High/Scope Youth PQA Technical Report

Findings From the Self-Assessment Pilot in Michigan 21st Century Learning Centers

Charles Smith
High/Scope Educational Research Foundation

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

Contents

Summary **4**

Part I. The Youth PQA Self-Assessment Pilot Study in Michigan 21st Century Programs **5**
The Program Self-Assessment Method 6
Pilot Study Sites and Staffing 7

Part II. Quality Ratings in 21st Century Programs **9**
Interpretation of Subscale and Item Scores 10
Conclusion: The Self-Assessment Process Yields Interpretable Data 12

Part III. Psychometric Performance of the Youth PQA Using the Program Self-Assessment Method **14**
Score Distributions 14
Scale Reliability 15
Concurrent Validity 16
Conclusion: The Self-Assessment Process Yields Psychometrically Acceptable Data 17

Part IV. Evaluation of the Self-Assessment Process by 21st Century Staff **18**
Survey of Site Administrators From the Pilot Study 18
Interviews With Site Administrators From the Pilot Study 18
Conclusion: The Self-Assessment Process Supports Organizational Learning and Change 19

Part V. Use of the Youth PQA and Youth PQA Data for Program Improvement **21**
Improvement Models Developed Based on Pilot Study Data 21
Conclusion: The Self-Assessment Process Supports System-Level Decision Making About Practices and Policies That Raise Quality 21

References **22**

Appendix A. Interview Transcripts **23**

Appendix B. Program Improvement Models **30**

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

Summary

Overall 24 sites within 17 grantees participated in the self-assessment pilot study by assembling staff teams to collect data and score the Youth Program Quality Assessment (PQA).

At each site an average of 5 staff spent an average of 13 staff hours to complete the self-assessment process.

Whether using an absolute standard or group norms as a benchmark for interpretation of data from the Youth PQA Self-Assessment Pilot Study (hereafter called the Pilot Study), quality scores were very positive for participating programs and also reflected the tendency of self-assessment scores to be biased toward higher quality levels.

The quality scores followed the same pattern as outside observer scores in other samples, highest on for issues of safety and staff support and lowest on higher order practices focused on interaction and engagement.

Youth PQA data collected using the self-assessment method demonstrated promising patterns of both internal consistency and concurrent validity with aligned youth survey responses.

Two thirds or more of sites reported that the observation and scoring process helped the self-assessment team to have greater insight into the operation of their programs, talk in greater depth about the program quality than usual, and have more concrete understanding of program quality.

Site directors and local evaluators said that the self-assessment process was a source of good conversations about program priorities and how to meet them. In almost all cases, concrete action followed from the self-assessment process.

Site directors and local evaluators demonstrated the ability to improvise the self-assessment method to fit local needs.

Program directors, site coordinators, and local evaluators have used the Youth PQA and statewide Youth PQA data to generate statewide program change models, suggesting that the instrument and data are useful for setting system-level improvement priorities.

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

Part I. The Youth PQA Self-Assessment Pilot Study in Michigan 21st Century Programs

The Youth Program Quality Assessment (PQA) is an assessment of best practices in afterschool programs, community organizations, schools, summer programs, and other places where youth have fun, work, and learn with adults. The Youth PQA consists of seven subscales: *safe environment, supportive environment, interaction, engagement, youth-centered policies and practices, high expectations for all youth and staff,* and *access*. Administration of the Youth PQA employs direct observation for the first four subscales and a structured interview for the remaining three. The instrument is structured by item-level measurement rubrics that consist of multiple indicators. Indicators are scored using observation and interview evidence and then averaged up to the item and subscale (multi-item) levels. The Youth PQA was developed and validated during a 4-year validation study funded by the W. T. Grant Foundation.[1]

Since 2003 the High/Scope Educational Research Foundation has collaborated with the Michigan Department of Education 21st Century Community Learning Centers program to develop a *low-stakes accountability and improvement system* based on the Michigan Model Standards for Out of School Time Standards.[2] The system uses the Youth PQA as a primary accountability and improvement measure. A low-stakes accountability system requires that grantees report on development and implementation of high-quality, data-driven improvement plans and emphasizes organizational learning and development of staff competencies based on model standards of practice. Low-stakes accountability is an alternative to higher stakes approaches that focus on licensing compliance and/or year-end child outcomes but which provide little guidance to staff about how to actually change and improve organizational performance.

The Youth PQA is a dual-purpose instrument, designed as both a rigorous measure of quality and as an effective support to organizational learning and staff development. A recently completed validation study established the reliability and validity of Youth PQA scores produced using intensive data collection procedures that employed trained outside observers and multiple ratings at each site.[3] This report focuses on instrument performance under a different data collection protocol, the *program self-assessment data collection method*, designed specifically for use in Michigan 21st Century afterschool programs. The program self-assessment method was developed to make both the assessment process more cost-effective and to support a process of organizational learning by building the program staff into the process of data collection, scoring, and interpretation of findings.

The Youth PQA Self-Assessment Pilot Study (hereafter called the Pilot Study) was conducted to provide validation evidence for the Youth PQA when using the program self-assessment data collection method. One central set research questions is focused on psychometric performance. Does the self-assessment data achieve widely agreed upon benchmarks for score reliability and

---

[1] For more information on the Youth PQA, visit the Youth Development Group page at the High/Scope Web site, youth.highscope.org. For findings from the 4-year Youth PQA Validation Study, see the report entitled *Youth Program Quality Assessment Validation Study: Findings for Instrument Validation*, also available at youth.highscope.org.

[2] The Michigan Model Standards for Out of School Time are available at http://www.michigan.gov/documents/OST_Standards_43292_7.pdf.

[3] The original Youth PQA Validation Study employed trained outside observers and required completion of three separate offering-level ratings at each participating organization. The three separate ratings were averaged together to create an offering quality score for an entire organization. This method was too invasive (using outside observers not connected to the program), time consuming, and expensive for use in Michigan's 21st Century programs.

validity (see discussion in Part III)? However, the primary focus of this study is to determine how successfully program staff are able to use the data to improve personal and organizational performance. We want to know if the scores are fair (reliable) and meaningful (valid) when the instrument is used as a self-assessment and how these findings compare to similar tests of data precision from trained outside observers. We also want to know if the self-assessment process actually produces organizational learning about quality[4] so we are really validating an assessment process that uses the Youth PQA rather than just an assessment tool alone.

Primary research questions that guided the Pilot Study include:

- What does the Pilot Study tell us about program quality in Michigan 21st Century afterschool programs?
- What are the psychometric characteristics of data collected using the program self-assessment data collection method?
- Did use of the Youth PQA support high-impact conversations about quality among staff?
- Were local teams able to improvise with the self-assessment protocol to make the process supportive, informative, and doable for program staff?
- Did afterschool administrators find the Youth PQA useful for local accountability purposes such as evaluation of line staff performance or communication with school personnel about best practices in afterschool?
- Were cross-site groups of local program leaders able to use the Youth PQA and aggregated Pilot Study scores for system-level program improvement planning?

The Program Self-Assessment Method

The program self-assessment method was developed for use of the Youth PQA in Michigan 21st Century programs. For the Pilot Study, program staff attended a 1-day training that was focused on learning how to collect objective anecdotal evidence and which provided practice scoring each of the instrument's 103 indicators using actual data from the validation study. Participating grantees were advised to conduct the assessment at only one site, preferably at middle schools or sites with older elementary children.[5] The self-assessment method consisted of the following five steps:

- *Step 1:* Select a team of staff for data collection. The team will collect observational data by observing each other and then complete the administrative interview. The

[4] We do not take this analysis toward the critical discussion of whether or not it is possible to have validity without reliability — we are trying to achieve at least marginal measurement reliability and consequential validity at the same time. See Moss (1994).

[5] The Younger Youth (Grades K–6) version of the Youth PQA is currently the subject of its own validation study.

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

team should consist of at least two direct service staff and one program supervisor or site coordinator. More are welcome.

- *Step 2:* Select at 2-week time period (roughly) when anecdotal data will be collected. Set a goal of 10–15 anecdotal records for each person that will result in a total of 40–50 anecdotal records for the whole team. One anecdotal record is roughly the amount of written record that will fit on a 2 x 4 inch post-it note.

- *Step 3:* Start collecting data. Have members of the team observe and collect anecdotal records while the other members of the team are leading youth in the program. Do not use staff names.

