widely accepted that attendance at after-school sites, i.e. dosage and frequency alone, cannot sufficiently explain outcomes (Hirsch, Mekinds & Stawicki, 2010; Lauer et al 2006; Roth, Malone and Brooks-Gunn, 2010). Instead, we must take into account various aspects of quality, a multi-dimensional construct that occurs at multiple levels of an after-school organization.

A synthesis of the literature conducted by the Harvard Family Research Project (2008) summarized three critical, and interrelated factors for achieving successful outcomes in after-school that point to one conceptualization of overall quality; a) access to and sustained participation in a program, b) quality programming and staffing, and c) promoting strong partnerships among the after-school program and other places where students are learning, such as their schools, families, and other community institutions.

A further articulation of commonalities in conceptualizing and measuring afterschool quality comes from a review by Yohalem & Wilson-Ahlstrom (2010) of nine commonly used assessments of after-school quality. They first identify six core concepts that were present in all tools they reviewed and four additional concepts that were present in only some (see Table One). Additional conceptualizations of a multi-dimensional conceptualization of quality occurring at multiple levels of the youth participation, program, and overall organization are provided below.

Youth perception.

One aspect of quality is the subjective experience of youth participants themselves. Indeed, Vandell (2013) summarizes the quality of students' experiences in after-school as occurring in three key areas: positive interactions with program staff, with program peers, and an overall interest and engagement in program activities. This sentiment fits well with the notion that participation in afterschool is something quite different than attendance alone, and understanding youth engagement in activities, qualities of relationships with others in the site, etc., must be taken into account (Hirsch et al 2010).

Program features.

At the programmatic level, common domains of quality include supportive social environments, such as having a welcoming and inclusive environment; program organization and structure, such as having flexible but well organized activities; and opportunities for engagement in learning, including youths' active involvement (Tracy, Charmaraman, Ceder, Richer & Surr, 2016). The Harvard synthesis reflects these same domains by describing "quality programming and staffing" as including adequate supervision, a well-prepared staff, and intentional programming.

Organizational characteristics.

At the organizational level, certain practices such as large-scale (city wide) advocacy, and partnering with families and schools (Hirsch, Mekinds & Stawicki, 2010), using ongoing evaluation to inform best practices, and even setting acceptably high salaries to recruit qualified staff are all considered part of a quality afterschool organization (HFRP, 2008). Locally, efforts by DAS include partnering with other OST providers to ensure a full network of access to programming, including identifying and addressing "program deserts."

What impacts does quality afterschool have for students?

Quality OST settings, particularly quality after-school programs, have been shown to positively impact the academic, social, and emotional development of participating youth (HFRP, 2008; National Institute on Out-of-School Time, 2009) and are associated with increased academic achievement, increased school engagement, and decreased involvement in risky behaviors (Greene, Lee, Constance, & Hynes, 2013; Moore, Murphey, Bandy, & Cooper, 2014.) The sites included in this study provide a hybrid mix of mentoring and general supports, some self-described as Social & Emotional Learning (SEL), alongside targeted homework help and academic interventions, and more general supports related to youth development. This mixed approach, of providing both social and academic activities, has demonstrated positive academic effects (Lauer, Akiba, Wilkerson, Apthorp, Snow & Martin Glenn, 2006; HFRP, 2008; Vandell, 2013). As such, we review the literature on multiple different kinds of programs and evidence of their impacts on proximal and distal academic outcomes. While the various types of programs are presented independently below, many programs have a blend of the values and approaches described and rarely does a program fit into a single category.

Academic supports.

Numerous studies have documented the association between OST programming—particularly high quality programing—and positive impacts on student academic outcomes including self-reported attitudes and behaviors as well as achievement metrics such as test scores and grades (Durlak et al, 2011; Eccles & Gootman, 2002; Lauer et al, 2006; Mahoney Vandell, Simpkins & Zarrett, 2009; Vandell, 2013). Lauer, Akiba, Wilkerson, Apthorp, Snow & Martin Glenn's (2006) meta-analysis concludes that OST programs can indeed have positive effects on the achievement of at-risk students in particular, that targeted small group tutoring has particularly clear impacts on reading outcomes. Importantly,their finding that relatively small to moderate effect sizes serves as a reminder that while positive impacts are plausible results of OST programming, they are likely not sufficient for closing the achievement gap in and of themselves. That is, other mullti-level and muti-setting factors are also at play when predcting academic outcomes and even exceptionally strong OST programming must work in concert with these other influences in order to truly close persistent gaps.

Social and emotional learning.

An emerging area of focus for OST programs is in social and emotional learning. Many providers recognize the opportunity to provide these "non-cognitive" supports that are not typically provided during school hours (though this is beginning to change as more schools and districts adopt large scale SEL curricula). A meta-analysis by Durlak, Weissberg & Pachan (2010) synthesized the available literature on SEL programs in OST settings and concluded that positive impacts for youth were achievable. These impacts included positive self-perception and social behaviors, and some proximal academic domains such as school bonding, as well as secondary, or more distal, impacts on academic test scores and grades.

Development.

Other programs purposefully focus on programming that is arguably non-academic, nor does it fit the (emerging) mold of SEL programming per se. These programs are concerned broadly with youth development (Halpern, 2002; Larson, 2000), and may include exposure to supportive adults, arts and culture, career opportunities, sports, etc. These programs generally operate on the premise that students, especially those in failing schools, *don't need more of the same* and instead need—and deserve—exposure to a variety of experiences, in different settings. These models tend to caution against an over-emphasis on academic pursuits that may deter access to much-needed developmental activities (Fulani & Kurlander, 2009; Riggs & Greenberg, 2004). Many programs with a developmental focus argue that these non-academic supports are of merit in and of themselves and may also indirectly impact academic gains by enacting on underlying skills that may be expanded through these kinds of experiences, including interest, confidence and self-awareness (Larson, 2000; Shernoff & Vandell, 2008; Vandell, 2013).

How is quality afterschool promoted?

While we know quite a bit about the association between after-school programming and a range of student outcomes, we know quite a bit less about how quality elements can effectively be promoted, at scale (Durlak & DuPre, 2008; Gardner, Roth & Brooks-Gunn, 2009; Granger 2010; Sheldon et al 2010; Yohalem, N & Wilson-Ahlstrom, 2010). However, emerging evidence is pointing to the value of systematically applying supports for afterschool sites in order to ensure that key aspects of quality programming are indeed being implemented.

Role of support organizations.

Support organizations that ensure quality programming are emerging as a key part of the equation for successfully providing widespread and equitable access to effective OST. Whether referred to as parent or intermediary organizations, “backbones,” or “quarterbacks.” these organizations have a strategic role in coordinating services, advocacy and fundraising, training and technical support, and promoting effective practices across diverse sites (for example, Collaborative for Building Afterschool Systems, 2005). The role of support organizations is an important one; across disciplines, there is a growing recognition that effective curricula or program models are *necessary but not sufficient* (Wandersman, Alia, Cook, Hsu & Ramaswamy (2016) for scaling best practices and that taking truly effective work to scale requires an application of an array of implementation supports that directly influences whether the provision of programming is of sufficient quality to produce intended outcomes (Durlak, 2013; Meyers, Durlak & Wandersman, 2012). These supports often come in the form of trainings, provision of tools, technical assistance or coaching, and evaluation including continuous quality improvement (Wandersman, Chien, Katz, 2012).

Models for supporting quality implementation at scale necessarily include multiple partners, or systems, working simultaneously on different aspects of the work. The Interactive Systems Framework (ISF) for example, posits a three-way partnership where 1) delivery systems (in this case, afterschool sites across the Dallas area) provide direct programming, 2) support systems (in this case, Dallas Afterschool) support programming through strategic services that combine training, provision of tools, coaching and technical assistance, and access to evaluation services and 3) a synthesis and translation system (in this case, CORE), that helps translate research-derived best practices into practical applications, monitor ongoing implementation, and evaluate emerging impacts (Wandersman, Duffy, Flaspohler, Noonan, Lubell, Stillman et al, 2008). The research partnership between DAS and CORE actively uses the ISF to frame our roles and work.

