



**OUT-OF-SCHOOL TIME (OST)
OBSERVATION INSTRUMENT**

Report of the Validation Study

Ellen M. Pechman
Monica B. Mielke
Christina A. Russell
Richard N. White
North Cooc

Policy Studies Associates, Inc.

February 2008

Prepared with the support of
The Charles Stewart Mott Foundation

OUT-OF-SCHOOL TIME (OST) OBSERVATION INSTRUMENT

Report of the Validation Study

Policy Studies Associates, Inc. (PSA) developed the Out-of-School Time (OST) Observation Instrument to support evaluations of several OST initiatives conducted during the past eight years. PSA designed the instrument to collect consistent, objective observation-based information on OST programs, whether within schools or in other settings. Two sets of assumptions about high-quality OST programs guided the instrument's development: (1) successful programs employ a qualified staff, offer youth both academic and interpersonal skill-building, and operate with adequate space and resources; and (2) successful programs offer varied instructional activities that are content-based and mastery-focused and that encourage positive youth-to-youth and youth-to-staff relationships (Birmingham, Pechman, Russell, & Mielke, 2005; Durlak & Weissberg, 2007; Reisner, White, Russell, & Birmingham, 2004; Yohalem & Wilson-Ahlstrom, 2007).

This paper describes the instrument's purpose and how it has been used, and it reports the results of reliability and validity analyses from recent PSA studies, including a follow-up study of The After-School Corporation (TASC) programs (Birmingham et al., 2005) and evaluations of the New York City Department of Youth and Community Development's Out-of-School Time Program for Youth initiative (Russell, Mielke, & Reisner, 2008; Russell, Reisner, Pearson, Afolabi, Miller, & Mielke., 2006) and New Jersey After 3 (Kim, Miller, Reisner, & Walking Eagle, 2006; Walking Eagle, Miller, Reisner, Johnson, Mielke, Edwards, & Farber, 2007).

Overview of the OST Observation Instrument

Purposes

The OST Observation Instrument is a practical data-collection tool that can be used in varied OST contexts, including programs in schools or other settings, and with youth participants who are in kindergarten through twelfth grade. It provides researchers and practitioners with a theory-based resource for recording consistent and objective data about the quality of OST programs. With this instrument, study teams can unobtrusively observe youth and staff in OST programs, concentrating on the strategies that staff employ and the instructional and interpersonal interactions that occur among youth participants and between participants and staff.

Collecting data while activities occur, but without intruding on youth experiences, enables observers to consistently document activities and interactions. The instrument was developed primarily for research purposes and is not designed to assign overall quality scores to specific programs or individual staff members. However, by using this instrument to document important information about activity variety, teaching and learning practices, and staff and participant interactions, evaluators and program leaders can obtain objective information that can support reflective practice, professional development, monitoring, and evaluation.

PSA researchers and their collaborators have used a version of the OST Observation Instrument successfully and reliably in several studies. It has also been shared by other researchers (Penuel & McGhee, forthcoming) and with program teams that have used it to implement their own internal monitoring and evaluation (Center for After-School Excellence, 2007). The experiences of these research teams demonstrate that the instrument is easy to use and that it can be adapted to varied programs settings.

Theoretical Framework

The OST Observation Instrument is grounded in assumptions about the characteristics of high-quality after-school programs that have been demonstrated by a growing body of research (Eccles & Gootman, 2002; Marzke, Pechman, Reisner, Vandell, Pierce, & Brown, 2002; McLaughlin, 2000; Miller, 2003; Mott Foundation Committee on After-School Research and Practice, 2005; Pittman, Irby, & Ferber, 2001; Vandell, Reisner, Pierce, Brown, Lee, & Bolt, 2006). Following McLaughlin (2000), this body of research finds that good after-school programs are not “happenstance.” Positive outcomes occur when adults deliberately create opportunities where activity content and instructional processes are both knowledge- and youth-centered and when adults use both structured and unstructured teaching strategies to promote learning and mastery (Bransford, Brown, & Cocking, 1999; Durlak & Weissberg, 2007; Grossman, Campbell, & Raley, 2007). To reflect these principles, the OST Observation Instrument measures activity content and structure, the quality of interpersonal relationships, and the degree to which activities focus on skill development and mastery, all factors that encourage positive youth outcomes.

