

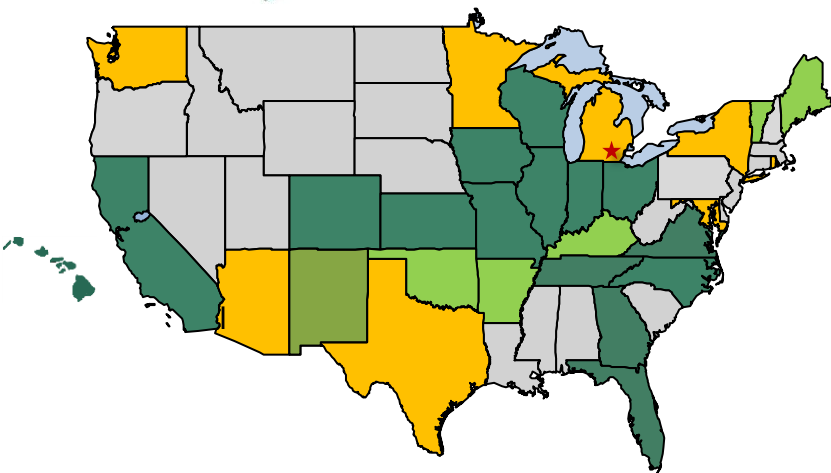
Does Quality Matter in Afterschool? Expanded Learning and Youth Outcomes

*Charles Smith, Weikart Center for Youth Program Quality
Neil Naftzger, American Institutes for Research*

April 23, 2014



Scale of Adoption (2013-2014)



Light Green- full-state

Dark Green- place-based

Gold- full-state + place-based

Policy
Setting

90 Networks/ Settings

Organization
Setting

>3650 Sites

>23,725 Staff

Estimate based on mean of 6.5
staff per site in YPQI Study Sample

Point of
Service
Setting

>310,250 Child & Youth

Estimate based on mean daily attendance of 85
youth per day in YPQI Study Sample



#Rb21NM



all youth ready for college work and life



Multi-Billion Dollar Questions about “More” and “Better” Learning

- Does participation in afterschool have a positive average effect on achievement? (More)
- Does the effect of participation on achievement depend on qualities of afterschool programs? (Better)

Two “Aging” Public Narratives on Effect of Afterschool Participation

- **2004 Mathematica Study**
 - No average effect on achievement
(neg effect on behavior)
- **2007 Durlak et al. Meta-Analyses**
 - Substantial achievement effect
depends on “qualities” (SAFE)

Primer on Effect Size

Helps evaluate the size of a difference.

Cohen's d :

$$d = \frac{\text{mean}_1 - \text{mean}_2}{SD}$$

where SD is the pooled standard deviation

Labels for Values of d

Value of d	Label
0.20	Small
0.50	Medium
0.80	Large
1.10	Very Large
1.40+	Extremely Large



Substantively Important Effect Sizes in Education

- The typical effect size associated with the school day's impact on achievement test scores ranges from .33 to .50
- Taking program dosage into consideration, would expect an effect size of around .05 related to academic achievement in afterschool programs

Average Effects: Summary of Findings from Five Statewide Evaluations by AIR

Mean Effects Sizes by Youth Outcome and Level of 21st CCLC Participation

Program Attendance Level	Outcomes Related to Academic Achievement					Outcomes Related to Behavior	
	State Assessment - Reading	State Assessment - Mathematics	Grade Promotion	Cumulative CPA	Credits Earned	School-Day Absences	Disciplinary Incidents
30-days or more	Mean ES = 0.011 (4 state reports) Range ES -0.054 to 0.041	Mean ES = 0.035 (5 state reports) Range ES 0.021 to 0.044	ES = 0.354 (1 state report)	No significant effect	No significant effect	Mean ES = -0.345 (3 state reports) Range ES -0.146 to -0.578	Mean ES = -0.308 (2 state reports) Range ES -0.058 to -0.557
60-days or more*	Mean ES = 0.032 (3 state reports) Range ES 0.030 to 0.033	Mean ES = 0.043 (6 state reports) Range ES 0.029 to 0.063	ES = 0.384 (1 state report)	ES = 0.195 (1 state report)	ES = 0.144 (1 state report)	Mean ES = -.487 (3 state reports) Range ES -0.159 to -0.664	Mean ES = -0.266 (3 state reports) Range ES -.096 to -0.583