Support strategies.

Strategies for increasing site level afterschool quality can largely be summarized as “continuous quality improvement”: 1) training, 2) technical assistance or coaching, 3) tools and resources such as resource “kits” and 4) provision of data to support program decision making and/or to evaluate progress. An evaluation of the Communities Organizing Resources to

Advance Learning (CORAL) sites for example, examined the provision of targeted training, on-site coaching, and ongoing data collection and analysis and found that the quality of literacy instruction being provided at the participating sites increased with participation in CQI and that children exposed to the higher quality literacy instruction showed more gain in reading than those exposed to low quality literacy instruction (Sheldon, Arbreton, Hopkins & Grossman, 2010).

A second major contribution to this literature has been a study by the Weikart Center for Youth Program Quality which has successfully demonstrated that CQI supports including training and on-site coaching (their Youth Program Quality Intervention) produced site-level changes in a reasonable amount of time, about 18 months. Changes included more routinization of best practices, especially continuous quality improvement, at the site level. The CQI approach to supporting afterschool quality utilized by Dallas Afterschool (DAS) and tested in this study, is called the Program Quality Initiative (PQI) and it focuses on both organizational and programmatic level domains of quality.

Research Questions

Three major questions are asked and answered in this research; 1) how is afterschool quality being operationalized and measured, 2) does the delivery of training and coaching services to sites increase afterschool quality, and 3) does afterschool quality explain academic outcomes for students?

Methods

Measures

Program Quality Initiative (PQI).

Dallas Afterschool "quality advisors" (n=4) maintain logs about the amounts and types of training and coaching they provide to each site. Similar to other successful approaches to continuous quality improvement (CQI), the DAS PQI approach includes setting goals, providing targeted coaching, generalized training, and sustaining an ongoing evaluation of impacts for sites and for youth. DAS quality advisors maintain a log of all PQI activities provided to each participating site..

Assessing Afterschool Quality (AQuA).

Dallas Afterschool's Assessing Afterschool Quality (AQuA) assessment was originally adapted from the New York State Network for Youth Success's Quality Self-Assessment (QSA) Tool (http://networkforyouthsuccess.org/qsa/). Both the original QSA and AQuA are modeled after identifiable elements of quality in afterschool and reflect a hybrid approach in the field of OST measurement to include assessments of program management practices that affect staff-youth interactions as well as directly assessing staff-youth interactions themselves (see Table One).

When organizations first join the DAS network, they participate in an initial, baseline AQuA assessment. Following that, once annually, quality advisors interview organizational leaders, review documents, observe the site and score the site. AQuA scores are then used formatively to guide training and target coaching. AQuA scores are also provided to CORE for ongoing evaluation and monitoring purposes.

Representing the multi-dimensional and multi-level aspects of quality, AQuA consists of 74 total items, representing ten separate elements that span both organizational and programmatic features of quality: 1) Environment and Climate, 2) Relationships, 3) Programming and Activities, 4) Administration and Organization, 5) Staffing and Professional Development, 6) Linkages between Day and After School, 7) Youth Participation and Engagement, 8) Parent, Family, and Community Partnerships, 9) Program Sustainability and Growth, 10) Measuring Outcomes and Evaluation. An element score is calculated, as is an overall quality score combining all 10 elements[1].

Academic outcomes-student self-report.

A survey administered to 3rd-8th graders in participating sites measured four sub-dimensions of engagement comprised of three items each: valuing school, self-belief, persistence, and disengagement. The 4 point Likert Scale items were derived from the Motivation and Engagement Scale (Martin, A. J., 2012). An additional three items measuring global academic self-efficacy were also measured (Appleton, Christenson, Kim, & Reschly, 2006). 2015-16 was the first year this survey was administered to DAS students.

Academic outcomes-extant school district data.

1 CORE recommends use of all 10 element scores.

Through a data sharing agreement with the local school district of schools attended by student participants, CORE and DAS receive individual student data about a range of outcomes, including attendance, discipline, and scores on standardized assessments.

The Iowa Test of Basic Skills (ITBS) is a standardized norm-referenced achievement test given to all kindergarten through second grade students in the school district at the end of the 2014-2015 school year. The TerraNova is a series of standardized norm-referenced tests that were given to all kindergarten through second grade students in the school district at the end of the 2015-2016 school year after the district changed the assessment that would be given. We use the 2015 ITBS national percentile scores for language arts and math, and the 2016 TerraNova national percentile scores language arts and math for analyses. National percentile scores for each of the two years were converted to standard scores for each year (z scores).

The State of Texas Assessment of Academic Readiness (STAAR) is a series of state-required standardized assessments used in Texas public schools to assess students' knowledge of content taught based the state's curriculum standards. The annual assessment in reading and math is administered in grades 3–8 at the end of each academic year. The scaled scores used for analysis represent the conversion of the raw score onto a scale that is common to all forms of the assessment accounting for difficulty of the individual items; it quantifies student performance to relative proficiency standards.

Controlling for school quality.

As our dependent variables of interest (student engagement and efficacy, and student academic achievement) are highly influenced by characteristics of the schools each student attends, and because the DAS student sample represents schools of varying quality (113 different schools were attended by the 999 students enrolled in the study in 2015-2016), we included a score of school performance indicators. This score, derived from publicly available school performance indicators, specifies the number of state accountability ratings standards met by each public school in the State of Texas. Schools are required to meet standards in four areas of the performance index framework: student achievement, student progress, closing performance gaps, and postsecondary readiness. The index score indicates the number of standards met by the school a student attends (min=0, max=4). We downloaded index scores from the Texas Education Agency website (https://rptsvr1.tea.texas.gov/perfreport/account/2015/) and included a score in the dataset representing the school that each student in the SOP sample attended. For the purposes of this investigation, the number of indices met will be referred to as the School Quality Index (SQI).

Participants

Dallas Afterschool serves a full network of 134 afterschool sites. Not every member participates fully in all PQI services. Therefore, AQuA data about quality (at two time points) and PQI data about coaching and training provided for one full year are available for all 54 sites. A subset (n=15) of these 54 also participate in the Student Outcomes Project (SOP) research study, where student level data is also being collected. Throughout the presentation of results, we specify whether the full sample, or the SOP sub-sample is being described.

Parents of students in the SOP sites were invited to participate in the study, and students assented to participate. The study was reviewed and approved by SMU's Institutional Review Board. In total, 999 students representing 26 different afterschool sites were enrolled in the 2015-

16 SOP study; 372 had school data and 181 had valid gain scores and could be included in key analyses. A description of student demographics is presented in Table Two.

Results

First, descriptive statistics are provided, followed by results from planned analyses.

PQI Types and Dosage

Across 54 sites, a total of 751 hours of PQI was provided over one academic year. These hours were logged as either planning or delivery; delivery of PQI is further broken down into two additional categories: Coaching and Training (see Table Three) .

AQuA.

An overall AQUA score and 10 element indicator scores (all ranging 0-4) for all 54 sites was calculated at Time1 and Time2 (see Table Four). Prior to taking other factors into account, paired samples t-tests indicated that the overall change in AQUA was not significant and slight declines on elements four ($t(53)=4.34$, $p<.01$), five ($t(53)=5.95$, $p<.01$), six ($t(53)=3.31$, $p<.01$), eight ($t(53)=6.25$, $p<.01$), and ten ($t(53)=4.76$, $p<.01$), were statistically significant.

As an overall proxy for students' exposure to the relative quality of a site over the span of the 2015-16 school year, average AQuA scores were calculated and were used in planned analyses as an independent predictor of academic gains only (n=22 sites). These averages are presented in Table Five.

In one case, only one administration of AQuA was available, and the single score from spring 2016 was used as the indicator of site quality for the academic year.

Program attendance.

Each SOP site records daily attendance of enrolled students. An attendance rate for the 2015-16 year, or percent of days attended (*M*=0.76, *SD*=0.23, min=0, max=1.0), was derived from the number of days each student was enrolled in a program and the number of days actually attended for n=961 (96%) of the enrolled SOP students.