The instrument captures data on three major structural components of after-school programs: (1) the types of activities that engage youth; (2) the structures that facilitate activities (e.g., spaces used, staffing, number of participants and their grade levels, adequacy of adult supervision, and materials); and (3) the quality of interactions among participating youth and the adults who work with them. For each activity, observers first record information about the number and grade levels of participants, the type and number of staff, and the activities and primary skills targeted. They then rate the quality of interactions among youth and between youth and adults in five domains: youth relationship-building; youth participation; relationship-building among staff and youth; instructional strategies; and activity content and structure.

In 2007, the developers of the OST Observation Instrument added a supplementary insert for documenting research-based academic features of reading/language arts, mathematics, and technology activities. This dimension of the instrument responds to the increasing focus of OST programs on enrichment opportunities that link to school-day academic learning in literacy and mathematics (Penuel & McGhee, forthcoming).

The OST Observation Instrument is a flexible resource that can help OST programs determine if their program is characterized by the youth development, academic, and enrichment qualities that promote positive participant outcomes.

Procedures for Using the OST Observation Instrument

The OST Observation Instrument is available for review and downloading at: www.policystudies.com.

The instrument's cover sheet records program activity types, documenting specific activities (e.g., tutoring, visual arts, music, sports, community service); spaces in which activities take place (e.g., classroom, gym, library, cafeteria, auditorium, hallway, playground); the primary learning skill targeted (e.g., artistic, physical, literacy, numeracy, interpersonal); the number and education level of staff leading in the activity; and the number, gender, and grade level of participants.

As seen in Exhibit 1, the instrument captures information about five youth development domains and allows observers to rate the quality of indicators of youth interaction, quality of youth participation, relationships among youth and staff, staff instructional strategies, and activity content and structure. Evidence that activities employ adequate supervision, space, and materials is also recorded in the section called "environmental context." Last, the observation record sheet includes basic information about the observer, program, date, and time of the record.

Rating observed indicators. Observers rate each indicator on a scale from 1 to 7, in which 1 means that the indicator was not evident during the observation period and 7 means that the indicator was highly evident and consistent (see Exhibit 2). A score of 5 means either that the exemplar was evident or implicit, indicating that the desired behavior was somewhat present but not actively initiated or emphasized. For example, under youth relationship-building, if youth relaxed together and enjoyed one another's company but the activity did not involve a high level of socializing, the rating for "youth are friendly with each other" would be 5 (evident or implicit). Likewise, under staff-directed relationships, if staff promoted the participation of all youth in an inclusive manner but there was no need to re-engage an isolated child or group because every child was functioning well and appeared included, the rating for "promote participation of all" would be a 5 (evident or implicit). Indicators receive a rating of 6 or 7 if the behavior within the indicator was highly evident, consistent, and constructively promoted learning and engagement.

The conceptual design of the OST Observation Instrument assumes that, within domains, some indicators will occur routinely in a high-quality program, such as youth being friendly to others, staff using positive behavior management, and staff verbally recognizing youth efforts and accomplishments. Other indicators, although important in ensuring high-quality learning and developmental opportunities over time, might occur less frequently and only in appropriate contexts. Examples of these are giving youth meaningful choices, providing leadership responsibilities and roles, encouraging youth to work together or collaborate in teams, and requiring activities that would promote analytic thinking. Not every activity necessarily provides these kinds of opportunities; thus, domain scores will vary, with the expectation that most of these indicators will occur in programs over time and, in high-quality programs, will be rated between 5 and 7.

Exhibit 1

OST Observation Instrument: Domains and Indicators

Domain: Youth Relationship Building

Youth ...

- Are friendly and relaxed with one another
- Respect one another
- Show positive affect to staff
- Assist one another
- Are collaborative

Domain: Youth Participation

Indicators: Youth...

- Are on-task
- Listen actively and attentively to peers and staff
- Contribute opinions, ideas, and/or concerns to discussions
- Have opportunities to make meaningful choices
- Take leadership responsibility/roles

Domain: Staff Relationship Building

Indicators: Staff...