- *Step 4:* At the end of the first week, do a check in. Make sure that you are collecting datum that is a "fit" with the items in the Youth PQA Form A booklet (program-offering items). Keep collecting data. Fitting data refers to the process of checking to make sure that you have data that apply to the scoring indicators on the Youth PQA.

- *Step 5:* Schedule a 3-hour scoring meeting with the team. Make sure that you clearly mark one "score" for each indicator in both of the booklets (unless it really doesn't apply to your program or you really just disagree with it). Do not score the booklets beyond the indicator levels until all of the indicators have been scored.

    - Start by scoring the indicators for items in Form A (program-offering items) using the only the anecdotal data that you have collected. Try not to use your prior knowledge — only the data from the anecdotal records. If you need more data, go back and collect more if you can.

    - Next, score the indicators for items in Form B (organization items). Score these indicators by discussing each indicator and selecting the best score that the site team can reach a consensus on.

Self-assessment is designed to represent enough of the reality of a youth program to support powerful conversation and decision making about program polices and staff practices. It relies on staff decisions about what to watch and an explicitly consensual process for completion of the organizational-level form. This method was designed to first maximize effective participation and secondarily to ensure objective measurement. Reliability and validity of the data produced when using the self-assessment method is discussed in Part III. The effects of the Youth PQA and the self-assessment data collection method for staff learning and program improvement are discussed in Parts IV and V.

Pilot Study Sites and Staffing

The Pilot Study was conducted at 24 afterschool sites within 17 of Michigan 21st Century grantees. Each participating grantee used the program self-assessment method to complete a Youth PQA rating for at least one afterschool site per grantee. Some grantees elected to complete Youth PQA ratings for multiple sites. The sample included afterschool programs at 7 elementary schools, 14

middle schools, and 1 high school (2 of the reports were not labeled). Self-assessment teams were primarily composed of program directors, line staff, and local evaluators and over two thirds of the teams contained at least one of each of these persons. On average, sites required 13 hours of staff time to complete Forms A and B of the Youth PQA. The average number of frontline staff involved in the self-assessment process was five.

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

Part II. Quality Ratings in 21st Century Programs

Program quality ratings for the Pilot Study sample are provided in Table 1 with a comparison to two other samples where Youth PQA data were collected in the state of Michigan. The first four subscales — safe environment, supportive environment, interaction, and engagement — were completed using data collected through observation in afterschool settings. The final three subscales — youth-centered policies and practices, high expectations for all youth and staff, and access — were completed using data collected through a group interview. For the comparison samples, the more rigorous *outside observer data-collection method* was used to produce Youth PQA scores: at the offering-level scores are an average of multiple observational ratings (usually three) for each organization and the organization-level scales were completed with evidence gathered by a trained interviewer. The second column presents data from a small sample of 21st Century afterschool programs for which outside observer data were available. The right-most column present scores for the entire Youth PQA Validation Study sample.[6]

Table 1. Youth PQA Scores for the Pilot Study in Comparison With Two Additional Samples Using an Outside Observer Data-Collection Method

	Self-Assessment Pilot Study	*21st Century Sites in the Youth PQA Validation Study*	*Youth PQA Validation Study — Total Sample*
	Average Scores *N*=24	*Average Scores N=11*	*Average Scores N=56*
I. Safe Environment	4.39	*4.16*	*4.35*
II. Supportive Environment	4.16	*3.91*	*3.74*
III. Interaction	3.73	*2.83*	*3.11*
IV. Engagement	3.37	*2.83*	*2.83*
Observation Total Score (subscales I–IV)	3.99	*3.43*	*3.51*
V. Youth-Centered Policies & Practices	3.20	*3.77*	*3.92*
VI. High Expectations for Youth & Staff	3.91	*3.83*	*3.86*
VII. Access	4.18	*3.00*	*3.86*
Interview Total Score (subscales V–VII)	3.76	*3.53*	*3.88*

[6] For findings from the 4-year Youth PQA Validation Study, see the report entitled *Youth Program Quality Assessment Validation Study: Findings for Instrument Validation*, available at youth.highscope.org.

Interpretation of Subscale and Item Scores

Interpretation of the data in Table 1 can proceed in at least two directions. Scores can be interpreted as degree of attainment of a predetermined performance standard. For example, when High/Scope certifies a classroom or teacher in use of its methods, an overall score of 4 is required for endorsement. Second, scores can be interpreted as a level of performance in relation to norms set by other programs. If most of the programs in the field are scoring near a 3 on the engagement subscale, then a score slightly over 3 might be seen as positive. Using either form of interpretation, the self-assessment scores are quite positive. All of the scores exceed scores in the other samples with the lowest scores occurring for the engagement and youth-centered policies and practices subscales.

Interpretation of these ratings is complicated by the fact that the self-assessment data collection method is known to introduce systematic bias into program quality ratings.[7] More specifically, self-assessment is believed to produce ratings that are higher than those produced by outside observers. Table 1 supports this belief, although the scores were not collected in the same programs so the issue cannot be resolved definitively by data from these samples.

Perhaps most important, scores descend incrementally from subscale I to subscale IV suggesting that items in the interaction and engagement subscales are likely candidates for improvement in these organizations. At the organization level, the self-assessment sample scored most poorly on the cluster of items that make up the youth-centered policies and practices subscale. Taken together, the three lowest scoring subscales — interaction, engagement, and youth-centered policies and practices — suggest an improvement agenda for 21st Century afterschool programs in the state of Michigan. By examining the item- and indicator-level data that were used to construct these scores, the improvement agenda can be much more precisely specified as a set of concrete organizational and staff practices.

Tables 2–3 present item- and indicator-level data that were used to construct the subscale scores presented in Table 1. This detail is provided so that the areas of relative strength and weakness in subscale scores can be examined in greater detail. Table 2 provides scores for the offering-level items, and Table 3 provides scores for the organization-level items.

[7] See Part III.

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

Table 2. Pilot Study Youth PQA Item Scores for the Offering Level (Observation Data)

*Items scores for the **Safe Environment** subscale*	Mean Score	Range
I.A. Psychological and emotional safety are promoted	4.26	*max 5 min 1*
I.B. The physical environment is safe and healthy for youth	4.67	*max 5 min 3*
I.C. Appropriate emergency procedures and supplies are present	3.88	*max 5 min 1*
I.D. Rooms and furniture accommodate activities	4.73	*max 5 min 4*
I.E. Healthy food and drinks are provided	New item	
*Item scores for the **Supportive Environment** subscale*		
II.F. Staff provides a welcoming atmosphere	4.32	*max 5 min 2.5*
II.G. Session flow is planned, presented, and paced for youth	4.39	*max 5 min 1.4*
II.H. Activities support active engagement	4.30	*max 5 min 1.7*
II.I. Staff support youth to build new skills	4.41	*max 5 min 2*
II.J. Staff support youth with encouragement	3.81	*max 5 min 1*
II.K. Staff use youth-centered approaches to reframe conflict	3.78	*max 5 min 1*
*Item scores for the **Interaction** subscale*		
III.L. Youth have opportunities to develop a sense of belonging	3.89	*max 5 min 2.5*
III.M. Youth have opportunities to participate in small groups	3.78	*max 5 min 1*
III.N. Youth have opportunities to act as group facilitators and mentors	3.39	*max 5 min 1*
III.O. Youth have opportunities for adult-youth partnership	3.96	*max 5 min 1*
*Items scores for the **Engagement** subscale*		
IV.P. Youth have opportunities to set goals and make plans	3.29	*max 5 min 1*
IV.Q. Youth have opportunities to make choices based on interests	3.67	*max 5 min 1*
IV.R. Youth have opportunities to reflect	3.22	*max 5 min 1*

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

Table 3. Pilot Study Youth PQA Item Scores for the Organization Level (Interview Data)

*Items scores for the **Youth-Centered Policies and Practices** subscale*	Mean Score	Range
V.A. Staff qualifications support a positive youth development focus	4.35	*max 5 min 3.8*
V.B. Offerings tap youth content interests to build multiple skills	3.99	*max 5 min 3.7*
V.C. Youth have influence on setting & activities in the organization	2.06	*max 5 min 1*
V.D. Youth have influence on structure & policy in the organization	2.38	*max 5 min 1*
*Item scores for the **High Expectations for Youth and Staff** subscale*		
VI.E. Organization promotes staff development	3.70	*max 5 min 2.2*
VI.F. Organization promotes supportive social norms	4.01	*max 5 min 1*
VI.G. Organization promotes high expectations for youth	4.04	*max 5 min 2*
VI.H. Organization is committed to ongoing program improvement	3.86	*max 5 min 1.8*
*Item scores for the **Access** subscale*		
VII.I. Staff availability & longevity support youth-staff relationships	4.39	*max 5 min 3.5*
VII.J. Schedules are in effect	4.39	*max 5 min 2.3*
VII.K. Barriers to participation are addressed	3.93	*max 5 min 2*
VII.L. Organization communicates w/ families, schools, & organizations	4.00	*max 5 min 2.3*

Conclusion: The Self-Assessment Process Yields Interpretable Data

The data suggest two broad implications for program improvement. First, the engagement subscale had lower scores than other Form A subscales and suggests areas for improvement related to youth planning and choice. Second, the lowest scoring subscale in the entire assessment is youth-centered policies and practices, suggesting improvements in providing opportunities for youth input in organization-level decision making.