MES.

Students in grades 3 through 8 at 20 sites completed surveys to self-assess their frequency of positive behaviors in sub-dimensions of academic engagement and efficacy on scale of zero to four, with higher scores indicating higher self-rating of exhibiting positive academic behaviors (one scale, Disengagement, was reverse coded, as higher scores for this sub-dimension indicated high tendency to exhibit negative academic behavior). Average scores for each sub-dimension and the overall scale were calculated (see Table Six).

Academic outcomes.

National percentile scores for grades K-2, scale scores for 3-8, and gain scores used in planned analyses for all grade levels are presented in Table Eight. Paired samples t-tests indicated that the overall change in K-2 language arts ($t(93)=1.92$, $p=0.058$) and K-2 math ($t(55)=0.81$, $p=0.43$) were not significant. Gains in grades 3-8 reading ($t(158)=-4.73$, $p<0.001$) and 3-8 math ($t(152)=-5.49$, $p<0.001$) were statistically significant.

School Quality Index.

A School Quality Index (SQI) score was assigned to each student in the SOP sample that attended a Texas public school or charter school in 2015-2016 (n=563 students). Scores are based on which school they were enrolled in during the 2015-16 school year; the school district data provided only the most recent school a student was enrolled in so transfers to multiple schools within the 2015-16 year are not accounted for. SQI scores represent 102 different public schools that SOP students attended (M=3.5, SD=0.89, min=1 max=4).

Question 1: Measuring afterschool quality.

To answer the first research question of how afterschool quality is being operationalized and measured, CORE conducted exploratory factor analyses on DAS's AQuA tool, representing n=86 sites, all of whom received an AQuA rating in the fall of 2015, but may or may not have also participated in PQI services during the 2015-16 school year. The factor analyses utilized polychoric correlations as suggested by Zumbo, Gadermann, and Zeisser (2007) when the items are categorical. This was done in order to prove a psychometrically sound tool for measuring the quality of an after school program. Instead of using the traditional Cronbach's Alpha coefficient, which should be calculated with continuous items, ordinal Alpha (above .7 is desired) and the total Omega coefficients (Revelle, 2015) were calculated for the remaining items of each element. The ordinal coefficient is used as a reliability estimate because it takes into account the ordinal nature of the Likert-types response choices. The Fit and RMSE indexes for the factor structures of the each element are also reported with these reliability coefficients presented in Table Eight. Based on these results, it was determined that each element has acceptably high fit scores and could be used for the planned analyses. Additionally, CORE and DAS collaboratively made editing decisions to move or remove some items that were not contributing to the Fit and RMSE indexes. This resulted in a 74 item revised AQuA tool that is used in the current analyses.

Question 2: Does coaching and training improve afterschool quality?

In order to understand the relationship between training and coaching (PQI) and afterschool quality (AQuA), correlational and regression analyses were run using a sample of n=54 sites with these data. As an initial look at the relationship between PQI and AQUA, Pearson's bivariate correlations were run between AQUA overall, each of the 10 elements and the total number of PQI hours provided (see Table Nine). This provided sufficient evidence that there was a relationship (albeit negative) between PQI services and AQuA outcomes that should be further explored through our planned regression models.

. First, the total number of PQI hours was regressed onto overall gain in AQuA from Time1 to Time2; Time1 AQUA was also entered into the equation to account for each site's "starting point." Results showed that the overall model was significant (R^2=0.14, $F(3, 50)$=4.03, $p<.05$). However, the total number of PQI hours was not a significant predictor and Time1 AQUA was. A second regression split PQI hours into coaching and training and still entered Time1 AQUA as a predictor. This second model was not significant, indicating that dividing PQI into coaching and training activities does contribute additional explanation to the relationship between PQI and AQuA (see Table Ten).

Third, an additional series of regressions were run substituting overall gain in AQUA for gain on each of the 10 elements (see Table Eleven) and still controlling for Time1 AquA. Here, differences by elements were seen; element three (Programming & Activities) showed a positive and significant relationship between total hours of PQI and gain on that element (i.e. as PQI hours increased, the element score also increased). Second, some elements showed the opposite, a negative but significant relationship (i.e. as PQI increased, the element score decreased). Elements in this category were: five (Staffing and Professional Development), seven (Youth Engagement) and eight (Parent, Family & Community Partnerships). There was not a significant relationship between PQI and the other (n=6) elements of AQuA.

Question 3: Does afterschool quality predict student outcomes?

Last, we sought to understand the relationship between sites' AQuA scores and student outcomes; these analyses represent n=22 sites with valid site and student-level data..

Site quality and student surveys.

First, Pearson's bivariate correlations were run between AQuA overall, each of the 10 elements and the comprehensive score for the full academic engagement and efficacy survey (see Table Twelve). Self-reported academic engagement and efficacy is negatively correlated with program site quality scores for link to day and afterschool (element 6) (r=-0.28, n=135, p=0.001) and Parent, Family & Community Engagement (element 8) (r=-0.29, n=135, p<0.001), indicating that at sites where there was a more highly rated link to day and afterschool and parent, family, and community engagement, students tend to rate themselves as exhibiting *less* positive behaviors of academic engagement and overall academic efficacy[2].

Next, a series of regression models were used to examine the relationship between self-reported academic engagement and efficacy and program site quality as measured by AQuA. For each model, the AQuA score for the site each student attends is accompanied by two additional predictors: (1) to control for student exposure to the site programming, students' 2015-16 program attendance rate e is included, (2) the student's program attendance rate, to account for exposure to program services, and (2) and quality of the school attended by each student (SQI), to account for variability in school level impacts on academic outcomes. The first model was not significant (R^2=0.04, F(3, 74)=1.04, p=0.38). Second, a multiple regression model with all 10 AQuA elements, program attendance and school quality index as predictors was also not significant (R^2=0.14, F(12, 65)=0.92, p=0.5). None of the individual 10 element scores, nor program attendance or school quality index significantly contributed to the model after controlling for the other predictors included (see Table Thirteen).

Site quality and student academic outcomes.

[2] Additional Pearson's bivariate correlations between the overall AQuA score, 10 AQuA elements and each of the 6 subscales of the student survey revealed no additional significant findings indicating a single subscale was accounting for the observed relationship for the full scale and AQuA.

Finally, we looked at whether the AQuA score for a site that a student attended predicted their actual academic achievement, as measured by annual standardized reading and math assessments. Pearson's bivariate correlations were run between AQuA overall, each of the 10 elements and the math and reading gain scores (see Table Fourteen). No significant correlations were observed between early elementary (kindergarten – 2nd grade) math gain and overall site quality nor any of the 10 individual elements. For Element 5, change in early elementary literacy is negatively correlated with Element 5 (program staffing and professional development quality), $r=-0.27$, $n=59$, $p=0.04$.

For upper elementary math, the overall AQuA score ($r=-0.20$, $n=100$, $p=0.04$), Element 2 score ($r=-0.30$, $n=100$, $p=0.003$), Element 4 score ($r=-0.24$, $n=100$, $p=0.02$), Element 7 score ($r=-0.24$, $n=100$, $p=0.02$), Element 9 score ($r=-0.30$, $n=100$, $p=0.003$), and Element 10 score ($r=-0.21$, $n=100$, $p=0.04$) were all negatively correlated with upper elementary and middle school math gains, suggesting that students at programs with higher quality AQuA scores on a range of elements have lesser gains in math performance year-over-year. Similarly, for Element 9, change in upper elementary and middle school literacy is negatively correlated with program sustainability and growth, $r=-0.20$, $n=103$, $p=0.04$.

A series of regression models were used to examine the relationship between math and literacy student gains and program site quality as measured by AQuA. For each model, the AQuA score for the site each student attends is accompanied by three additional predictors: (1) to control for time 1 achievement, students' spring 2015 standardized test score is included and (2) performance of the school attended by each student (SQI), to account for variability in school level impacts on academic outcomes, and (3) program attendance rate, to account for the amount of program “dosage” each student received at the site he/she attended.