- Use positive behavior management techniques
- Promote the participation of all
- Show positive affect toward youth
- Actively listen to and/or observe youth
- Encourage youth to share ideas, opinions, and concerns about the content of the activity
- Engage personally with youth
- Guide peer interactions

Domain: Staff Instructional Strategies

Indicators: Staff...

- Communicate goals, purpose, expectations
- Verbally recognize youth's efforts and accomplishments
- Assist youth without taking control
- Ask youth to expand on their answers and ideas
- Challenge youth to move beyond their current level of competency
- Employ varied teaching strategies
- Plan for/ask youth to work together

Domain: Activity Content and Structure

Indicators: Activity...

- Is well organized
 - Challenges students intellectually, creatively, developmentally, and/or physically
 - Involves the practice/a progression of skills
 - Requires analytic thinking
-

Exhibit 2
Indicator Rating Scale for OST Observation Instrument

----1----	----2----	----3----	----4----	----5----	----6----	----7----
Exemplar is not evident		Exemplar is rarely evident		Exemplar is evident or implicit		Exemplar is highly evident and consistent

Recording observation data. Observers describe the content and context of what is occurring in the OST programs using checklists, quality scales, and written observations. The cover sheet, which describes the activity, is a checklist that captures descriptive information about the observed activity, including the content of the activity, skills targeted, type of space in which the activity occurs, participants’ grade level(s), types of staff, the number of total participants (by gender), and how they are organized (by age, grade, interest, or including all attendees). Observers track additional descriptive information about the content of the literacy, mathematics, and technology activities in the “OST Academic and Technology Features” section. By recording literacy, mathematics, and technology features as “present” or “not present,” observers catalog the details of activities’ academic features. At the end of an observation, researchers summarize the adequacy of the space and resources available to implement the activity in three yes/no questions in the Environmental Context section following the Domain Item Ratings section. The instrument also provides inserts for recording brief written notes about the observed quality features.

Training. Prior to conducting observations in the field, the observation teams in each study participate in intensive internal training to ensure that raters interpret domain indicators and activities similarly. During this training, team members observe and rate videotapes of after-school activity segments and compare their interpretations and ratings. Through discussions, they arrive at a common understanding of the theoretical framework and terminology. Observer pairs then visit OST program sites, and following each day’s observation, team members debrief their observations to clarify any interpretive differences. After several observer teams collect observational data in the field, observers meet again to clarify and align their interpretation of observation indicators and terminology to ensure continuing inter-rater reliability.

Coordinated instruments. PSA has used the OST Observation Instrument with a series of participant, staff, program coordinator, and executive director surveys and with guides for conducting interviews with participants, staff, and program coordinators. The surveys, interview guides, and observations have been used to collect data across several studies. Only the youth participant surveys were used in validating the OST Observation Instrument.

Technical Properties of the OST Observation Instrument

In establishing the psychometric properties of the OST Observation Instrument, the research team followed the technical standards for the development of psychometric instruments described in Yohalem and Wilson (2007). Exhibit 3 summarizes these elements of validity and their application to the OST instrument.

Exhibit 3 Technical Procedures Used to Validate the OST Observation Instrument¹

Element	Technical Definition	Procedures Used
Construct Validity	The degree to which an instrument accurately measures the construct of concern	The instrument's organizing constructs were grounded in research on OST programs that relate to positive developmental and behavioral outcomes for youth (Durlak & Weisberg, 2007; Eccles & Gootman, 2002; Marzke et al., 2002; Vandell et al., 2006).
Score Distributions	The dispersion of scores from multiple observations for a specific item or scale	Distributions of scale scores were examined. The scores varied in skewedness and overall distribution, and demonstrated neither floor nor ceiling effects.
Inter-rater Reliability	Whether assessments by different trained raters agree when observing the same program at the same time	Inter-rater reliability analyses using both Pearson correlations and intra-class correlations were conducted to verify overall- and domain-based inter-rater agreement.
Internal Consistency	The cohesiveness of items forming an instrument's scales	Cronbach's Alpha measured scale reliability, underscoring the conceptual coherence of the defined domains.
Concurrent Validity	When results from an instrument compare favorably with those from a similar measure (preferably one with demonstrated validity strengths)	Participant survey scales were compared with observation scales to confirm that the observation instrument was measuring similar constructs. The scales were compared using rank order correlations with Spearman's Rho statistic.
Predictive Validity	When an instrument successfully predicts related outcomes	Outcome information is not yet available to the study team, so analyses of predictive validity were not conducted for this phase of validation. Future analyses of the relationship between program quality and program enrollment, attendance, and youth behavior and academic outcomes are planned.
Validity of Scale Structure	When individual scales adequately measure the concepts they claim to measure	The reliability of the scale structure was measured using Cronbach's Alpha. Several scale structures were demonstrated as statistically reliable and strong. A factor analysis was conducted that demonstrated the consistency of the OST scales with the SAFE structure (Sequenced, Active, Focused, and Explicit, from Durlak & Weisberg, 2007).