The engagement subscale divides into three items that measure the presence of opportunities for youth to make plans, make choices, and reflect on their experiences. Programs could improve these scores by creating such opportunities for youth in program offerings. Youth organizations might invest in staff training in these areas and set expectations for staff to implement training content resulting in raised Youth PQA scores. An individual organization may also look at its particular low-scoring indicators and address those issues specifically. For example, an organization might focus on

indicator IV-R.1 and begin to offer opportunities for youth to "reflect on their participation in activities."

The youth-centered policies and procedures subscale measures youth involvement in organization-level decision making. For many of the indicators contained within items in this subscale, a high percentage of organizations received the lowest possible score, a 1. For example, indicator V-C.1, for which 76% of organizations scored a 1, reads, "Youth and adults share decision-making responsibility for design and use of physical environment." Increasing the score, and therefore the quality of youth experience in this area, would be fairly easy for an organization to do.

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

Part III. Psychometric Performance of the Youth PQA Using the Program Self-Assessment Method

The primary purpose of the Youth PQA when using the self-assessment method is to support effective conversation, planning, and action within groups of administrators and line staff. When measurement efficacy is a paramount concern, the outside observer data collection method should be used.[8] We know from our work with the Preschool Program Quality Assessment[9] that when quality ratings produced by trained outside observers are compared to quality ratings produced through self-assessment, several patterns are consistent: self-assessment ratings are higher, have smaller standard deviations, and contain less explained variance in statistical models. This kind of systematic bias raises concerns about interpretation of self-assessment data.

Several potential sources of bias are introduced by the self-assessment method. First, the program self-assessment method instructs data collectors to gather evidence for only 10–15 minute periods in each of several program offerings. Without sufficient observation time in each offering, it is possible that important staff practices and student experiences will be missed. Second, self-assessment uses a consensual and group-based method for data collection and scoring, introducing the possibility that norms of social-acceptability could influence the scoring process. Finally, because line staff were invited to observe each other in the Pilot Study, it is more likely that observers will allow prior experience to influence their scoring process, rather than relying solely on the observed evidence.

Given these potential sources of bias when using the Youth PQA as a self-assessment, it is important to know how reliable and valid the scores are. Three criteria were used to evaluate the data produced during the Youth PQA Pilot Study: score distributions, scale reliabilities, and concurrent relationships with youth survey data regarding program quality.

Score Distributions

Table 1 presented subscale mean scores for the Pilot Study in comparison to two additional samples: 21st Century programs assessed through multiple outside observations as part of the Youth PQA Validation Study (second column) and the total sample for the Youth PQA Validation Study. For the observational subscales, scores are about one half of a point higher in the Pilot Study, suggesting the expected positive bias in the scores. However, the general pattern of the observational scores descending steadily from highest on subscale I to lowest on subscale IV is the same across all three samples, suggesting that the Youth PQA is capturing relative levels of quality using the program self-assessment method, even if the scores are biased upward.

For the organization-level subscales, the comparative pattern is less clear. The interesting point is that for the youth-centered policies and practices subscale, the self-assessment teams appear to have scored themselves more critically than the outside interviewers in the validation study — unlike the pattern in the observational data where self-assessment scores were uniformly higher than outside observer scores.

[8] For a thorough description of the continuum of use and data collection procedures, see High/Scope Educational Research Foundation (2005).
[9] For a review of research on the High/Scope Preschool PQA, see High/Scope Educational Research Foundation (2003).

Tables 2 and 3 presented item-level scores for both the organization- and offering-level subscales with score ranges. The score ranges demonstrate that for most of the items the lowest part of the 5-point scale is being used. This alleviates some concern regarding bias toward the upper bound of the scale.

Scale Reliability

Table 4 presents alpha coefficients for Youth PQA subscales using data from the Pilot Study. Three of the observational scales — supportive environment, interaction, and engagement — achieve acceptable alpha levels[10] meaning that the items in each scale do appear to measure different dimensions of a related construct. In comparison to the much larger sample from the Youth PQA Validation Study, these alpha coefficients are similar — with the exception of the youth-centered policies and practices and high expectations for all students and staff scales that have low alphas in this sample. Scales I and VII were not expected to produce high alpha levels since they include items that are individually important to the named construct but are not likely to co-occur in all settings.

Table 4. Alpha Coefficients for Youth PQA Scales

		Self-Assessment Pilot Study N=24	*Youth PQA Validation Study — Wave 1 Sample N=22*	*Youth PQA Validation Study — Wave 2 Sample N=118*
I.	Safe Environment (5 items)	.45	.38	.43
II.	Supportive Environment (6 items)	.85	.85	.84
III.	Interaction (4 items)	.72	.72	.64
IV.	Engagement (3 items)	.71	.71	.70
V.	Youth-Centered Policies and Practice (4 items)	.50	NA	.71
VI.	High Expectations for All Students and Staff (4 items)	.49	NA	.68
VII.	Access (4 items)	.50	NA	.45

Sample size, which is small at N=24, is of course a concern. However, given the small sample size and use of the program self-assessment method, these are fairly positive results for psychometric performance of the instrument.

[10] Nunally's (1978) criteria of .70 is widely used as the acceptable standard for scale reliability, although Nunally's earlier work (1967) and research by Davis (1964) view alpha's as low as .60 as being acceptable. For a thorough review, see Peterson (1994).

Concurrent Validity

Concurrent validity of the Youth PQA data collected using the program self-assessment method was evaluated using student responses to survey items collected as part of the annual statewide evaluation of 21st Century programs.[11] Table 5 provides Pearson-r correlation coefficients for Youth PQA self-assessment data and aligned subscales from a survey of youth in these programs. Youth survey responses for Grades 4–12 are aggregated to the site level for a total sample size of 12 pilot sites that had both Youth PQA and youth survey data available.

Table 5. Bivariate Correlation: Youth PQA Subscales and Aligned Youth Survey Subscales for Children Grades 4–12

Youth PQA Scales	Aligned Youth Survey Subscale and Pearson-*r* Correlation Coefficient	Items on the Aligned Youth Survey Subscales
I. Safe Environment	NA	
II. Supportive Environment	Staff Support = .83**	Staff Support: (1) Staff here keep their promises; (2) Staff here try to be fair; (3) When staff tell me not to do something, I know they have a good reason; (4) Staff listen to our ideas about how to make the program better; (5) Staff here care about me; (6) I feel safe and comfortable with staff.
III. Interaction	NA	
IV. Engagement	Program Governance=.40++	Program Governance: (1) Kids get to decide what goes on with this program; (2) Staff and students decide together what the rules will be; (3) Kids help plan what they will do; (4) Kids are asked what they think should happen in the program; (5) Staff and students decide together how to do activities.
Offering Total Score	Academic Support = .50+	Academic help: This program…(1) helped me stay caught up with my homework; (2) helped me understand what I was doing in class; (3) matched the things we were doing in class; (4) Helped me learn school subjects in new and interesting ways
	Peer Relations = .42++	Peer relations: (1) Kids help each other even if they aren't friends; (2) Kids treat each other with respect; (3) Kids work together to solve problems; (4) When I'm having a problem other kids help me; (5) Kids here really care about each other.