*In one state, the higher attendance condition was defined as 70-days or more

Cells with lighter green shading outline **small** average effect sizes

Cells with darker green shading outline **moderate** average effect sizes

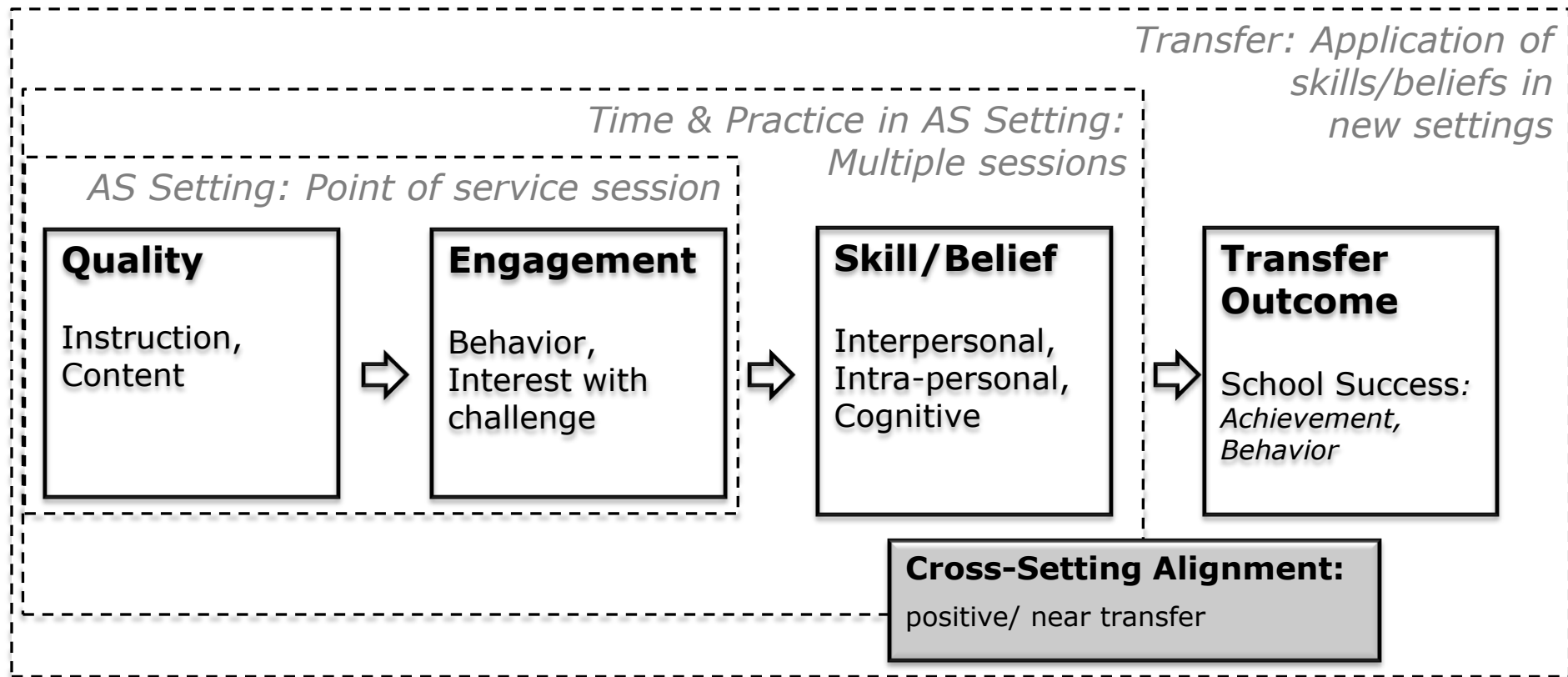
Commonly Used in Early
Warning Systems



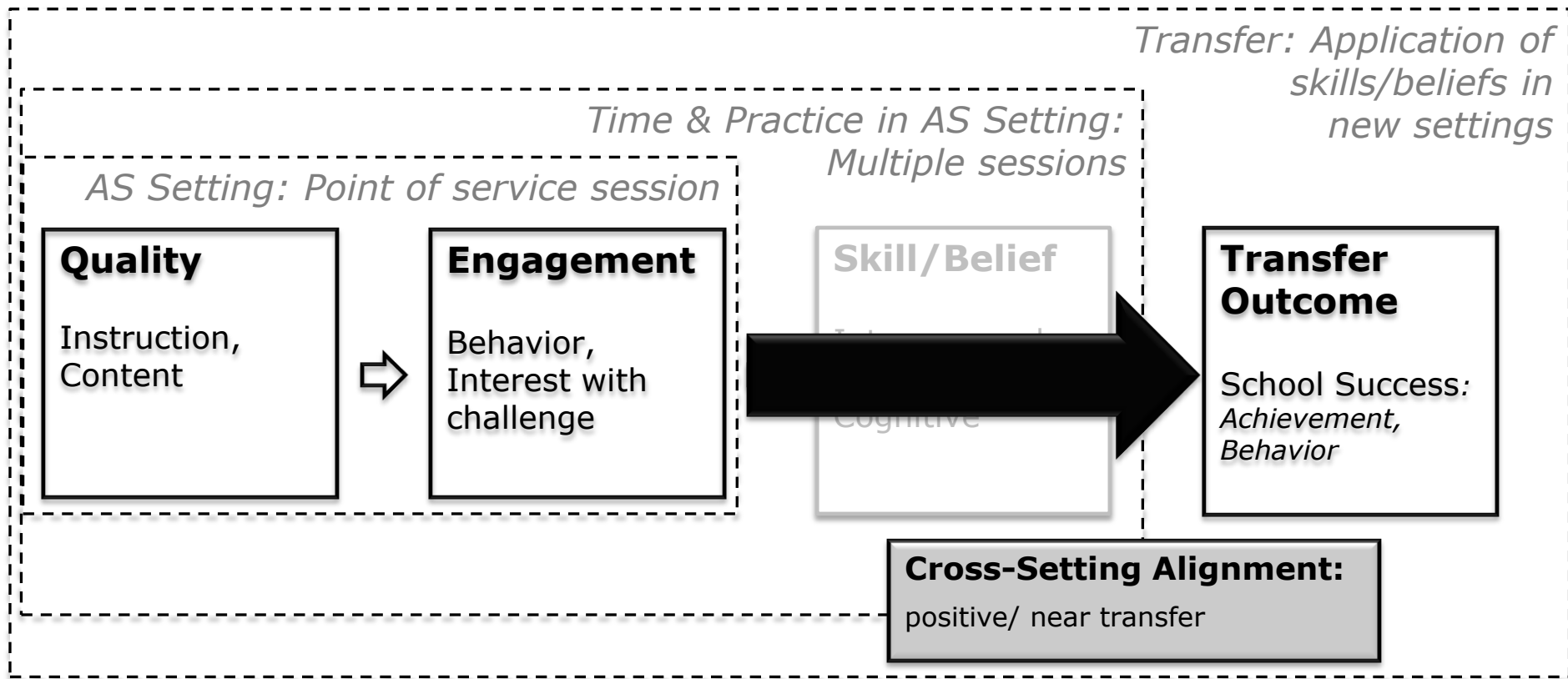
Questions?