Early Elementary Mathematics

The first model, with overall AQuA, time 1 mathematics score, school quality index, and program attendance rate as predictors and mathematics standardized assessment z score gain from 2015 to 2016 as the dependent variable, was not significant (R^2=0.22, F(4, 30)=2.17, p=0.09). School quality index score had a significant negative regression weight, indicating that students at schools with higher quality index scores had lower gains after controlling for other variables in the model. Second, a multiple regression model with all 10 AQuA elements, time 1 math score, school quality index and program attendance as predictors was not significant (R^2=0.24, F(7, 27)=1.24, p=0.32). School quality index and elements 1, 2, 4, 7 and 9 were omitted from the model due to being collinear with other predictor variables[3]. In this adjusted model, none of the individual element scores significantly contributed to predicting early elementary math gains after controlling for the other predictors included (see Table Fifteen).

Early Elementary Literacy

The first model, with overall AQuA, time 1 literacy score, school quality index, and program attendance rate as predictors and literacy standardized assessment z score gain from 2015 to 2016 as the dependent variable, was statistically significant (R^2=0.34, F(4,53)=6.73, p<0.001). The time 1 literacy score and school quality index score had significant and negative

[3] Multicollinearity between two or more predictor variables increases the difficulty to attribute contribution to influence on the independent variable to any one predictor.

regression weights, indicating that students at schools with higher quality index scores and with higher time 1 literacy scores had less gain after controlling for other variables in the model. The 2015-16 overall AQuA score had a significant positive regression weight, indicating that higher quality programs are expected to produce greater literacy gain scores after controlling for the other predictors in the model. For every 1 point increase in AQuA quality score (on a 0-4 scale), gain scores are predicted to increase by 4.0 standard deviations (gain scores are standardized z scores). However, the range of AQuA scores for sites only ranges 0.65 points, so a more reasonable interpretation would be to conclude that an increase of 0.25 points on AQuA predicts a 1 standard deviation increase in literacy gain scores.

Second, a multiple regression model with all 10 AQuA elements as individual predictors was also significant (R^2=0.40, F(9, 48)=3.5, p=0.002). Elements 1, 3, 4, and 7 were omitted from the model due to being collinear with other predictor variables. In this adjusted model, one individual element, Element 6: Link to Day and Afterschool, significantly contributed to predicting early elementary literacy gains after controlling for the other predictors included (see Table Sixteen), while time 1 literacy z score had a significant negative relationship, indicating that students with higher time 1 literacy z scores had lower gains after controlling for other variables in the model.

Upper Elementary and Middle School Mathematics

The first model, with overall AQuA, time 1 mathematics scale score, program attendance rate and school quality index as predictors and math standardized scale score gain from 2015 to 2016 as the dependent variable, was statistically significant (R^2=0.26, F(4, 93)=8.28, p<0.001). The time 1 math scale score had a significant negative regression weight, indicating that students with higher time 1 math scale scores had lower gains after controlling for other variables in the model. The overall AQuA score did not significantly contribute to math gain scores. Second, a multiple regression model with all 10 AQuA elements as predictors was also significant (R^2=0.42, F(13, 84)=4.59, p<0.001). In this adjusted model, none of the individual element scores significantly contributed to predicting upper elementary/middle school math gains after controlling for the other predictors included (see Table Seventeen), while time 1 math scale scores maintained a significant negative relationship, indicating that students with higher time 1 math scale scores had lower gains after controlling for other variables in the model.

Upper Elementary and Middle School Reading

The first model, with overall AQuA time 1 reading scale score and school quality index as predictors and reading standardized scale score gain from 2015 to 2016 as the dependent variable, was statistically significant (R^2=0.17, F(4, 96)=4.92, p<0.001). The time 1 reading scale score had a significant negative regression weight, indicating that students with higher time 1 reading scale scores had lower gains after controlling for other variables in the model. The overall AQuA score did not significantly contribute to reading score gains. Second, a multiple regression model with all 10 AQuA elements as predictors was also significant (R^2=0.22, F(13, 87)=1.86, p=0.04). In this adjusted model, none of the individual element scores significantly contributed to predicting reading gains after controlling for the other predictors included (see Table Eighteen), while time 1 reading scale scores maintained a significant negative relationship, indicating that students with higher time 1 reading scale scores had lower gains after controlling for other variables in the model.

Summary of Results: AQuA and Student Outcomes

Individual AQuA elements alone, after simultaneously controlling for the effects of each of the other 10 elements, did not significantly predict gains in math achievement, upper elementary and middle school reading achievement, nor self-reported academic engagement and efficacy. A single element, Element 6, was alone a significant positive predictor of early literacy gains, after controlling for the effects of the other 9 elements, school quality, time 1 achievement, and program attendance. However, in models including all AQuA elements, time 1 scores, school quality indices, and program attendance were sound models, significantly predicting 14% of variability in self-reported academic engagement and efficacy, 24% of variability in early math gains, 40% of variability in early literacy gains, 42% of variability in upper elementary/middle school math gains, and 22% of the variability in upper elementary/middle school reading gains.

Interpretation and Conclusions

Analyses indicated that provision of PQI positively impacted only one element of afterschool quality, Activities and Programming, and that higher overall quality was related to positive gains for literacy in lower elementary aged students. These findings partially support the theory of change being tested here, that training and technical assistance can increase site quality and that site quality in turn, drives positive student outcomes. However, the findings are not robust overall and additional analyses with a stronger research design are warranted. Below, interpretation for the findings related to each research question is provided.

Question 1: Measuring Quality. Mixed results in these analyses point to a need for further refinement and clarity around the construct and measurement of quality. Results from the factor analyses do point to a psychometrically sound tool that can be used to estimate quality at single time points and to track change over time. However, elsewhere in the paper AQuA is alternatively used as an outcome (of PQI efforts) and as a predictor (of student outcomes). It is clear that all underlying elements are not equal; some may be more responsive to PQI under certain circumstances, and some may change more quickly than others. Each of these are important implications when considering AQuA both as an outcome of training and coaching and also as a reliable predictor of student outcomes.

For instance, all 10 AQuA elements contributed some explanation, particularly when looking at impacts of PQI, but when exploring the relationship between quality and student outcomes, the single, overall measure of AQuA was a more robust predictor. Issues related to conceptualizing and measuring quality are pervasive in the OST field right now, leading us to conclude that AQuA is likely as strong of a measurement instrument as is currently available and as the field further refines this process, DAS should continue its CQI process and adapt and improve the AQuA tool as warranted.

Some of the best evidence that AQuA is a robust indicator of quality stems from results from Questions Two and Three, combined. That is; while PQI seems to act on individual, underlying elements of quality (Activities and Programming for example), an overall single quality score was a better predictor of student outcomes. This suggests that tailored coaching and training must address different underlying aspects of quality at a given site, paying attention to the time it takes for certain practices to be taken up and the “level” at which new or expanded practices may occur. And yet, youth seem to experience quality as a gestalt, represented by the overall AQuA score, and are less directly impacted by the particular nuances that adults in the system concern themselves with. A recommendation, then, for utilizing quality indicators is to use a multi-dimensional metric when examining afterschool quality as an outcome of a capacity or quality-building intervention and a single metric when using quality to predict student-level outcomes.

Question 2: PQI and AQuA. Results showed a significant impact of PQI on only one element; Activities and Programming (Element 3). The staff behaviors and site or program-level organizational features that are associated with Programs and Activities may have received more focus during PQI (DAS's quality advisors are constantly refining and improving their monitoring and tracking process), and may have been most malleable and actionable. Importantly, the activities associated with this element may also be most temporal and likely to show up in the change in quality over just one year. Change in other elements (perhaps those that encompass community partnerships, family engagement or staff turnover) likely take for the impacts of PQI to be seen.