**p≤.01; *p≤.05; +p≤.1; ++p≤.2

The staff support and program governance subscales were reasonable matches with the Youth PQA's supportive environment and engagement subscales. Levels of correlation were substantial for both the supportive environment and engagement subscales and the aligned youth survey measures, although not always statistically significant at the p<.05 level. There were no questions on the youth

[11] Michigan's statewide evaluation of the 21st Century program is conducted by a team of researchers at Michigan State University. See the description at http://outreach.msu.edu/ucp/initEdu.asp.

survey that provided a concurrent test of Youth PQA's safe environment or interaction subscales.

The final row of Table 5 provides other bivariate correlations of interest for the offering-level Youth PQA data. The Youth PQA's global quality rating for the offering level, the total score for subscales I–IV, is positively associated with youth survey items tapping support for academic work and feelings of attachment between peers suggesting that in high-quality programs youth experience positive relationships with their peers and are able to do work that supports school success.

Although there were no concurrent youth responses for Youth PQA items at the organization level, students' overall ratings of satisfaction with the program were also positively associated with the youth-centered policies and practices subscale (r=.46, p=.14).

Conclusion: The Self-Assessment Process Yields Psychometrically Acceptable Data

The findings presented above provide reason for optimism concerning the reliability and validity of Youth PQA self-assessment data. The three observational scales with greatest relevance to the field of afterschool — supportive environment, interaction, and engagement — produced measures of internal consistency that paralleled findings for samples collected by outside observers. However, it also appears that self-assessment data for the observational scales may be biased toward higher scores. The observational subscales I–IV demonstrated positive patterns of bivariate association with aggregated responses from youth at pilot study sites, suggesting a concurrent evidence of quality from the perspective of both outside observers using the Youth PQA and youth participants in the programs.

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

Part IV. Evaluation of the Self-Assessment Process by 21st Century Staff

Following the data collection phase of the Pilot Study, surveys and interviews were used to gather feedback from site directors and local evaluators about how the self-assessment process went at their sites. This section provides feedback from 12 of the 17 grantees that participated in the Pilot Study, nine survey responses and five interviews (with two of the sites completing both).

Survey of Site Administrators From the Pilot Study

Survey responses were gathered from multiple staff at 9 of the 17 participating pilot sites. Self-assessment teams were primarily composed of program directors, line staff, and local evaluators, and more than two thirds of the teams contained at least one of each of these persons.

In general, the self-assessment process supported good conversation and learning among staff at the pilot sites. When asked how the process of observing impacted them, site staff reported that they gained greater insight into the operation of their programs (89%), talked in greater depth about program quality (67%), and had a more concrete understanding of program quality (78%). Similar responses followed regarding the process of scoring the instrument. Staff reported that scoring the instrument led to greater insight into the operation of their programs (78%), talking in greater depth about program quality (89%), and development of a more concrete understanding of program quality (78%).

Site directors and local evaluators reported that that the Youth PQA training did effectively prepare them to collect useful anecdotal evidence; however, 56% stated that they wanted their line staff to attend the training. Site directors reported the following when asked how line staff responded to the self-assessment process: "Line staff who work directly with program activity offerings were not involved beyond providing input"; "Next step is to provide additional PD to them"; "Helpful/insightful...time consuming"; "Interesting, gave them a model to run program effectively"; "It was helpful in that it showed areas of improvement but skewed since it was self-evaluation"; "Misunderstood at first, but eventually understood and got a deeper understanding of purpose"; "Time consuming but useful/had to work at condensing anecdotes."

Site directors reported that next year they would change the process of collecting observational data and scoring of the instrument in the following ways: "Collect any data vs. positive data"; "Use the instrument to directly gather info — not collect anecdotes. It is too time consuming"; "More staff and time"; "Not sure yet. We at least want this on a computer database for direct entry. We are currently planning for next year and will incorporate these things into our map of the year"; "Nothing — our process worked fine"; "Our goal is that site coordinators will lead this effort and facilitate site-bound PD using the Youth PQA"; "Start earlier, get more collectors, train!"

Interviews With Site Administrators From the Pilot Study

Interviews were conducted with program directors and local evaluators at 5 of the 17 grantees (see Appendix A for the interview transcripts/notes). Several themes emerged from the interviews, and nearly all were positive. Some interviewees thought that the process took too much time.

Like the survey, the interviews include repeated references to good conversations about program quality. Further, the interviews provide specific examples of substantive changes in program

operation that followed from reflection on Youth PQA data. The interviews also demonstrate that local sites were improvising with the data collection/reflection process to fit their own circumstances. Sites collected data and scored the instrument using numerous configurations of staff. Self-assessment teams included staff from the site, graduate students from a local university, program directors, and paraprofessionals. At other sites data were not collected by teams but by a local evaluator or the program director alone.

Also, different methods were used to return Youth PQA data to site directors and line staff: some just had a conversation about scores once the self-assessment process had been completed, while others used the process of scoring to generate conversation. One of the local evaluators described framing a few key issues for site directors to take action around. Another program director completed the data collection and scoring and then routed the data through building principals to discuss with line staff. At another site, a pilot Youth PQA was completed during a scoring session with all of the evidence stuck to the wall on post-it notes.

Conclusion: The Self-Assessment Process Supports Organizational Learning and Change

Although some site staff found the self-assessment process too time consuming, most of the feedback from staff was very positive. Site directors were able to customize the self-assessment process to fit their organizational constraints and reported that individual and organizational learning about program quality did occur as a result of the process. Several sites reported that concrete program change followed the self-assessment process.

The Youth PQA program-use vignette below was provided by a 21st Century program director on the back of one of the program surveys. This feedback reiterates several themes from the Pilot Study regarding use of the Youth PQA and the self-assessment method by program staff. First, staff value the opportunity to watch and gather evidence about what happens in their program. Second, the combination of detailed quality rubrics and evidence from the program defines pathways of action — Youth PQA users are able to make the jump from data to action steps that lead to improvement of quality.

The third and most important message from the vignette is that training and practice are critical to helping users get more efficient and more reliable. We know from extensive experience in preschool and out-of-school-time settings that staff can be become efficient producers of quality data and that quality assessment does not have to be a time burden on programs. We also know that subjectivity in the scoring process can be reduced to acceptable levels of rater agreement. Both of these issues emerge — efficiency and subjectivity — in a program where a program director with little prior assessment experience received minimal training (1 day) and then several months later trained staff to complete the instrument. From one perspective the vignette is very positive. Even when limited resources dictate minimal preparation for staff, a program with strong leadership but little prior assessment experience can complete the self-assessment process and learn from it. However, with more support (more training for staff or more local expertise to help), these critical start-up costs can be reduced dramatically.

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

Youth PQA Program-Use Vignette:

Informative — we were able to use the comparison to other programs in the state as a measuring stick, but the most useful piece was that it required us to observe the program from outside of our daily roles. The anecdotal evidence helped us to focus on specific areas of our program. We realized from our observations, and interviews of staff and students, that we were not offering activity choices for what our students could participate in. We also did not allow for students to choose "how" that performed a task or activity. With that information, we have decided to offer "clubs" that each student can join and have a choice in what activities they participate in. Just one example of how we were able to use the information. Again, the most beneficial aspect of the YPQA was that it required staff to self-evaluate the program and find proof or evidence of specific program areas. Some times what you think you do is different than what happens and the YPQA does not allow you to "fake" an evaluation.

Miserable — the YPQA was a significant time commitment. We also found that we had differing standards between our three program sites. What one site felt was a 3, the other was sure that it should be a 1 when we completed the program level booklet. I can only imagine that the same would be true if the sites evaluated each other along with their own. One site might give all 5's while the other might give all 3's. The other difficulty we had was that we did not do our evaluation until February (several months after our training). The staff (specifically me) was confused about much of the evaluation and how we were supposed to gather evidence and score it. After doing once, I know that we could have collected better information and scored the program more accurately.

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

Part V. Use of the Youth PQA and Youth PQA Data for Program Improvement

Improvement Models Developed Based on Pilot Study Data

On April 8, 2005, a meeting of approximately 75 program coordinators, site directors, and local evaluators from across the state engaged in a program improvement session using data from the Pilot Study to develop system-level program improvement models. These logic models provide a unique perspective on the thinking of program leadership across the state.