¹ Adapted from Yohalem & Wilson-Ahlstrom, 2007, p. 16.

Reliability and Internal Consistency

Statistical evidence collected from three PSA studies demonstrates the strength of the OST Observation Instrument. Data reported here reflect inter-rater reliability and internal consistency analyses conducted for the TASC study of features of high-performing programs (Birmingham et al., 2005), the New York City OST initiative evaluation, Year 1 and Year 2 (Russell et al., 2006; Russell et al., 2008), and the New Jersey After 3 evaluation, Year 1 and Year 2 (Kim et al., 2006; Walking Eagle et al., 2007). Several approaches to measuring reliability were used. Pearson correlations established ratings of co-observers; intra-class correlations established the correlations of ratings across sites and observers. Cronbach's alpha measured the degree to which groups of items assessed the same underlying construct. Across these analyses, the OST Observation Instrument reached high levels of inter-rater agreement and the scales within the instrument demonstrated strong reliability.

Inter-rater Reliability

Inter-rater reliability analyses confirm that the observation instrument scores are consistent irrespective of who is collecting the information. These reliability analyses were conducted at both the instrument and indicator levels, using Pearson correlation coefficients and, for data collected in 2007, intra-class correlation coefficients. Correlations were calculated for rater pairs and across all raters. Results indicated consistent, high inter-rater reliability.

The TASC study (Birmingham et al., 2005) was the first to report reliability analyses for the OST Observation Instrument as part of PSA's after-school research. The study involved co-observations of 31 activities in 10 after-school programs, and achieved an overall inter-rater reliability of 0.83. Researchers studying the New York City OST initiative in 2005-06 co-observed 40 activities and also achieved an inter-rater reliability of 0.83. Similarly, for Year 1 of the New Jersey After 3 evaluation, researchers co-observed 30 activities with an inter-rater reliability of 0.88. Exhibit 4 summarizes the inter-rater reliability analyses across the three studies. In Year 2 of the New York City OST and New Jersey After 3 evaluations, researchers calculated additional, more rigorous intra-class correlations that yielded a similarly respectable level of reliability between raters. (Intra-class correlation coefficients are shown in parenthesis in Exhibit 4.)

Exhibit 4
Observation Inter-rater Reliability Coefficients by Study and Domain

Measure	TASC 2005	DYCD –Y1 2006	NJA3 – Y1 2006	DYCD –Y2 2008	NJA3 – Y2 2007
Number of programs observed	10	15	10	15	10
Number of activities with co-observations	62	40	30	33	19
Overall reliability coefficients	0.83	0.83	0.88	0.73 (0.72)	0.76 (0.73)
Youth relationship-building	0.86*	0.84	0.86	0.83 (0.79)	0.75 (0.75)
Youth participation	*	0.89	0.90	0.80 (0.81)	0.87 (0.86)
Staff relationship-building	0.87	0.87	0.91	0.79 (0.64)	0.82 (0.78)
Staff instructional strategies	0.77	0.78	0.86	0.63 (0.58)	0.64 (0.52)
<i>Note:</i> Coefficients are Pearson's; parentheses provide intra-class correlations, which were calculated for studies conducted in 2007. *In TASC 2005, elements from the youth relationship-building and youth participation domains were combined.					