How Effects Depend on Qualities?

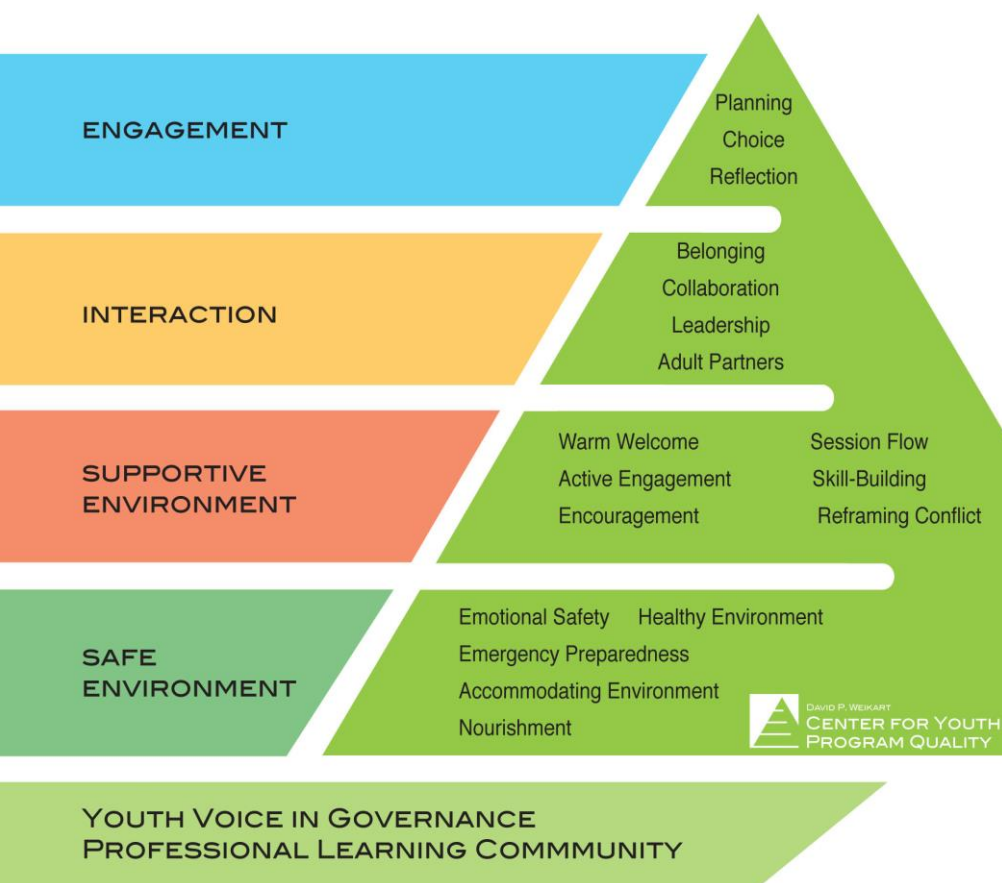
QuEST



Replicated Findings in Three Studies



Note on “Qualities”



Engagement:

- I was interested in what we did
- The activities were important to me
- I wished I was doing something else (Reversed)
- I tried to do things I have never done before
- I was challenged in a good way
- I really had to concentrate to complete the activities
- I was using my skills
- The activities were too easy (Reversed)



#Rb21NM



all youth ready for college work and life



Primary Research Question & Hypothesis

- What impact does participation in *higher quality* afterschool programs have on youth outcomes as compared to similar youth participating in *lower quality* afterschool programs?
- Primary Hypothesis - Youth who participate regularly in *higher quality* programs will demonstrate better functioning on a variety of youth outcomes as compared to similar youth attending *lower quality* programs.