Additionally, some negative impacts of PQI were seen (on Elements 5, 7 and 8). We suspect that a self-reference bias is in effect here, where both DAS quality advisors and site level staff who supply much of the information utilized in an AQuA assessment are susceptible. As site level staff work with their quality advisors and become more aware of what quality "looks like" in each domain and become generally more well-informed they may be giving lower ratings. This interpretation, however, begs the question of why some elements may be susceptible to the bias while others are not. Again, this may be related to how malleable the specific behaviors and practices embedded in each element really are, and in what timeframe.

Dividing PQI into its component parts of training and coaching did not help explain variation in site level AQuA outcomes though continued refinement in DAS's tracking of sub-types and focus areas of training and coaching may further illuminate this relationship in future analyses. An AQuA score at Time1, however, was a significant predictor of overall gain in AQuA (for n=54 sites that had two AQuA scores). This means that sites that are already performing somewhat well on AQuA (i.e. have some requisite level of capacity already in place) appear to benefit most from DAS's provision of PQI, supporting a hypothesis of site "readiness" for these support services, a concept that is gaining traction across multiple fields (CITE).

Further, prior to adjusting for Time 1 scores, the relationship between PQI hours and gain in AQUA is negative: as more hours were spent, the change score actually decreases. However, once adjusting for Time 1 AQUA scores, across elements, the relationship reverses. This underscores the importance of utilizing strong and defensible models to test for impacts, which unfortunately, are not always par for the course in OST evaluation work. Had we only looked at whether number of PQI hours was correlated with AQuA gain we would have concluded, erroneously, that PQI was actually *hurting* site quality. By utilizing a stronger model we were able to see the positive impacts, after one year, on a key element of afterschool quality, the Activities & Programming provided directly to youth.

Question 3: Quality Afterschool and Student Outcomes. The relationship between afterschool programming and student outcomes has been rigorously demonstrated elsewhere (Lauer et al, 2006), and quite arguably with more methodological rigor than was available to us in this sample. Our results showed that AQuA is not a significant predictor of self-reported academic engagement and efficacy, mathematics achievement gains for students of any grade level, nor of upper elementary/middle school literacy gains[4]. Our analyses did show however, that AQuA is a positive predictor of early elementary literacy gains. This in and of itself is a meaningful finding given the potency of early literacy skill sets as a foundation for later learning. Ongoing data collection in the partnership between CORE and DAS will allow us to follow these students over time and understand how their early elementary literacy scores may have long term implications for success in school.

In all models using extant school level data, two themes were consistent. First, the inclusion of the School Quality Index as a proxy for school level quality contributed very little to the model in and of itself, yet models where SQI was included alongside AQuA, program attendance rate, and time1 academic scores explained a significant amount of variability in outcomes. Though we hypothesized that the overall quality of the school would explain a large amount of variance in our outcomes of interest (therefore showing us what impacts quality OST may have above and beyond what influence a student's school may have), these results suggest that it is much more likely a complex interaction between school and site quality that best predict student outcomes.

Second, in each instance, the contribution of a time1 academic score was meaningful, and once again, the inclusion of multiple predictors in the models actually reversed the direction of findings. For instance, prior to adjusting for time 1 scores, the relationship between AQuA and early elementary literacy gains was slightly negative, yet not significant. However, after adjusting for time 1 early literacy standardized scores, program attendance, and school quality, the overall AQuA score was a positive predictor of early elementary gains in reading

[4] Previous CORE analyses have shown that DAS elementary students outperform non-DAS students in mathematics.

achievement. With early literacy, as well as upper elementary/secondary reading and math, lower time 1 scores, which are an indicator of student achievement prior to receiving after school services in 2015-16, predict higher gain scores. This indicates that students who start off lower are likely to make greater gains and supports an approach to providing access to the highest quality OST to the students most at risk for school failure.

This highlights a paradox in multi-level research where we see discrepant findings for students compared to sites. That is, sites that are already performing a little better seem to get the most out of PQI services, yet students who are under-performing seem to get the most out of high quality afterschool. Organizations that demonstrate moderate levels of preliminary quality benefit most by focusing more resources on students with low levels of preliminary performance. This underscores a need for our field—support agencies, programs that deliver services directly to youth and families, and the research-practitioners involved in ongoing quality improvement—to help transform a cohort of relatively capable OST sites into truly excellent providers such that greater numbers of students most at risk can be reached.

Limitations

There are a few notable limitations to this design. First, the PQI staff who rate sites on AQuA are also the same Quality Advisors providing coaching & training who may be subject to bias in reporting. Therefore, while their ratings are not those of a true external rater, this does constitute an improvement over a general reliance in the field on self-assessments. A second limitation, due to sample size, our analyses, pool all DAS SOP sites into one group, though there are different types of programming taking place at each site. For sample, some SOP sites in these analyses focus on reading, while others are more generalist, focusing on mentoring or SEL-type interventions. Variations in the type of programming provided is not linked to outcomes in these analyses. As additional SOP sites are added, we plan to conduct sub-analyses that will link reading programming to reading outcomes, math programs to math outcomes, and SEL/mentoring programs to non-cognitive outcomes. However, given the hybrid approach that many afterschool sites utilize, this kind of differentiation may end of being of limited utility. Last, though we believe it was a strength to leverage existing evaluation data into analyses, the realities of community-based data collection meant that data was not missing at random. This means that the full sample of DAS sites was not represented in the SOP study and even within the SOP study, academic indicators are only available for about 30% of enrolled students. There are undoubtedly meaningful differences between sites whom DAS recruits for participation in the SOP study and not, whether these differences show up in mean AQuA ratings, as well as differences between students whose parents consent for participation in SOP and those who do not. These non-random effects are a reality of community-based evaluation work, and do present limitations for generalizability of findings.

Next Steps

CORE and DAS plan at least two additional designs for future testing and call on the broader field to contribute to this growing body of knowledge as well. First, in order to strengthen a research design to examine effects of PQI on underlying elements of quality, a design where sites are randomized into different PQI conditions is recommended. Here, full PQI (all available training, tools, technical assistance and continuous quality improvements) could be provided to treatment sites while other sites receive only partial or no supports. Alternatively, in instances where a large enough sample and sufficient data is available, a propensity score matching technique could be used to rigorously address existing differences in sites that could be contributing to variability in outcomes. Second, in samples that are sufficiently large and where program characteristics themselves are well differentiated, linking program content to relevant outcomes (for example, reading programs to reading outcomes) and comparing this to a "pooled" approach where any quality afterschool program, regardless of content, is being used as a predictor for a range of outcomes, is appropriate. Last, our analyses did not show impacts on student-self reported attitudes related to academics. Further exploration to link proximal changes in student attitudes toward school to later academic performance in order to explicate an underlying results chain is warranted.

References

AfterSchool Alliance (2004). America after 3 PM: Afterschool programs in demand. Washington, DC: Afterschool Alliance.

Appleton, J.J., Christenson, S.L., Kim, D., & Reschly, A.L. (2006). Measuring cognitive & psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, 44, 5, 427-445.

Durlak, J. A., & Weissberg, R. P. (2007). The impact of after-school programs that promote personal and social skills. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.

Durlak, J.A., Mahoney, J.L., Bohnert, A.M. & Parente, M.E. (2010). Developing and improving after-school programs that seek to promote personal and social skills in children and adolescents: A special issue of AJCP. *American Journal of Community Psychology*, *45*, 285-293

Durlak, J.A., Weissberg, R., & Pachan, M. (2010). A meta-analysis of afterschool programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology*, *45*, 294-309.

Eccles, J., & Gootman, J.A. (Eds). (2002). Community programs to promote youth development. Washington, DC: National Academy Press.

Fulani, L.B., & Kurlander, G.L. (2009). Achievement gap or development gap?; "Outliers" and outsiders reconsider an old problem. Special Report. The All Stars Project.

Granger, R. (2010). Understanding and Improving the Effectiveness of After-School Practice. *American Journal of Community Psychology*, 45(3-4), 441-446.