Internal Consistency

Internal consistency analyses demonstrate the reliability of items within scales, indicating the extent to which the scales capture different aspects of a larger dimension. Driven by slightly different evaluation questions and change theories, the studies used the five domains and indicators in the OST instrument to create separate scales in 2005, 2006, and 2007. These scales, listed in Exhibits 5 to 8, represent an evolution of thought about how best to analyze data gathered with the OST Observation Instrument. They also present future users of the instrument with various possibilities for analyzing their observation data.

Because the studies created different scales in some years, scale reliability was analyzed separately for each of the studies. To compute all of the scales, evaluators averaged the items within each scale, yielding possible scale scores from one to seven. All scales showed strong reliability, with Cronbach's Alpha ranging from 0.73 to 0.91.

The TASC (2005) study combined the OST instrument's five domains and created four scales based on 173 independent observations in the 10 programs studied (Exhibit 5).

Exhibit 5
Internal Consistency of Observation Scales for TASC Follow-Up Study (2005)

Observation Scale	Indicators	Statistics
Youth Relationship-Building and Participation	<ul style="list-style-type: none"> ■ Youth are friendly and relaxed with one another ■ Youth respect one another ■ Youth show positive affect to staff ■ Youth are on-task ■ Youth listen actively and attentively to peers and staff ■ Youth have opportunities to make meaningful choices ■ Youth take leadership responsibility/roles 	Alpha: 0.73 Mean: 4.49 SD: 0.85
Staff-Youth Relationships	<ul style="list-style-type: none"> ■ Staff use positive behavior management techniques ■ Staff promote participation of all youth ■ Staff show positive affect toward youth ■ Staff attentively listen to and/or observe youth ■ Staff encourage youth to share their ideas, opinions, and concerns about the content of the activity ■ Staff engage personally with youth 	Alpha: 0.79 Mean: 4.58 SD: 1.40
Skill Building and Mastery	<ul style="list-style-type: none"> ■ Staff communicate goals, purposes, expectations ■ Staff verbally recognize youth's efforts and accomplishments ■ Staff assist youth without taking control ■ Staff ask youth to expand on their answers and ideas ■ Staff challenge youth to move beyond their current level of competency ■ Staff plan for/ask youth to work together ■ Staff employ varied teaching strategies 	Alpha: 0.83 Mean: 3.77 SD: 1.40
Activity Content and Structure	<ul style="list-style-type: none"> ■ The activity is well organized ■ The activity involves the practice/a progression of skills ■ The activity challenges students intellectually, creatively, and/or physically ■ The activity requires analytic thinking 	Alpha: 0.88 Mean: 4.35 SD: 1.69

Similarly, the New Jersey After 3 Year 1 study reported on four scales from the OST instrument, based on 179 observations across 10 programs (see Exhibit 6). While scales showed strong internal consistency across observations, analyses indicated that students' experiences and activity-quality scores varied, depending on the types of activities in which students were engaged (Kim et al., 2006).

Exhibit 6
Internal Consistency of Observation Scales for New Jersey After 3 Study (2006)

Observation Scale	Indicators	Statistics
Youth Relationship-Building	<ul style="list-style-type: none"> ■ Youth are friendly and relaxed with one another ■ Youth respect one another ■ Youth show positive affect to staff 	Alpha: 0.80 Mean: 5.86 SD: 0.81
Staff Relationship-Building	<ul style="list-style-type: none"> ■ Staff use positive behavior management techniques ■ Staff promote participation of all youth ■ Staff show positive affect to youth ■ Staff attentively listen to and/or observe youth ■ Staff encourage youth to share their ideas, opinions and concerns about the content of the activity ■ Staff engage personally with youth 	Alpha: 0.87 Mean: 4.50 SD: 0.87
Instructional Methods	<ul style="list-style-type: none"> ■ Staff communicate goals, purposes, expectations ■ Staff verbally recognize youth's efforts and accomplishments ■ Staff assist youth without taking control ■ Staff ask youth to expand on their answers and ideas ■ Staff challenge youth to move beyond their current level of competency ■ Staff employ varied teaching strategies ■ Staff plan for/ask youth to work together 	Alpha: 0.81 Mean: 3.60 SD: 1.23
Activity Content and Structure	<ul style="list-style-type: none"> ■ The activity is well organized ■ The activity involves the practice/a progression of skills ■ The activity challenges students intellectually, creatively, and/or physically ■ The activity requires analytic thinking 	Alpha: 0.83 Mean: 4.44 SD: 1.52