Staff worked in table groups to analyze Youth PQA presented in Table 1 and to develop logic models that described the improvement priorities for their cross-site teams. Teams were instructed to pick one priority area of quality and to develop a quality improvement logic model that described a sequence of action and outcomes that that would apply to their own programs. Each team was asked to create a three-column model, moving from left to right on the page: The middle column (center) was the quality improvement area that the team selected and represented the actions of line staff that could be assessed using the Youth PQA. The first column (left) described the action of administrators or other resources that could be used to initiate and support the quality improvement area(s) listed in the middle column. The third column (right) describes student-level outputs and/or outcomes that would follow from implementation of the quality improvement items listed in the second column.

Twelve program improvement models developed by the leadership staff and local evaluators of many of Michigan 21st Century grantees are presented as Appendix B. Interestingly, 6 of the 12 teams chose to focus on organizational-level factors that would give youth greater influence over program design and delivery. Two teams each focused on offering-level factors related to opportunities for active engagement and student reflection. One team focused on organization-level factors related to high expectations for youth. The final team focused on offering-level opportunities for youth to mentor and act as group facilitators.

Conclusion: The Self-Assessment Process Supports System-Level Decision Making About Practices and Policies That Raise Quality

As demonstrated in Appendix B, Youth PQA data from the Pilot Study were successfully interpreted by leadership teams from 21st Century programs across the state of Michigan and translated into generic quality improvement plans. Interestingly, the content of these plans converge with the sample program improvement agenda described in the second section of this report. In general, 21st Century leadership would like to see program change in the following areas: staff practices related to youth planning, decision making and reflection, and organizational policies related to how youth input is built into program operation.

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

References

Davis, F. B. (1964). *Educational measurements and their interpretation.* Belmont, CA: Wadsworth.

High/Scope Educational Research Foundation. (2003). *High/Scope program quality assessment — Preschool version* (2nd ed.). Ypsilanti, MI: High/Scope Press.

High/Scope Educational Research Foundation. (2005). *Youth program quality assessment.* Ypsilanti, MI: High/Scope Press.

Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher, 23,* 5–12.

Nunally, J. C. (1967). *Psychometric theory.* New York: McGraw-Hill.

Nunally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *The Journal of Consumer Research,* 21(2), 381–391.

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

Appendix A. Interview Transcripts

Site 1

Interview With Local Evaluator (LE)

How have you made the Youth PQA fit your program circumstances?

- Did not want to take time away from the program staff so LE and three graduate students collected the data and scored the Youth PQA.
- Basic feedback was given to the teachers involved in the elementary program.
- Feedback was given to the site coordinator of the middle school program.

Process — How did you collect the data and score the instrument, and who participated?

- LE and one student went to the elementary school site and two students went to the middle school site and collected data over a 2-week period of time totaling 2 hours of observation.
- They met during that time to review the anecdotes and decide what needed to be focused on (what information was missing).
- Data/anecdotes were typed, cut, and pasted by one person and then reviewed by the other.

What worked well in this process?

- Having one person decide where the anecdote fit.

What did you like about it?

- Using outside observers did not take time away from the program staff.

Did data collected or related conversations lead to any changes?

What changed as a result?

- Elementary school site: received feedback about the positive interactions observed and specific items that need improvement (e.g., 911 procedures).
- LE is not certain of changes that may have occurred from this feedback.

What changed as a result (e.g., people's perceptions about things and/or specific actions that you undertook as a result of the data)?

- Middle school site: A major problem was observed with inconsistent discipline and staff supervision and reported to the site coordinator. The coordinator reported that the week observed was not a usual week. He was out of town and was concerned that the Youth PQA was not an appropriate tool because of the small slice of time used to complete the assessment.

People's perceptions about things:

- BUT then the coordinator received an e-mail from another source (not privy to the Youth PQA information) with concerns about the "chaos" observed with the program. He then considered that the Youth PQA information was appropriate and accurate.

Specific actions that you undertook as a result of the data:

- They use a third party to provide the program and feel that the third party isn't hiring qualified individuals; grantee is considering retaining the third party but doing the hiring themselves.

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

Anything else?

Concerns:

- The process was cumbersome — cutting and pasting with scissors and paste. An electronic version would make things easier.
- They spent 2 hours observing and 14 hours organizing and scoring. They felt the 14 hours was excessive time spent analyzing a 2-hour block of time.
    - One of the students created a scaled down version of the Youth PQA in their opinion appropriate for elementary school children using more "general categories":
- Youth PQA appropriate for middle school and high school; suggest another version for elementary.
    - Examples of items not appropriate for elementary: Assisting with governing of group, conflict resolution
- Suggest adding an item specific to the "consistency of discipline from staff"; they observed staff not being consistent with the rules.

Site 2

Interview With the Program Director (PD)

How have you made the Youth PQA fit your program circumstances?

- PD and a paraprofessional attended the Youth PQA training. They serve an area that is larger than the state of Rhode Island, yet there is not much there — a grocery store and a blinking light. Some items did not fit their circumstances (e.g., partnering with others in the community).

Process — How did you collect the data and score the instrument, and who participated?

- PD and the paraprofessional took anecdotes for several weeks (6–8 hours each).
- The paraprofessional then sent her data to PD who then typed up the anecdotes (approximately 2 hours).
- PD, the paraprofessional, and the rest of the program staff members (totaling 8) sat down for lunch and placed the anecdotes in the Youth PQA. Then everyone collaborated on the scoring (approximately 4 hours).
- When PD and the paraprofessional started collecting the anecdotes, there was some concern from the staff but he assured them the assessment was for personal growth and he felt they were receptive to the process.

What worked well in this process?

- The group working together to score the items. PD felt that a lot of enlightenment took place. There were some areas they had never thought about.
- It helped them focus on their goals and objectives.

What did you like about it?

- The enlightenment that the staff experienced. "Oh my gosh! I never thought about that!" Their eyes were opened to some areas they had not thought of. (PD reported that many of the program staff are paid minimum wage and have little education.)

Did data collected or related conversations lead to any changes?

- Conversations were positive.

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

What changed as a result (e.g., people's perceptions about things and/or specific actions that you undertook as a result of the data)?
People's perceptions about things:
• The enlightenment of the staff to the many different areas of a self-evaluation/assessment.

Specific actions that you undertook as a result of the data:
• Talked about some improvements but were waiting for the scores to focus on areas of major improvement.
• Some changes/improvements were processed as they talked through the scoring of the Youth PQA. For example: The need to create an incident report and procedures and their staffing ratio (they were understaffed).

Anything else?
• PD would like an "outside observer" to use the Youth PQA in their program. (Maybe a trained individual from another school district.)
• They would like to use the Youth PQA twice a year: at the midway point and end of the year.
• "Youth PQA simple and straightforward; don't want an assessment to be too long or redundant." (PD)
• High/Scope was great with support through the process.
• They enjoyed the Youth PQA process. It was their goal to be the first in the state to complete the assessment. There was a lot of personal growth for them.

Site 3

Interview With Local Evaluator (LE)
How have you made the Youth PQA fit your program circumstances?
• Grantee has two different programs; elementary and middle school operated by different site coordinators. LE is a consultant and she observed, wrote up, and scored the Youth PQAs for both sites. She then met with the site coordinators to discuss the low scores that fell within the foundation items — items that drove other items.

Process — How did you collect the data and score the instrument, and who participated?
• LE spent about 3 ½ hours observing (total time for both sites) and then 2 hours to write up and score (total time for both sites). She felt that the write-up went quickly — "Best practices are described." She is also very familiar with the PQA.

What worked well in this process?
• She selected a couple of leverage points to address with each site coordinator and framed the follow-up discussions around those key points needing improvement.

What did you like about it?
• LE was able to quickly observe and complete the Youth PQA for each site. She then met alone with the site coordinators (no administrators) so she could be completely forward and honest.

Did data collected or related conversations lead to any changes?
• Good conversations came from the process. "Youth PQA captures what she (LE) has

learned in her gut over 35 years. The Youth PQA gives you the common language and position to have the discussions for improvement. It forces the difficult discussions that administrators usually ignore instead of address (e.g., adult sits in chair and watches youth — no interaction occurs)." (LE)

What changed as a result (e.g., people's perceptions about things and/or specific actions that you undertook as a result of the data)?
• Summer staff will meet to discuss the Youth PQA prior to the start of the program. They will focus on 2–3 troublesome areas, talk about strategies, and allow the staff to work through it themselves — come up with their own solutions. A follow-up meeting will host a discussion of what was tried and what worked.