Study Steps

- Assign Afterschool Programs Quality to *Higher* and *Lower Quality* Groups
- Create Meaningful Comparison Groups (Texas and Palm Beach)
- Conduct Outcome Analyses

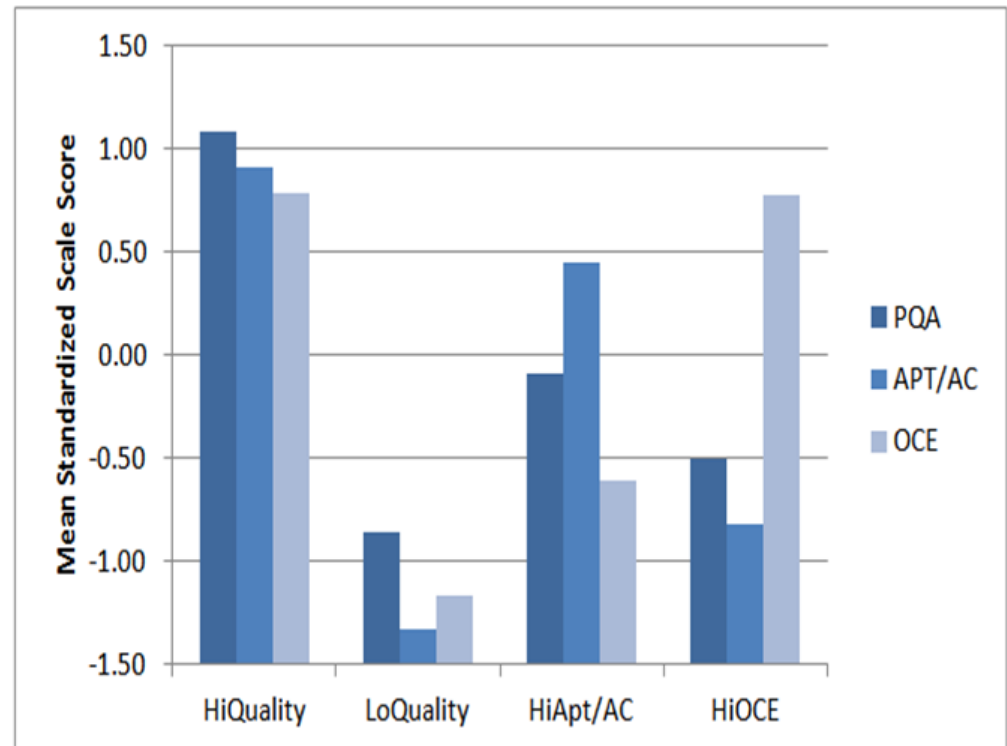
Quality Profile Creation

- Goal - Define quality profile types that were *different* from one another in substantive ways in terms of how youth engage in and benefit from afterschool programming
- Reliance on *Youth PQA* data to form profiles, specifically items from *supportive environment, interaction, and engagement*

Texas Quality Groups

- **Definition of quality:** Multi-variate profiles
 - Youth PQA total score
 - Academic Climate/Skill Building Opportunities
 - Observed youth engagement

Quality Profiles Based on Observation Data, Spring and Fall 2011



Source: $n=238$ activities observed at 80 programs

Study Designs

- One study design was *causal* (Palm Beach)
 - Significant, positive effects, if found as hypothesized, could be interpreted as participation in higher-quality programming *causing* a given outcome
- Two were *correlational* (Texas and Nashville)
 - *Cannot* definitively say that program quality caused a given outcome to happen

Youth Outcomes

- Outcomes examined included:
 - Afterschool program attendance
 - School day attendance/absences
 - School day disciplinary referrals
 - Grade promotion (lower probability of being retained in the same grade level)
 - Achievement in reading and mathematics (state assessments results and grades)

Results

- Three youth outcomes where our hypothesis was *confirmed* and *replicated* in at least two of the three studies:
 - Longer duration of attendance in afterschool programming (Texas and Palm Beach)
 - Fewer school-day disciplinary referrals (Texas and Nashville)
 - Enhanced likelihood of grade promotion (Texas and Palm Beach)



Results

- Results related to achievement were mixed:
 - *Lower quality* programs in Texas were found to have smaller effect sizes for *reading* state assessment scores
 - In Nashville, higher levels of program attendance combined with *higher quality* was related to greater improvement in *mathematics grades* during the span of the school year
 - In Palm Beach, enrollment in *higher quality* programs had a negative effect on *mathematics* state assessment scores relative to enrollment in lower quality programs

Replication: Summary of Findings

	AS Attend	School Attendance	School Behavior	Grade Promotion	Grades	Achievement
Texas 21st CCLC Evaluation¹ - <small>N=10,381 youth attending 40 centers; Quality based on observations and staff surveys</small>	Yes	No	Yes	Yes	NA	Yes (Literacy)
Palm Beach QIS Impact Study² – <small>N=1,332 youth attending 38 programs; Quality based on observations</small>	Yes	No	No	Yes	NA	No (Literacy)
Nashville Exploratory Youth Outcome Study¹ <small>N=539 youth attending 16 programs; Quality based on observations and youth surveys</small>	NA	Yes	Yes	NA	Yes	NA



#Rb21NM



all youth ready for college work and life



An Emerging Narrative

- Average Effects
 - Moderate and substantively important average effects on behavior, grade promotion, and attendance
 - Small average effects on achievement consistent with what would be expected given program dosage
- Quality Effects
 - Quality matters in terms of reducing negative behaviors and supporting grade promotion
 - Larger pattern of positive results
 - Average effects for reducing unexcused absences and disciplinary behaviors largest in states that have invested in quality improvement systems

Meta-Analysis

- Combine results from statewide evaluations conducted by AIR with *other robust evaluations* undertaken by states and local 21st CCLC grantees to support a comprehensive *meta-analysis* of program impacts

Meta-Analysis

- The average effect size on academic achievement in Durlak & Weissberg was .20 (achievement test scores) and .22 (grades) (for the programs exhibiting the SAFE characteristics) and the average effect size in Lauer et al. was XX.