In Year 1 of the New York City OST evaluation, researchers conducted a total of 238 independent observations in 15 OST programs. Results were reported on three quality scales that integrated indicators differently from previous studies. For this evaluation, evaluators reported data using three scales: youth and staff relationships, staff instructional strategies, and activity content and structure. As seen in Exhibit 7, Cronbach's Alpha statistics demonstrated high levels of internal consistency within each scale, indicating that the items strongly capture the underlying content.

In 2007, Durlak and Weissberg (2007) reported the results of a meta-evaluation of studies reporting outcomes from OST programs that sought to improve participants' personal social skills, demonstrating that programs can improve school and non-school outcomes by enhancing youth's personal and social skills. In their analysis, Durlak and Weissberg found programs that had achieved positive academic youth outcomes employed sequential, focused, explicit, and active learning strategies, described collectively as SAFE features:

Exhibit 7
Internal Consistency of Observation Scales for New York City OST Study (2006)

Observation Scale	Indicators	Statistics
Relationships	<ul style="list-style-type: none"> ■ Youth are friendly and relaxed with one another ■ Youth respect one another ■ Youth show positive affect to staff ■ Youth are on-task ■ Youth listen actively and attentively to peers and staff ■ All or most youth take leadership responsibility/roles ■ Staff use positive behavior management techniques ■ Staff show positive affect toward youth ■ Staff attentively listen to and/or observe youth 	Alpha: 0.88 Mean: 5.31 SD: 0.96
Instructional Strategies	<ul style="list-style-type: none"> ■ Youth contribute opinions, ideas, and/or concerns to discussions ■ Staff encourage youth to share their ideas, opinions, and concerns about the content of the activity ■ Staff communicate goals, purposes, expectations ■ Staff verbally recognize youth's efforts and accomplishments ■ Staff assist youth without taking control ■ Staff ask youth to expand on their answers and ideas ■ Staff challenge youth to move beyond their current level of competency ■ Staff employ varied teaching strategies 	Alpha: 0.87 Mean: 3.06 SD: 1.46
Activity Content and Structure	<ul style="list-style-type: none"> ■ The activity is well organized ■ The activity involves the practice/a progression of skills ■ The activity challenges students intellectually, creatively, and/or physically ■ The activity requires analytic thinking 	Alpha: 0.84 Mean: 4.26 SD: 1.67

- Sequenced (the activity builds on skills and content to achieve goals)
- Active (the activity focuses on developing positive relationships between youth and staff)
- Focused (the activity focuses on developing positive relationships among youth and with staff)
- Explicit (the activity explicitly targets specific learning and/or social development goals)

Because the Durlak and Weissberg SAFE framework was theoretically consistent with the domain structure of the OST Observation Instrument and also aligned with the theories of change for both the New York City OST and the New Jersey After 3 studies, evaluators used it to interpret their Year 2 findings. Analyses demonstrated that the OST indicators map well to

the SAFE framework and produce equally strong and reliable scales as those used in previous years. The resulting scale structure and combined statistics for both studies are reported in Exhibit 8.

Exhibit 8
Internal Consistency of Observation Scales for Studies of New York City OST (2008)
and New Jersey After 3 (2007)