People's perceptions about things:
• The elementary school site coordinator's perception changed right away. She caught on immediately and made positive and major changes — a major impact on the program. LE believes that the site coordinator will continue to make changes.
• The middle school site coordinator didn't get it as quickly. He wanted to argue about the problems observed. LE has encouraged him to listen and observe and see what interests the youth and work from that point. She isn't sure what changes will truly occur.

Specific actions that you undertook as a result of the data:
• Elementary school program:
• LE observed the elementary program and found "chaos" — youth were sitting and waiting for adults. They had nothing to do, and they were loud and physical. That high level of energy was taken into the program activities.
• She targeted (1) Engagement — there was nothing for them to do while they waited for the adults to start the activities, and (2) Choices — what options did they have other than choosing to annoy each other?
• The elementary school site coordinator purchased materials and created a system for snack and activities. The choices completely turned around the situation. The youth are able to make choices immediately and independently — without waiting for an adult to offer, or start, the process. Because they were actively engaged in activities of their choice, they did not have time to get loud and physical with each other.

• Middle school program:
• LE observed the middle school program and found youth waiting around for the adults to ready themselves for an activity to begin (gym person took 15 minutes to gather equipment) and that "kids creating knowledge and process isn't happening — needs to happen" (LE).
• She targeted the following areas for discussion of improvements: (1) Choices — youth were doing the same things everyday derived from a choice given at the start of the school year program. "They should have choices everyday — content choices, process choices...." (LE), and (2) Staff Interaction — the adult sits in the chair and kids work on the same thing everyday without any interaction — among themselves or with the adult.
• LE had the middle school site coordinator walk through the observation time period with her and then read what she wrote up. She felt it helped him see what needed to be changed — but there has been no evidence of change yet.

Anything else?

• It would be helpful to videotape a class or activity and have the program staff then write anecdotes, score, and discuss.

Site 4

Interview With Program Director (PD)

How have you made the Youth PQA fit your program circumstances?

• PD observed and scored the Youth PQA as an outside observer. She provided initial findings from the use of the Youth PQA to the grant administrator, site coordinator, and principals of the schools. The principals were to follow up with the program staff and/or provider.

Process — How did you collect the data and score the instrument and who participated?

• PD and two middle school administrative staff members collected the data and then scored together. Then, PD completed one on her own (high school), and they completed their own (middle school).

• PD met with the site coordinator and shared general findings. The general findings "matched with his impression of things — Youth PQA looking at the right things — or he was!" (PD)

What worked well in this process?

• The administrators were able to observe the program and learn things they could not have known from hearsay or assumptions. It "helped them with giving them a perspective on program happenings." (PD)

Did data collected or related conversations lead to any changes?

• "Youth PQA opens door for conversations for additional improvements." (PD)

• Conversations with the administrators and site coordinators made them think about doing things better and having the conversation with the program staff about improvements.

What changed as a result (e.g., people's perceptions about things and/or specific actions that you undertook as a result of the data)?

People's perceptions about things:

• The grant administrator and site coordinator thought the Youth PQA was a good measure to use for training of staff. It will help provide a sense of what type of staff they are looking for.

• Some Youth PQA items may be done "instinctively — or not at all — but they (program staff) may not think about them, particularly the traditionally trained people." (PD)

• Youth leadership opportunities — or lack thereof — popped up as the biggest issue. This seems to be related to the more traditionally trained staff.

Specific actions that you undertook as a result of the data:

• The site coordinator and principals used the Youth PQA findings to discuss and review whether the provider used was good for the youth at their schools. They believe that the provider is not up to par.

• They plan to/want to use the Youth PQA to build staff development for the next year.

Anything else?

© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

• The Youth PQA "form" was easy to use — but it was time consuming to observe and score — especially when a follow-up interview was necessary.
•The Organization piece was hard to use. Some questions weren't relevant to the afterschool program — their environment. The questions just didn't relate to their setting — perhaps a different set of questions would be appropriate for afterschool settings?

Site 5

Interview With the Program Director (PD)

How have you made the Youth PQA fit your program circumstances?

• They have embraced their local evaluator and asked them to also embrace and use the Youth PQA. Together they philosophically believe the Youth PQA is getting at what they need to/want to do and report that the changes are what the 21st Century Programs are looking to create in youth's lives.
• As a way of presenting the Youth PQA as a positive, low-key way to evaluate their programs and work toward improvement, PD and four staff members embarked upon a "mock Youth PQA." In January the group met and looked through the Youth PQA and then went back to their sites and collected a couple dozen anecdotes during the month. At the end of the month, they met for a "retreat." The collected anecdotes were posted into items, and then the group collectively scored the mock Youth PQA. Discussions included why the anecdotes fit and how the scoring was processed. This prompted many in-depth conversations and many insights were encouraging. The approach of the Youth PQA gave them a way to talk to program staff about thinking and planning programs — all details of the planning process.

Process — How did you collect the data and score the instrument, and who participated?

• The site staff collected the data, and one staff person placed and scored the anecdotes within the Youth PQA.
• The "mock Youth PQA" process was a terrific learning experience and truly set the stage for staff acceptance of it as a tool for evaluation and improvement.

What worked well in this process?

• At the site, it was easiest to have one person assign and score the anecdotes because it is a problem to get staff together to go over it all.
• The "mock Youth PQA" was a collective effort.

Did data collected or related conversations lead to any changes?

• The local evaluator put together a presentation, and there were many constructive conversations between the local evaluator and staff with regards to strengths and weaknesses; all geared toward improvement.
• PD believes that the Youth PQA "is a useful tool for really understanding their program quality and how to have conversations to improve it."

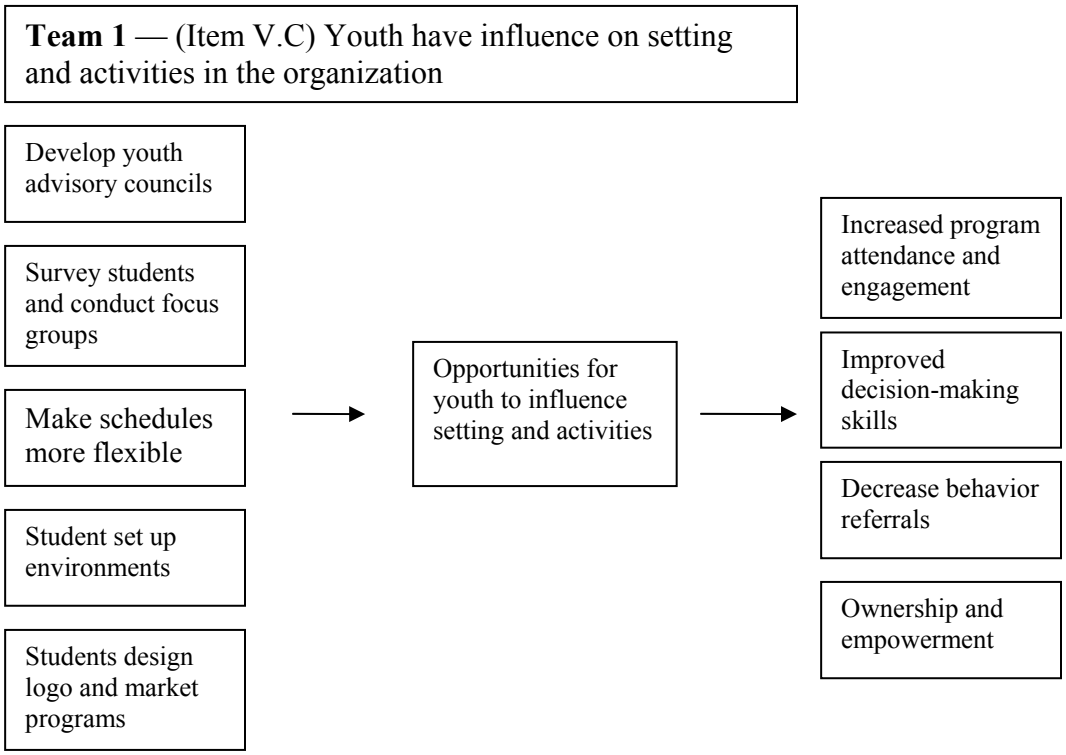
What changed as a result (e.g., people's perceptions about things and/or specific actions that you undertook as a result of the data)?