Gullotta, T.P. (2015). After-school programming and SEL. In: Handbook of Social and Emotional Learning; research and practice. Durlak, J.A., Domitrovich, C.E., Weissberg, R., & Gullotta, T.P. The Guilford Press, New York, NY. 260-266

Greene, K. M., Lee, B., Constance, N., & Hynes, K. (2013). Examining youth and program predictors of engagement in out-of-school time programs. *Journal of Youth and Adolescence,42* (10),1557-72.

Halpern, R. (2002). A different kind of child development institution: The history of after-school programs for low-income children. *Teachers College Record,* 104 (2). 178-211.

Harvard Family Research Project (2008). Afterschool programs in the 21st century; Their potential and what it takes to achieve it. *Issues and Opportunities in Out Of School Time Evaluations*, 10.

Hirsch, B. J., Mekinda, M. A., Stawicki, J. (2010). More than attendance: The importance of afterschool program quality. *American Journal of Community Psychology*, 45(3-4), 447-452.

Lauer, P.A., Akiba, M., Wilkerson, S. B., Apthorp, H.A. Snow, D., & Martin-Glenn, M.L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, *76*, 2, 275-313.

Larson, R. (2000). Toward a psychology of positive youth development. *The American Psychologist*, 55, 170-183.

Larson, R. W.., Walker, K. C. (2010). Dilemmas of Practice: Challenges to Program Quality Encountered by Youth Program Leaders. *American Journal of Community Psychology*, 45(3-4), 338-349.

Libbey, H.P, (2004). Measuring student relationships to school: Attachment, bonding, connectedness & engagement. *Journal of School Health*, *74*, 7, 274-283.

Mahoney, J.L., Vandell. D.L. Simpkins, S., & Zarrett, N. (2009). Adolescent out-of-school activities. In R.M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology: Vol 2, Contextual influences on adolescent development* (3rd ed. Pp 228-269). New York, NY. Wiley.

Martin, A. J. (2012). The motivation and engagement scale (12th ed.). Sydney, NSW.: Lifelong Achievement Group. Retrieved from http://www.lifelongachievement .com.

Moore, K.A., Murphey, D., Bandy, T, & Cooper, M. (2014). Participation in Out-of-School Time Activities and Programs. *Child Trends*. Retrieved from: http://www.childtrends.org/wp-content/uploads/2014/03/2014-13OutofSchoolActivities1.pdf

National Institute on Out-of-School Time (2009). Making the Case: A 2009 Fact Sheet on Children and Youth in Out-of-School Time. Retrieved from: http://www.niost.org/pdf/factsheet2009.pdf

Revelle, W. (2015). Procedure for psychological, psychometric, and personality research. *Package 'psych'*.

Riggs, N., & Greenberg, M.T. (2004). Afterschool youth development programs; A developmental-ecological model of current research. *Clinical Child and Family Psychology Review, 7*, 177-190.

Roth, J.L., Malone, L.M., Brooks-Gunn, J. (2010). Does the amount of participation in afterschool programs relate to developmental outcomes?; A review of the literature source. American Journal of Community Psychology, 45, 430-444.

Sheldon J., Arbreton A., Hopkins L., Grossman J. (2010). Investing in success: Key strategies for building quality in after-school programs. *American Journal of Community Psychology,* 45(3-4), 394-404.

Shernoff, D. (2010). Engagement in after-school programs as a predictor of social competence and academic performance. *American Journal of Community Psychology, 45,* 430-444.

Smith, C. Akiva, T. Sugar, S., Lo Y.J., Frank, K.A., Peck, S.C. Cortina, K.S. & Devany, T. (2012). Continuous quality improvement in afterschool settings: Impact findings from the Youth Program Quality Intervention study. Washington, DC: The Forum for Youth Investment.

Smith C., Peck Stephen C., Denault A,, Blazevski J., Akiva T. (2010). Quality at the point of service: Profiles of practice in after-school settings. *American Journal of Community Psychology*, 45(3-4), 358-369.

Vandell, D.L. (2013). Afterschool program quality and student outcomes: Reflections on positive key findings on learning and development from recent research. Expanding Minds & Opportunities, Recent Evidence of Impact. 180-186.

Wandersman, Alia, Cook, Hsu & Ramaswamy (2016) Evidence-based interventions are necessary but not sufficient for achieving outcomes in a complex world: Empowerment

Evaluation, Getting to Outcomes, and demonstrating accountability. *American Journal of Evaluation*, 1-18.

Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., Blachman, M., Dunville, R., & Saul, J. (2008). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. *American Journal of Community Psychology*, 41, 171-181.

Wndersman, A. Chien, V.,H., Katz, J. (2012). Toward an evidence-based system for innovation support for implementing innovations with quality: tools, training, technical assistance, and quality assurance/quality improvement. *American Journal of Community Psychology*, 50, 3-4, 445-59.

Weiss, H., Little, P., Bouffard, S. M., Deschenes, S. N., & Malone, H. J. (2009). *The federal role in out-of-school learning: After-school, summer learning, and family involvement as critical learning supports*. Cambridge, MA: Harvard Family Research Project.

Yohalem, N., Wilson-Ahlstrom, A. (2010). Inside the Black Box: Assessing and Improving Quality in Youth Programs. *American Journal of Community Psychology,* 45(3-4), 350-357.

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, *6*(1), 21-

Table One: Crosswalk of DAS's AQuA Elements and Synthesis of the Literature on Afterschool Quality Assessments (Yohalem & Wilson-Ahlstrom, 2010).

Dallas Afterschool's AQuA Elements	**Six Core Concepts* + Four Common Concepts**
1. Environment & Climate	Environment* – aspects of program climate and setting, including physical and emotional safety
2. Relationships	Relationships* – connections between and among youth and adults in the program
3. Programming & Activities	Skill building* – opportunities for participants to develop specific skills by participating in intentional learning activities
4. Administration & Organization	Program management
5. Staffing & Professional Development	Staffing – ratios, qualifications
6. Linkages Between Day & Afterschool	
7. Youth Participation & Engagement	Engagement* - extent to which children, youth and staff are meaningfully involved in program activities Youth leadership/participation
8. Parent, Family & Community Partnerships	Linkages to family and community
9. Program Sustainability & Growth	
10. Measuring Outcomes & Evaluation	
	*the common concepts, seen across all nine reviewed assessments are indicated with (*)

Table Two: Demographic Characteristics of Students in 2015-16 SOP Study

Variable	MES Survey (n=181)	K-2 Math (n=56)	K-2 Literacy (n=94)	3-8 Math (n=147)	3-8 Reading (n=150)
Gender (Female)	48%	41%	51%	55%	55%
Race (African American)	15%	16%	10%	8%	8%
Race (Hispanic/ Latino)	71%	68%	81%	73%	73%
Grade Level					
Kinder					
1st		50%	51%		
2nd		50%	49%		
3rd	31%			2%	2%
4th	24%			34%	33%
5th	25%			42%	41%
6th	13%			11%	11%
7th	2%			8%	8%
8th	4%			3%	5%
Grade level unknown	1%				
Number of SOP sites represented	20	9	11	17	17

Table Three: Descriptive Statistics on PQI Activities (n=54 sites)

	M	(SD)	Min	Max	Sum
Total Hours PQI	13.91	8.96	2.25	29.25	751.00
Planning	4.04	2.50	1.00	10.00	218.00
Delivery	9.87	6.84	1.25	23.25	533.00
Training	5.96	5.05	0	16	322.00
Coaching	8.00	5.40	2	25.75	432.25

Table Four: Descriptive Statistics on AQUA Scores (n=54 sites)

	T1		T2		Change from T1 to T2	
	M	(SD)	M	SD	M	(SD)
Overall AQuA Score	2.74	.18	2.72	.15	-0.13	.21
E1: Environment & Climate	2.81	.22	2.76	.37	-0.05	.34
E2: Relationships	2.81	.27	2.75	.39	-0.07	.35
E3: Programming & Activities	2.71	.25	2.64	.41	-0.07	.27
E4: Admin & Organization	2.97	.28	2.86	.36	-0.10*	.17
E5: Staffing & PD	2.92	.17	2.70	.34	-0.23*	.27
E6: Link Day & Afterschool	2.66	.32	2.44	.41	-0.22*	.48
E7: Youth Engagement	2.45	.23	2.42	.42	-0.03	.32
E8: Parent, Family, Community	2.63	.30	2.27	.42	-0.34*	.42
E9: Sustainability & Growth	2.70	.30	2.69	.32	-0.01	.09
E10: Outcomes & Evaluation	2.64	.26	2.46	.42	-0.18*	.27