Observation Scale	Indicators	Statistics
Sequenced	<ul style="list-style-type: none"> ■ Activity involves the practice or a progression of skills ■ Activity requires analytic thinking ■ Activity challenges students intellectually, creatively, developmentally, and/or physically ■ Staff employs varied teaching strategies ■ Staff challenges youth to move beyond their current level of competency ■ Staff assists youth without taking control ■ Staff verbally recognizes youth efforts and accomplishments 	Alpha: 0.91 Mean: 3.63 SD: 1.60
Active	<ul style="list-style-type: none"> ■ Youth are collaborative ■ Youth take leadership responsibilities and roles ■ Youth have opportunities to make meaningful choices ■ Youth assist one another ■ Youth contribute opinions, ideas, and concerns to discussions ■ Staff plan for and ask youth to work together ■ Staff encourages youth to share their ideas, opinions and concerns ■ Staff asks youth to expand on their answers and ideas 	Alpha: 0.76 Mean: 1.90 SD: 1.00
Focused	<ul style="list-style-type: none"> ■ Youth show positive affect to staff ■ Youth are friendly and relaxed with one another ■ Youth respect one another ■ Staff shows positive affect toward youth ■ Staff engages personally with youth ■ Staff guides for positive peer interactions ■ Staff uses positive behavior management techniques ■ Staff is equitable and inclusive 	Alpha: 0.89 Mean: 4.32 SD: 0.74
Explicit	<ul style="list-style-type: none"> ■ Activity is well organized ■ Youth are on task ■ Youth listen actively and attentively to peers and staff ■ Staff communicates goals, purposes, and expectations ■ Staff attentively listens to and/or observes youth 	Alpha: 0.77 Mean: 5.05 SD: 1.16

Finally, a separate factor analysis was conducted to explore the alignment of Durlak and Weisberg’s SAFE framework with the OST Observation Instrument’s domains and indicators. This analysis distributed the indicators into four factors consistent with the SAFE structure. The elements of the four scales were identified with a principal component analysis of the data generated during the observations in the 25 programs studied in 2006-07 in New York City and New Jersey. The factors were further delineated through a varimax orthogonal rotation that converged in eight iterations. The Kaiser-Meyer-Olkin measure of sampling adequacy was .89, indicating that the sample size was adequate to confirm that these factors are reliable; if a follow-up analysis were conducted, a similar categorization of items would likely result.

Concurrent Validity

The research team explored the concurrent validity of the OST Observation Instrument by comparing it with the other instruments used in the New York City OST and New Jersey After 3 evaluations in 2006 (Year 1) and 2007 (Year 2). These analyses combined observation data from the two studies to compare program-level observation ratings and program-level scale scores on participant surveys.

Researchers used rank-order correlations to examine the extent to which observer ratings using the OST Observation Instrument aligned with participant survey results. For these analyses, researchers examined the scales created in Year 1 of the New York City OST study (see Exhibit 7 for a list of indicators in the scales.) In the combined data from Year 1, the relationships and instructional strategies scales correlated strongly with four participant survey scales relating to exposure to new experiences, sense of belonging, and staff and peer interactions. The instrument’s activity and content scale was not associated with any of the participant survey scales. Results summarized in Exhibit 9 demonstrate concurrent validity in Year 1.

Exhibit 9
Correlation between Observation and Participant Survey Scales in
New York City and NJA3 Evaluations

Observation Scale	Participant Survey Scale	2006	2007
Relationships	Exposure to new experiences	0.44*	0.44*
Relationships	Sense of belonging	0.56**	0.24
Relationships	Interactions with staff	0.43*	0.47**
Relationships	Interactions with peers	0.59*	0.45*
Relationships	Academic benefits	0.01	0.52**
Instructional Strategies	Exposure to new experiences	0.39*	-0.15
Instructional Strategies	Sense of belonging	0.37*	-0.20
Instructional Strategies	Interactions with staff	0.39*	0.06

Note: All correlations are Spearman’s Rho

*p<0.05 , **p<0.01 (1-tailed)

In Year 2, the correlation analyses also indicated statistically significant support for the instrument's construct validity with the relationship scales but did not support the construct validity of the instructional strategies scale. While the participant surveys measure relationships among youth and with staff, they do not explicitly measure instructional strategies or activity quality. It is therefore reasonable that the OST Observation Instrument and participant survey were more likely to correlate on the relationship scale than on other scales.

Conclusion

Overall, the reliability, internal consistency, and construct validity analyses conducted for the OST Observation Instrument demonstrate strong, consistent evidence of the instrument's usefulness in documenting the quality features of OST programs. The five studies using the instrument since 2005 included 35 OST program sites implemented by different local provider organizations over several years. Across the studies, observers obtained high inter-rater reliability ratings. Additional statistical analyses demonstrated strong reliability and validity of the evidence collected using this measure. Internal consistency assessments demonstrated that the instrument's indicators can be combined effectively in various ways to assess different program emphases and change theories. Finally, youth participant surveys demonstrated the alignment of OST relationship domains with survey evidence of youth and staff interaction, youth sense of belonging, academic benefits, and youth exposure to new experiences.