• Plan to encourage staff to use the Youth PQA as an orientation for new program staff.
• They will "keep it (Youth PQA) in front of us to continue conversations around quality" (PD) and use for the purpose of planning.

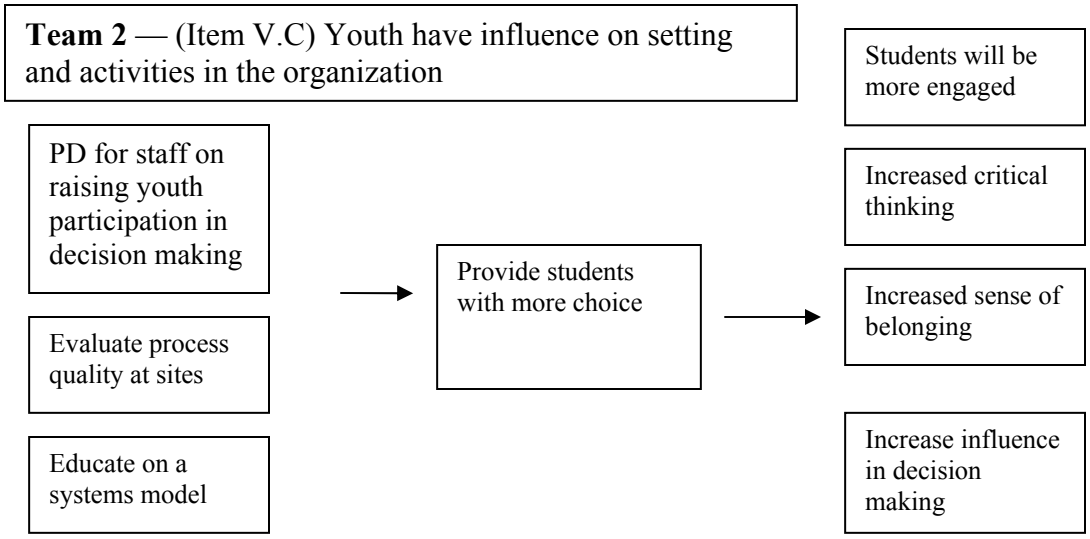
Specific actions that you undertook as a result of the data:
• Plan to incorporate the Youth PQA with the site visits for communication and feedback for local improvement efforts.

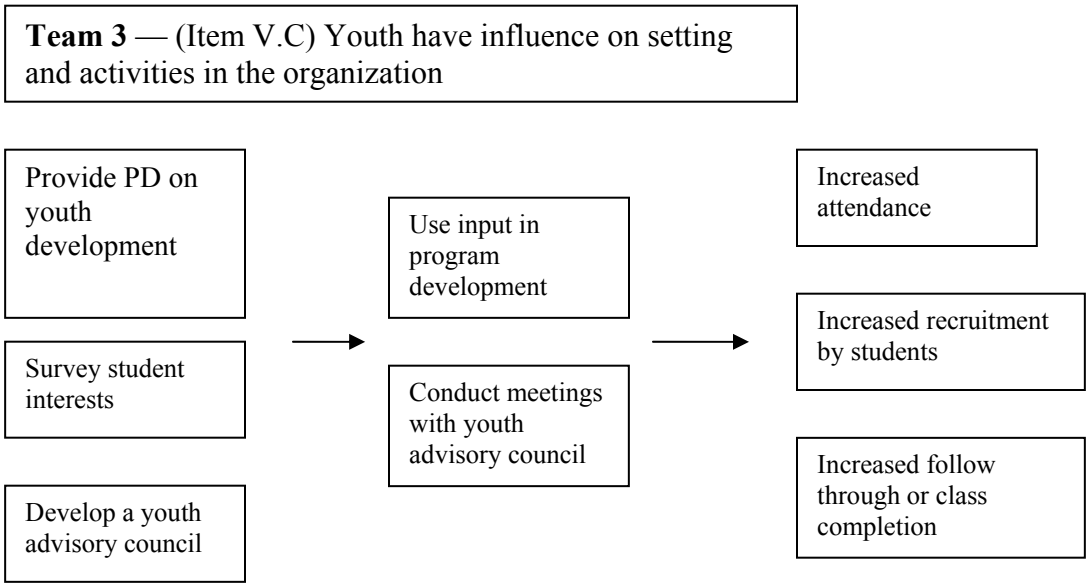
© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org

Appendix B. Program Improvement Models

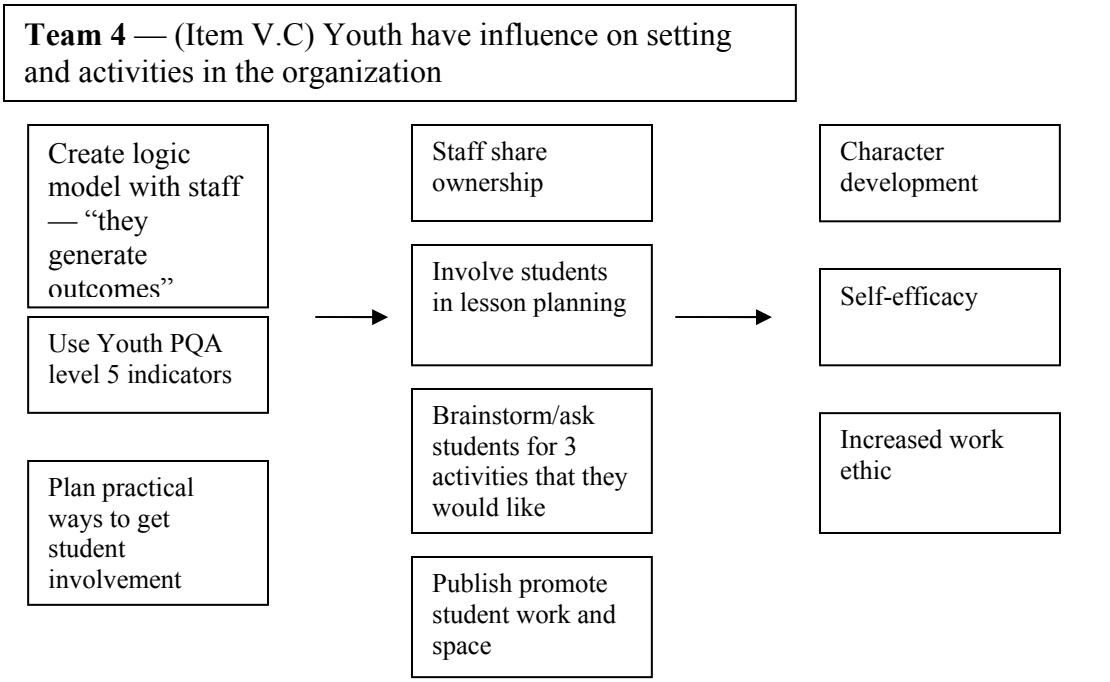


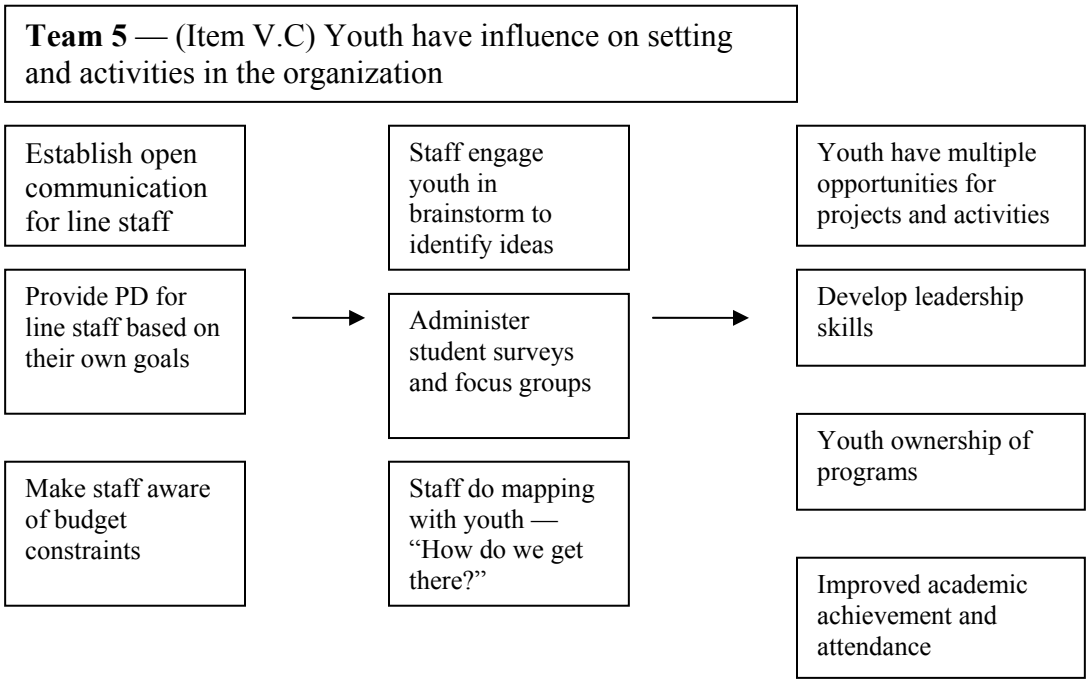
© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org