P<.01

Table Five: Descriptive Statistics of 2015-16 average AQuA scores (n=22 sites)

	M	(SD)	Min	Max
Overall AQuA score	2.60	0.22	2.26	2.91
E1: Environment & Climate	2.74	0.23	2.40	3.20
E2: Relationships	2.70	0.28	2.15	3.14
E3: Programming & Activities	2.65	0.22	2.20	2.92
E4: Admin & Organization	2.69	0.35	2.11	3.34
E5: Staffing & PD	2.73	0.23	2.32	3.01
E6: Link Day & Afterschool	2.55	0.24	2.10	3.20
E7: Youth Engagement	2.37	0.27	1.80	2.80
E8: Parent, Family, Community	2.32	0.33	1.84	3.00
E9: Sustainability & Growth	2.87	0.34	2.33	3.67
E10: Outcomes & Evaluation	2.55	0.43	1.69	3.13

Table Five: Descriptive Statistics of 2015-16 average AQuA scores

	Question 1 (n=86 sites)		Question 2 (n=54 sites)		Question 3 (n=22 sties)		Difference between Q2 & Q3 AQuA Sample
	M	SD	M	SD	M	SD	*p* value
Overall AQuA score	2.69	0.34	2.62	0.27	2.60	0.22	0.76
E1: Environment & Climate	2.79	0.29	2.73	0.28	2.74	0.23	0.88
E2: Relationships	2.80	0.36	2.72	0.33	2.69	0.28	0.71
E3: Programming & Activities	2.69	0.41	2.63	0.33	2.65	0.22	0.80
E4: Admin & Organization	2.94	0.43	2.87	0.34	2.69	0.35	0.04*
E5: Staffing & PD	2.85	0.31	2.78	0.26	2.73	0.23	0.44
E6: Link Day & Afterschool	2.58	0.44	2.47	0.31	2.55	0.24	0.28
E7: Youth Engagement	2.48	0.42	2.39	0.33	2.37	0.27	0.80
E8: Parent, Family, Community	2.43	0.42	2.34	0.36	2.32	0.33	0.82
E9: Sustainability & Growth	2.72	0.39	2.69	0.31	2.87	0.34	0.03*
E10: Outcomes & Evaluation	2.62	0.40	2.52	0.35	2.55	0.43	0.75

*Difference between two samples' AQuA scores is significant at $p<0.05$.

Table Six: Descriptive Statistics of student survey (n=181 students)	M	(SD)	# of items
Global Academic Efficacy	3.36	.51	3
Valuing School	3.49	.53	3
Disengagement	3.06	.92	3
Persistence	3.31	.60	3
Self Belief	3.54	.53	3
Full Scale	3.36	.41	3

Table Seven: Descriptive Statistics of student academic outcomes

	T1 (Spring 2015)		T2 (Spring 2016)		Change from T1 to T2	
	M	(SD)	M	SD	M	(SD)
K-2 Math** (n=56)	55.75	25.37	50.32	24.78	-0.10	0.89
K-2 Language Arts** (n=94)	53.89	27.71	48.75	25.99	-0.20	0.99
3-8 Math (3-8) (n=147)	1478.51	165.25	1581.35	144.75	102.84*	94.62
3-8 Reading (3-8) (n=150)	1445.07	166.18	1524.92	152.40	79.90*	83.62

*p<.01
**Gain scores of national percentile scores were not used for analyses and, therefore, were not calculated. Changes in standardized conversions of national percentile scores scores (z scores) are presented.

Table Eight. Fit Indexes and Reliability Coefficients for AQuA (n= 86 sites)

	RMSE	FIT	Ordinal Alpha	Omega Total
E1: Environment & Climate	0.26	0.51	0.33	0.84
E2: Relationships	0.08	0.96	0.79	0.85
E3: Programming & Activities	0.09	0.97	0.92	0.94
E4: Admin & Organization	0.17	0.86	0.84	0.95
E5: Staffing & PD	0.18	0.86	0.84	0.95
E6: Link Day & Afterschool	0.04	.98	0.58	0.71
E7: Youth Engagement	0.07	0.97	0.83	0.88
E8: Parent, Family, Community	0.07	0.89	0.54	0.71
E9: Sustainability & Growth	0	1	0.35	0.42
E10: Outcomes & Evaluation	0.12	0.93	0.83	0.91

Table Nine: Pearson's Correlation between PQI Hours and AQUA Element Gain Scores (n=54)

	R
Overall AQuA Score	-0.19
E1: Environment & Climate	0.16
E2: Relationships	-0.48
E3: Programming & Activities	-0.06
E4: Admin & Organization	-0.23*
E5: Staffing & PD	-0.47*
E6: Link Day & Afterschool	-0.02
E7: Youth Engagement	-0.34*
E8: Parent, Family, Community	-0.14
E9: Sustainability & Growth	-0.05
E10: Outcomes & Evaluation	-0.03

*Correlation is significant at p<0.05.

Table Ten: Regression for PQI Hours on Overall AQuA Gain (n=54)

Predictor Variable	Model 1			Model 2		
	B	*SE B*	β	*B*	*SE B*	β
Total Number PQI Hours	.001	.004	.05	--	--	--
Time1 AQUA	0.47	0.20	0.40*	0.45	0.20	0.39*
Total Number PQI Hours - Training				0.001	0.007	0.020
Total Number PQI Hours - Coaching				0.001	0.006	0.021
R^2		0.14			0.084	
F for change in R^2		4.03*			2.61	

*Model is significant at p < .05.

Table Eleven: Regression for PQI Hours and Gain on Each AQUA Element (n=54)								
Dependent Variable			Time 1 Element Score			Total # PQI Hours		
	R^2	*F for* R^2	*B*	*SE B*	β	*B*	*SE B*	β
E1: Environment & Climate	0.05	1.28	-0.26	0.24	-0.17	0.00	0.01	0.081
E2: Relationships	.080	2.18	-0.40	0.20	-0.31*	-0.01	0.01	-0.196
E3: Programming & Activities	0.33	12.76*	0.62	0.12	0.67*	0.01	0.01	0.279*
E4: Admin & Organization	0.06	1.65	0.06	0.10	0.10	-0.00	0.00	-0.175
E5: Staffing & PD	0.23	7.44*	-0.12	0.21	-0.08	-0.02	0.01	-.501*
E6: Link Day & Afterschool	0.30	10.98*	-0.84	0.18	-0.57*	-0.01	0.01	-0.18
E7: Youth Engagement	0.11	3.23*	0.03	0.19	0.02	-0.02	0.01	-0.33*
E8: Parent, Family, Community	0.21	6.69*	-0.64	0.18	-0.46*	-0.02	0.01	-0.29*
E9: Sustainability & Growth	0.01	0.19	0.03	0.06	0.09	0.01	0.01	0.12
E10: Outcomes & Evaluation	0.05	1.25	0.24	0.15	0.23	0.01	0.01	0.05

*Model is significant at p < .05.