Together this evidence confirms that the OST Observation Instrument produces reliable and valid ratings of OST programs that can be used with confidence by program administrators, developers, and advisors. Although the studies reported here did not involve the instrument's use by on-site program directors or staff, the analyses suggest that program-level users can be confident that, with appropriate training, the instrument can be a valuable resource in assessing the presence or absence of research-based quality indicators.

PSA will continue to conduct and report analyses as the instrument is used and, with each evaluation, additional statistical properties will be reviewed and reported.

References

- Birmingham, J., Pechman, E. M., Russell, C. A., & Mielke, M. (2005). *Shared features of high-performing after-school programs: A follow-up to the TASC evaluation*. Washington, DC: Policy Studies Associates.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Center for After-School Excellence. (2007). *21st Century Community Learning Center evaluation report (2005-06)*. New York, NY: TASC/The Center for After-School Excellence.
- 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 (CASEL); University of Illinois at Chicago.
- Eccles, J. S., & Gootman, J. A. (Eds.). (2002). *Community programs to promote youth development*. Washington, DC: National Academy Press.
- Grossman, J., Campbell, J., & Raley, B. (2007). *Quality time after school: What instructors can do to enhance learning*. Philadelphia, PA: Public/Private Ventures.
- Kim, J. H., Miller, T. D., Reisner, E. R., & Walking Eagle, K. P. (2006). *Evaluation of New Jersey After 3: First-year report on programs and participants*. Washington, DC: Policy Studies Associates.
- Marzke, C., Pechman, E. M., Reisner, E., Vandell, D., Pierce, K. M., & Brown, B. B. (2002). *Study of promising programs: Theory of change guiding the research*. Washington, DC: Policy Studies Associates.
- McLaughlin, M. W. (2000). *Community counts: How youth organizations matter for youth development*. Washington, DC: Public Education Network.
- Mott Foundation Committee on After-School Research and Practice. (2005). *Moving towards success: Framework for after-school programs*. Washington, DC: Collaborative Communications Group.
- Penuel, W., & McGhee, R. (forthcoming). *21st Century Community Learning Centers study of the quality of academic instruction*. Arlington, VA: SRI International.
- Pittman, K., Irby, M., & Ferber, T. (2001). Unfinished business: Further reflections on a decade of promoting youth development. In *Youth development: Issues, challenges and directions*. Philadelphia, PA: Public/Private Ventures.

- Reisner, E. R., White, R. N., Russell, C. A., & Birmingham, J. (2004). *Building quality, scale, and effectiveness in after-school programs: Major findings of the TASC evaluation*. Washington, DC: Policy Studies Associates.
- Russell, C. A., Mielke, M. B., & Reisner, E. R. (2008). *Evaluation of the New York City Department of Youth and Community Development Out-of-School Time Programs for Youth Initiative: Results of efforts to increase program quality and scale in Year 2*. Washington, DC: Policy Studies Associates.
- Russell, C. A., Reisner, E. R., Pearson, L. M., Afolabi, K. P., Miller, T. D., & Mielke, M. B. (2006). *Evaluation of the Out-of-School Time Initiative: Report on the first year*. Washington, DC: Policy Studies Associates.
- Vandell, D. L., Reisner, E. R., Pierce, K. M., Brown, B. B., Lee, D., & Bolt, D. (2006). *The Study of Promising After-School Programs: Examination of longer term outcomes after two years of program experiences*. Madison, Wisconsin: University of Wisconsin.
- Walking Eagle, K. P., Miller, T. D., Reisner, E. R., Johnson, J. C., Mielke, M. B., Edwards, S. K., & Farber, M. H. (2007). *Evaluation of New Jersey After 3: Second-year report on program and participants*. Washington, DC: Policy Studies Associates, Inc.
- Yohalem, N., & Wilson-Ahlstrom, A. (2007). *Measuring youth program quality: A guide to assessment tools*. The Forum for Youth Investment.