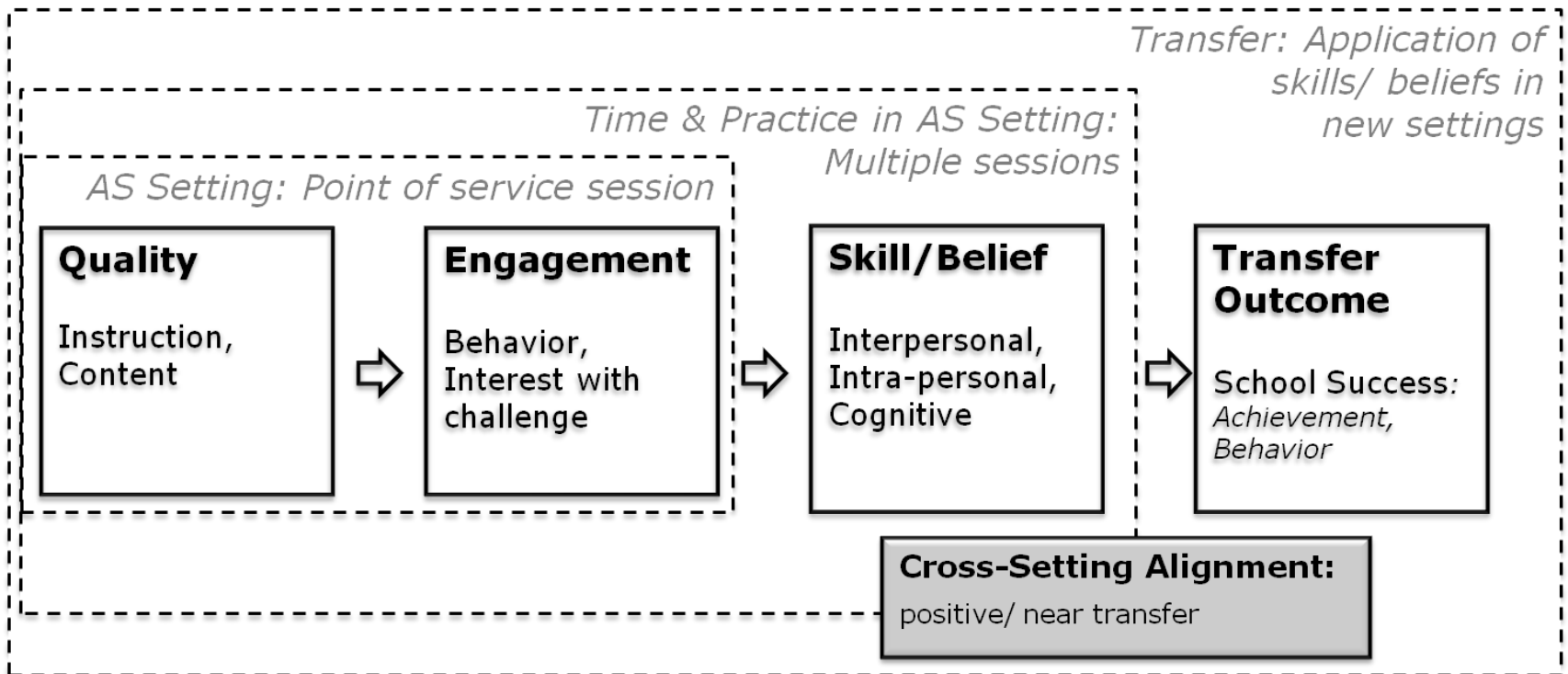
Durlak, J. A., & Weissberg, R. P. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology*, 45(3-4), 294-309.

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

Questions?



So... What are the skills that transfer and why do they matter for behavior and retention?



Question for Afterschool Field

- Is social and emotional learning the path from afterschool experience to improved school day outcomes?

Stay Connected

- Presentation materials will be posted at www.readyby21.org/nationalmeetingonline
- Tweet about your session! #Rb21NM

A Summary of Three Studies Exploring the Relationship between Afterschool Program Quality and Youth Outcomes

**Conference Paper Presented at the 2014 Ready by 21
National Meeting – Northern Kentucky**

April 2014

Neil Naftzger



AMERICAN INSTITUTES FOR RESEARCH®

1120 E. Diehl Road, Suite 200
Naperville, IL 60563-1486
630.649.6500 | Fax: 630.649.6700

www.air.org

Copyright © 2014 American Institutes for Research. All rights reserved.

Introduction

As schools and communities struggle to close the persistent achievement gap as well as meet the social and emotional needs of young people, a growing body of research has focused in recent years on the impact afterschool programs are having on the youth who attend them. Although there is a great deal of research pointing toward the benefit of afterschool, questions remain about which types of programs are most effective and how often young people need to attend them in order to see benefits.

With those questions in mind, the American Institutes for Research (AIR), in conjunction with the David P. Weikart Center for Youth Program Quality (Weikart Center), has taken some preliminary steps to explore the relationship between afterschool program attendance, program quality, and school-related outcomes. During the span of the past year, AIR has conducted three studies oriented at answering the following primary research question (Naftzger et al., 2013; Naftzger, Hallberg, & Tang, 2014; Naftzger, Devaney, & Foley, 2014):

- What impact does participation in *higher quality* afterschool programs have on youth outcomes as compared to similar youth participating in *lower quality* afterschool programs?