Table Twelve: Pearson's Correlation between student survey average score and 2015-16 AQUA element scores (n=18 sites and n=135 students)							
	R						
	Total Score (all sub-scales)	GAE	SB	P	VS	VOST	D
AQUA overall	-0.12	-0.10	-0.14	-0.19*	0.01	0.01	-0.06
E1: Environment & Climate	-0.10	-0.07	-0.12	-0.14	-0.01	0.02	-0.07
E2: Relationships	-0.17	-0.19*	-0.13	-0.24*	-0.02	-0.05	-0.06
E3: Programming & Activities	-0.05	-0.02	-0.14	-0.11	-0.03	0.04	0.04
E4: Admin & Organization	0.11	0.01	0.01	-0.03	-0.16	0.12	0.11
E5: Staffing & PD	-0.11	-0.07	-0.11	-0.16	-0.01	0.05	-0.13
E6: Link Day & Afterschool	-0.28*	-0.21*	-0.23*	-0.29*	-0.15	-0.17	-0.10
E7: Youth Engagement	-0.02	-0.05	-0.09	-0.10	0.05	0.03	0.03
E8: Parent, Family, Community	-0.29*	-0.21*	-0.18*	-0.29*	-0.12	-0.11	-0.24*
E9: Sustainability & Growth	0.07	0.11	-0.01	-0.02	0.08	0.09	0.02
E10: Outcomes & Evaluation	-0.09	-0.06	-0.10	-0.12	0.04	-0.04	-0.09

*Correlation is significant at p<0.05
Sub-scales: GAE=global academic efficacy; SB=self-belief; P=persistence; VS=valuing school; VOST=valuing OST; D=disengagement (negative item, reverse coded)

Table Thirteen: Regression for 2015-2016 AQuA and Academic Engagement & Efficacy scores

Predictor Variable	Model 1			Model 2		
	B	*SE B*	β	*B*	*SE B*	β
AQuA Overall	-0.04	0.19	-0.03	--	--	--
School quality index score	0.03	0.07	0.05	0.11	0.12	0.27
Program attendance rate	0.38	0.23	0.19	0.52	0.32	0.19
Element 1				0.22	0.56	0.13
Element 2				0.33	0.49	0.32
Element 3				-0.43	0.62	-0.19
Element 4				0.39	0.37	0.44
Element 5				0.01	0.37	0.01
Element 6				0.20	0.48	0.13
Element 7				-0.99	0.84	-0.74
Element 8				-0.27	0.29	-0.24
Element 9				0.57	0.38	0.50
Element 10				-0.28	0.45	-0.36
R^2		0.04			0.14	
F for change in R^2		1.04			0.90	

Table Fourteen: Pearson's Correlation between academic gain scores and 2015-16 AQUA element scores (n=22 sites; n=372 students)

	R			
	K-2 Math	K-2 Literacy	3-8 Math	3-8 Reading
AQUA overall	-0.16	-0.05	-0.20*	-0.11
E1: Environment & Climate	-0.11	0.04	0.06	0.03
E2: Relationships	-0.12	0.10	-0.30**	-0.04
E3: Programming & Activities	-0.21	0.04	-0.18	-0.11
E4: Admin & Organization	-0.13	-0.03	-0.24*	-0.08
E5: Staffing & PD	-0.31	-0.27*	-0.01	-0.16
E6: Link Day & Afterschool	-0.06	-0.07	0.02	-0.08
E7: Youth Engagement	-0.15	0.03	-0.24*	-0.07
E8: Parent, Family, Community	-0.03	-0.01	-0.17	-0.08
E9: Sustainability & Growth	0.02	0.02	-0.30**	-0.20*
E10: Outcomes & Evaluation	-0.15	-0.14	-0.21*	-0.15

*.Correlation is significant at $p<0.05$.
**.Correlation is significant at $p<0.01$.

Table Fifteen: Regression for 2015-2016 AQuA and early elementary math achievement

Predictor Variables	Model 1			Model 2		
	B	*SE B*	β	*B*	*SE B*	β
AQuA Overall	5.54	2.84	1.38	--	--	--
Time 1 math z score	-0.29	0.17	-0.31	-0.17	0.25	-0.18
Program attendance rate	0.20	0.95	0.04	0.31	1.07	0.06
School quality index score	-2.97	1.42	-1.46*	--	--	--
Element 1	--	--	--	--	--	--
Element 2	--	--	--	--	--	--
Element 3	--	--	--	-0.48	5.91	-0.07
Element 4	--	--	--	--	--	--
Element 5	--	--	--	-1.96	1.04	-0.56
Element 6	--	--	--	1.96	1.98	0.37
Element 7	--	--	--	--	--	--
Element 8	--	--	--	1.79	1.99	0.49
Element 9	--	--	--	--	--	--
Element 10	--	--	--	-0.59	2.65	-0.26
R^2		0.22			0.24	
F for change in R^2		2.17			1.24	

*Predictor is significant at p < .05.

Table Sixteen: Regression for 2015-2016 AQuA and early elementary literacy achievement

Predictor Variables	Model 1			Model 2		
	B	*SE B*	β	*B*	*SE B*	β
AQuA Overall	4.00	1.65	0.85*	--	--	--
Time 1 literacy z score	-0.43	0.13	-0.42**	-0.37	0.14	-0.36**
Program attendance rate	-0.30	0.62	-0.06	-0.54	0.79	-0.10
School quality index score	-2.08	0.85	-0.87*	5.87	4.29	2.45
Element 1	--	--	--	--	--	--
Element 2	--	--	--	2.02	1.16	0.47
Element 3	--	--	--	--	--	--
Element 4	--	--	--	--	--	--
Element 5	--	--	--	-3.78	2.84	-0.95
Element 6	--	--	--	4.64	2.11	0.91*
Element 7	--	--	--	--	--	--
Element 8	--	--	--	4.00	2.24	0.89
Element 9	--	--	--	-0.76	2.62	-0.18
Element 10	--	--	--	-9.09	5.40	3.28
R^2		0.34			0.40	
F for change in R^2		6.73**			3.50**	

*Significant at p < .05.
**Significant at p<.01.

Table Seventeen: Regression for 2015-2016 AQuA and upper elementary/secondary math achievement

Predictor Variables	Model 1			Model 2		
	B	*SE B*	β	*B*	*SE B*	β
AQuA Overall	-28.18	44.21	-0.07	--	--	--
Time 1 math scale score	-0.29	0.06	-0.50**	-0.31	0.07	-0.52**
Program attendance rate	15.75	39.99	0.04	44.62	42.56	0.10
School quality index score	15.35	15.79	0.10	23.29	24.51	0.15
Element 1	--	--	--	73.74	82.29	0.15
Element 2	--	--	--	139.78	149.76	0.42
Element 3	--	--	--	-427.11	400.84	-0.53
Element 4	--	--	--	11.73	188.99	0.05
Element 5	--	--	--	356.25	222.84	0.93
Element 6	--	--	--	205.81	254.05	0.41
Element 7	--	--	--	-125.68	182.14	-0.32
Element 8	--	--	--	-245.18	119.81	-0.10
Element 9	--	--	--	207.78	217.29	0.79
Element 10	--	--	--	-322.55	443.21	-1.51
R^2		0.26			0.42	
F for change in R^2		8.28**			4.56**	

**Significant at p<.01.

Table Eighteen: Regression for 2015-2016 AQuA and upper elementary/secondary reading achievement						
Predictor Variables	Model 1			Model 2		
	B	*SE B*	β	*B*	*SE B*	β
AQuA Overall	1.10	36.86	0.003	--	--	--
Time 1 reading scale score	-0.20	0.05	-0.42**	-0.22	0.06	-0.50**
Program attendance rate	-6.53	33.61	-0.02	-5.74	39.76	-0.02
School quality index score	1.92	13.46	0.02	-13.97	22.88	-0.11
Element 1	--	--	--	-19.81	74.77	-0.05
Element 2	--	--	--	142.29	134.69	0.52
Element 3	--	--	--	-267.99	374.10	-0.41
Element 4	--	--	--	120.91	176.52	0.67
Element 5	--	--	--	187.06	208.14	0.61
Element 6	--	--	--	79.98	237.24	0.19
Element 7	--	--	--	39.98	164.63	0.13
Element 8	--	--	--	25.41	110.84	0.07
Element 9	--	--	--	130.96	202.17	0.64
Element 10	--	--	--	-341.62	413.49	-2.01
R^2		0.17			0.22	
F for change in R^2		4.92**			1.86*	

*Significant at $p < .05$.
**Significant at $p<.01$.