In each of the three studies, afterschool program quality was primarily defined by quality ratings produced using the Youth Program Quality Assessment (Youth PQA; Smith & Hohmann, 2005), an observation-based quality assessment tool developed and supported by the Weikart Center. The Youth PQA is made up of a series rubric-based items organized into four broad domains - (1) safety, (2) supportive environment, (3) interaction, and (4) engagement – and can be used to produce quality ratings for instructional best practices in afterschool programs.

Two of the studies were conducted by analyzing quality ratings, afterschool program attendance, and school-related youth outcome data associated with afterschool programs supported by afterschool intermediaries in Palm Beach County, Florida (Prime Time of Palm Beach County, Inc.) and Nashville, Tennessee (Nashville After Zone Alliance or NAZA). Both of these intermediaries have a core mission of helping afterschool programs in their community progress to higher levels of program quality by using the Youth PQA to help programs understand what constitutes afterschool program quality and how well they measure up to these criteria. This information is then used to support the development of afterschool staff to better design and deliver programming in a fashion consistent with the quality criteria articulated in the Youth PQA.

The final study was conducted as part of the statewide evaluation of the 21st Century Community Learning Centers (21st CCLC) program in Texas (branded as the Texas Afterschool Centers on Education or ACE by the Texas Education Agency). In this instance, the Youth PQA

was combined with other observation measures of setting-level quality to assess the level of quality at a random sample of centers funded by Texas ACE.

The primary purpose of the paper is to: (1) summarize the analyses undertaken to answer the primary research question underpinning each study, (2) summarize key findings, and (3) discuss what these results mean for future efforts oriented at understanding the relationship between afterschool program quality and youth outcomes.

What steps were taken to define higher and lower quality program?

In each of the three studies, a critical first step was to classify afterschool programs into higher and lower quality groups based on quality ratings collected in each system. In both Nashville and Palm Beach, extant quality data collected by external raters employed by each intermediary were used to support the classification of programs into higher and lower quality groups. Quality ratings data was produced by each intermediary organization as part of quality improvement processes that provide feedback to afterschool staff on how well they were doing in implementing quality programming and where there were opportunities for improvement. In addition, in Nashville, self-assessment ratings provided by the programs themselves were also available and utilized in the formulation of quality groups.

In Texas, quality rating data was collected directly by members of the evaluation team in a random sample of 40 centers funded by the program, with a subset of observations conducted by rater pairs to assess inter-rater reliability and control for rater bias. In addition to the Youth PQA, two additional observation instruments were scored by observers - the (1) Observation of Child Engagement (OCE) and (2) portions of the Afterschool Practices Observation Tool (APT-O) related to supports provided by staff and tasks undertaken by participating youth to practice specific academic skills.

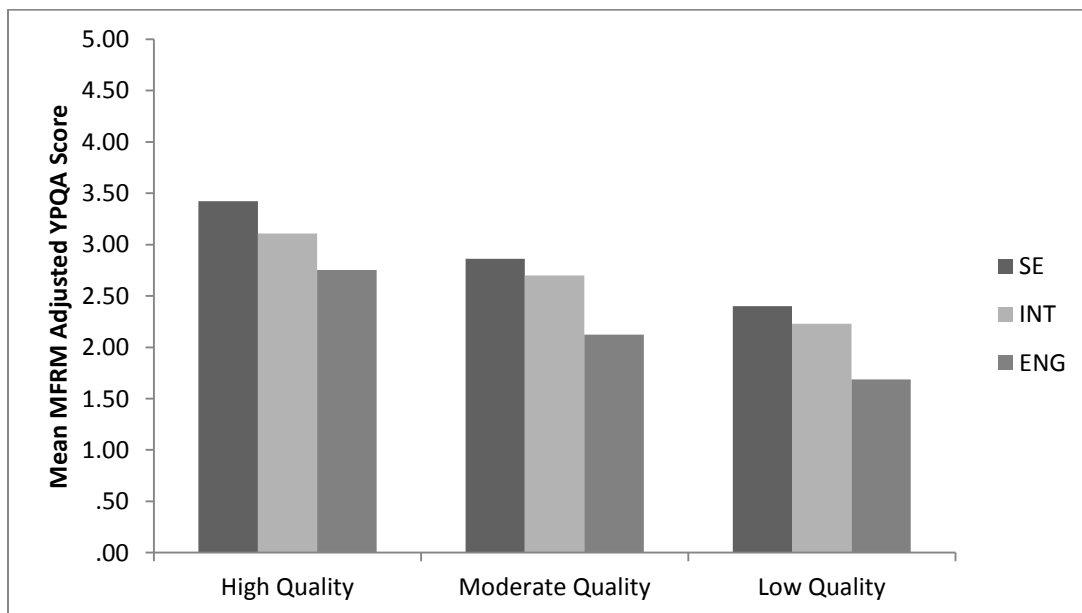
For all studies, scores for the Youth PQA were obtained by running a series of Rasch-based analyses. For each study, this allowed the research team to identify and control for a variety of elements which may have served to bias Youth PQA ratings, including:

1. *Rater bias* - Some raters scoring the Youth PQA are systematically more lenient or severe in their ratings. In Texas, steps were taken to identify and control for rater bias. This was possible given that a subset of observations had two paired raters scoring the Youth PQA.
2. *Bias introduced by scores obtained through self-assessment* - Scores derived from self-assessment were found to be higher on some domains of the Youth PQA than those obtained from external observers. Steps were taken to control for this type of bias in Nashville since self-assessment data was used in that particular study to classify programs into higher and lower groupings.

3. *Bias introduced by the type of activity observed* – Some afterschool activities simply score better on the Youth PQA than others, and not controlling for these differences can have significant impacts on how a given program is rated. Significant differences have been consistently found to exist between enrichment activities which score systematically higher on the tool and recreational and overt academic activities which score systemically lower (Smith, Peck, Denault, Blazeovski, & Akiva, 2010; Naftzger et al., 2014; Naftzger, Hallberg, & Tang, 2014). If a program’s offerings are not carefully sampled, especially in a program which offers many different types of activities, then a program-level rating could be biased.

Once Youth PQA scores were obtained, steps were then taken to classify programs into higher, moderate, and lower quality groupings using hierarchical cluster analysis (see Figure 1 as an example of the clusters created in Palm Beach). Rasch-derived scores on the *supportive environment*, *interaction*, and *engagement* domains of the Youth PQA were included in these analyses. Scores from the *safety* domain were not included given little variation in these scores across programs. Additional steps were then taken to refine the programs assigned to the higher and lower quality groups in order to ensure there was a significant difference in the level of performance between the two groups, resulting in some lower performing centers being removed from the higher quality group and higher performing programs removed from the lower quality group. The goal was to maximize the contrast between higher and lower quality programs.

Figure 1. An Example of Youth PQA Cluster Analysis Results



How were the study designs different?

While a relatively consistent approach was used to create higher and lower quality groups, there were some important differences about the design of each study that have ramifications for the robustness of each study's results.

1. *Palm Beach County* – The strongest study design was employed in the Palm Beach County study. In this case, a propensity score stratification approach was used which allowed the research team to more closely estimate the *causal effect* of attending a higher quality program on youth outcomes relative to a comparison group made up youth attending lower quality programs. Like random assignment, this approach better controlled for selection bias which may have differentiated youth that chose to attend a higher as opposed to a lower quality program. As a result, significant, positive effects, if found as hypothesized, could be interpreted as participation in higher-quality programming causing a given outcome.
2. *Texas* – A slightly less robust design was used in Texas. Here, propensity score matching was used to create effect sizes for school-day outcomes at each afterschool program by comparing youth attending that center with non-participating youth attending the same school during the day. Then, multiple regression analyses were run to assess if participation in a higher quality program was significantly related to higher effect sizes on the outcomes examined. In this sense, these analyses were correlational, as opposed to causal.
3. *Nashville* – The least robust design was used in conducting the Nashville study. Here, factorial ANOVAs were run to explore the direct effect of quality on youth outcomes, the direct effect of higher program attendance on youth outcomes, and how quality and attendance interacted to produce desirable effects. Fewer efforts were taken to control for pre-existing differences between youth attending higher and lower quality programs. Like Texas, these analyses were correlational in nature.

What were the results?

Results from each of the three studies are summarized in Table 1. Across the three studies, the following domain of youth outcomes was examined:

1. Afterschool program attendance
2. School day attendance/absences
3. School day disciplinary referrals
4. Grade promotion (lower probability of being retained in the same grade level)
5. State assessment results in reading and mathematics

As mentioned previously, the primary hypothesis underpinning these studies was that participation in higher quality programs would be more likely to be related to desirable youth outcomes. There were three youth outcomes where a positive relationship was found with

enrollment in higher quality programs and each of these quality-outcome relationships were replicated in at least two of the three studies:

1. Longer duration of attendance in afterschool programming (Texas and Palm Beach)
2. Fewer school-day disciplinary referrals (Texas and Nashville)
3. Enhanced likelihood of grade promotion (Texas and Palm Beach)

Results related to state assessment scores in reading and mathematics were mixed. Lower quality programs in Texas were found to have smaller effect sizes for reading state assessment scores (consistent with what would be hypothesized). However, the opposite was found to be true in Palm Beach where enrollment in higher quality programs had a negative effect on mathematics scores relative to enrollment in lower quality programs, and no discernable effect on literacy. It is worth mentioning that school-based programs were heavily represented in the lower quality group, while programs in the higher quality group were more evenly split between community-based and school-based programs. While speculative, it may be the case that school-based programs facilitated the alignment of afterschool programming with school-day content in a way that supported the achievement of desirable academic outcomes. Since lower quality programs were overwhelmingly school-based programs, this may have resulted in the finding related to mathematics achievement. This Palm Beach finding certainly requires additional exploration.

However, findings from the Nashville study demonstrated a positive relationship between higher quality and mathematics achievement. In this case, higher levels of program attendance combined with higher quality was related to greater improvement in mathematics grades during the span of the school year. An effort is currently underway in Palm Beach to gain access to grades in order to further explore the relationship between participation in higher quality programs and an improvement in grades.

While these results are promising, there are several limitations the reader should be aware of when drawing conclusions from these data. First, in the case of Texas and Nashville, these analyses were only correlational in nature and the research team only partially explored demographic differences between youth enrolled in higher or lower quality programs or who attended programs more frequently. It is possible that if significant differences in outcomes are found to exist between youth in higher and lower quality programs, the differences may have more to do with the demographic differences associated with youth enrolled in each type of program than the level of quality. In other words, while the findings described in this report demonstrate a relationship between program quality and outcomes, we cannot definitively say that program quality caused a given outcome to happen.

While the design employed in Palm Beach was more rigorous, the study overall was underpowered given the small n sizes involved, which may have impeded the ability of the research team to detect meaningful effects.

Table 1. Summary of the Studies Completed by AIR Examining the Relationship Between Program Quality and Youth Outcomes

	AS Attendance	School Attendance	School Behavior	Grade Promotion	Grades	State Assessment Results
Texas 21st CCLC Evaluation¹ - N=10,381 youth attending 40 centers Quality based on observations and staff surveys	No relationship found between quality and hours of participation Youth in higher quality programs attended programming for a longer duration	No significant relationship found	Higher quality programs were found to have higher effect sizes in terms of lower disciplinary referrals	Higher quality programs were found to have higher effect sizes in terms of supporting grade promotion	Not examined	Lower quality programs were found to have lower effects sizes in terms of reading state assessment results
Palm Beach QIS Impact Study² – N=1,332 youth attending 38 programs Quality based on observations	No relationship between quality and days of participation Higher quality programs were more likely to retain youth in programming across multiple years	No significant relationship found	No significant relationship found	Participation in higher quality programming reduced the likelihood that a student would be retained in the same grade for the next school year	Not examined	Some analyses demonstrated a negative relationship between enrollment in higher quality programs and state assessment scores in mathematics
Nashville Exploratory Youth Outcome Study¹ N=539 youth attending 16 programs Quality based on observations and youth surveys	Not examined	No significant relationship found between enrollment in higher quality programs and the percentage of school days attended Higher levels of program attendance combined with higher quality (defined by observation and youth survey data) was related to fewer school-day tardies	Youth enrolled in higher quality programs were found to have fewer disciplinary referrals	Not examined	Higher levels of program attendance combined with higher quality was related to greater improvement in mathematics grades during the span of the school year	Not Examined

¹ Analyses connecting program quality to youth outcomes were correlational. This is worthy of note since there is some evidence that youth that attend lower quality programs are often different both demographically and on pre-treatment youth outcomes than youth attending higher quality programs. These possible differences are not adequately controlled for in some of the correlational models included here, particularly the Nashville study, so while a relationship may exist between quality and youth outcomes, we cannot rule out that this is an artifact of pre-existing differences in the youth served in higher and lower quality programs. The Texas analyses related to education-related outcomes are substantially more robust in this regard since propensity score analyses were first used to match youth based on pre-treatment characteristics and then correlational models were run to look for differences in effect sizes by quality groupings.

² Analyses connecting program quality to youth outcomes were causal, with the exception of those related to program attendance which were correlational

Recommendations

Based on the results demonstrated across the three studies detailed in this paper, there are two primary recommendations that would seem to flow from this pattern of results.

1. *There is a need for further study.* While the results from the three studies paint a promising picture of the relationship between higher, setting-level program quality and youth outcomes, there is still a need to conduct more robust, adequately-powered research studies to better quantify the effect of quality on host of youth outcomes across time; how these effects interact with other aspects of program design and delivery; and how these effects vary for different age-levels. This work outlined here should be considered a starting point for future analyses, not an endpoint.
2. *Process quality matters and warrants investment on the part of state and local systems.* However, despite the need for further research, the results outlined here are promising enough that state and local systems should consider using the scarce resources available to them to fund the development and implementation of quality improvement systems predicated on tools like the Youth PQA as a strategy for enhancing the likelihood of achieving desired youth outcomes, particularly those outcomes related to positive school-related behaviors.

References

- Naftzger, N.; Devaney, E.; & Foley, K. (2014) *Summary of Analyses Relate to NAZA Program Outcomes*. Naperville, IL: American Institutes for Research
- Naftzger, N., Hallberg, K. & Tang, Y. (2014) *Exploring the relationships between afterschool program quality and youth outcomes: Summary of findings from the Palm Beach County quality improvement system*. Naperville, IL: American Institutes for Research.
- Naftzger, N., Nistler, M., Manzeske, D., Swanlund, A., Shields, J., Rapaport, A., Smith, C., Gersh, A., & Sugar, S. (2013). *Texas 21st Century Community Learning Centers: Year two evaluation report*. Naperville, IL: American Institutes for Research.
- Naftzger, N., Vinson, M., Liu, F., Zhu, B., & Foley, K. (2014). *Washington 21st Century Community Learning Centers Program Evaluation: Year 2*. Naperville, IL: American Institutes for Research.
- Smith, C., & Hohmann, C. (2005). Full findings from the Youth PQA validation study *High/Scope Youth PQA Technical Report*. Ypsilanti, MI: High/Scope Educational Research Foundation.
- Smith, C., Peck, S. J., Denault, A., Blazeovski, J., & Akiva, T. (2010). Quality at the point of service: Profiles of practice in afterschool settings. *American Journal of Community Psychology*, 45, 358-369.