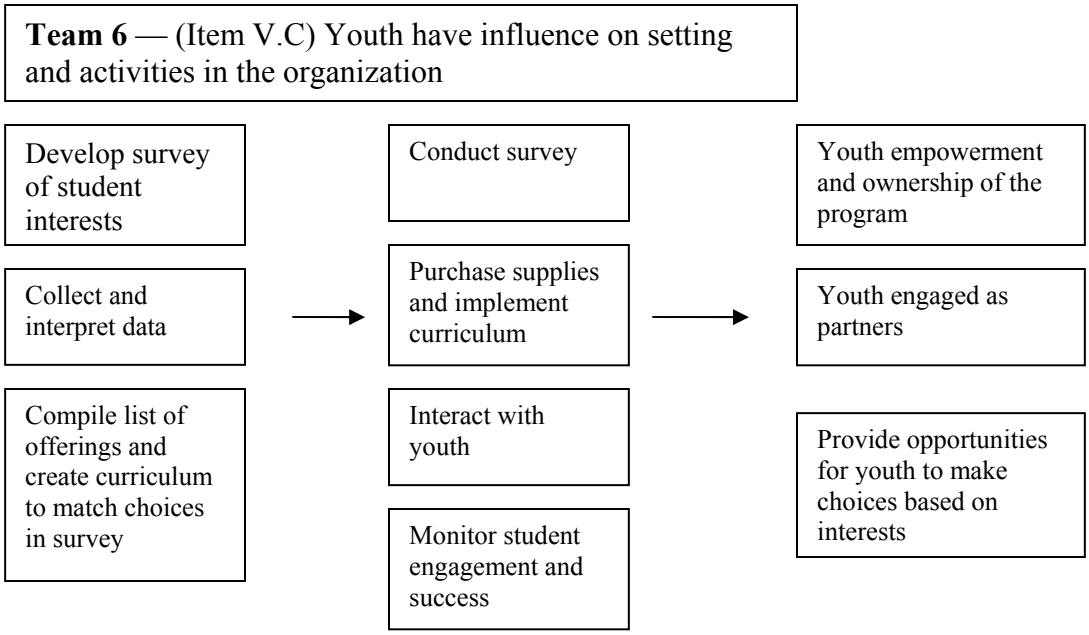


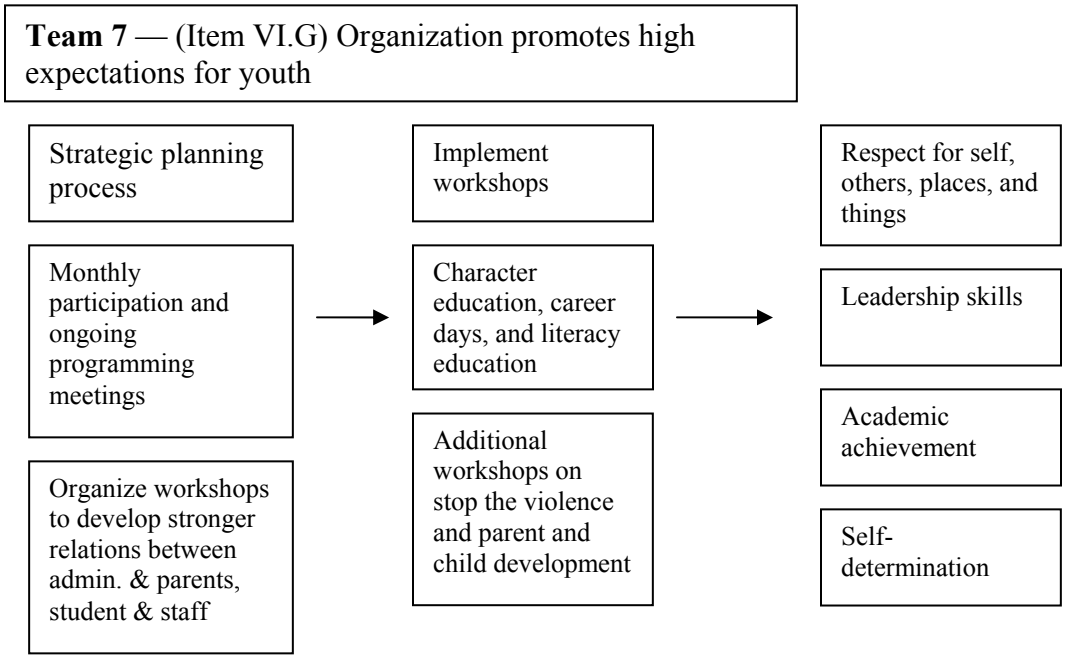
© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org



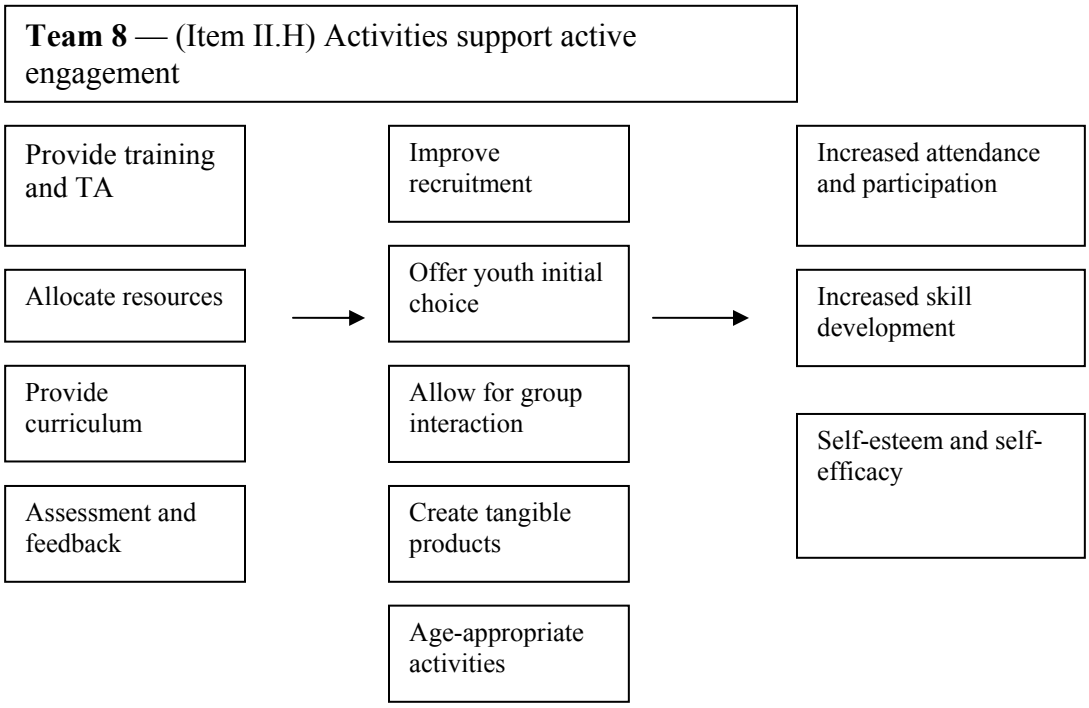


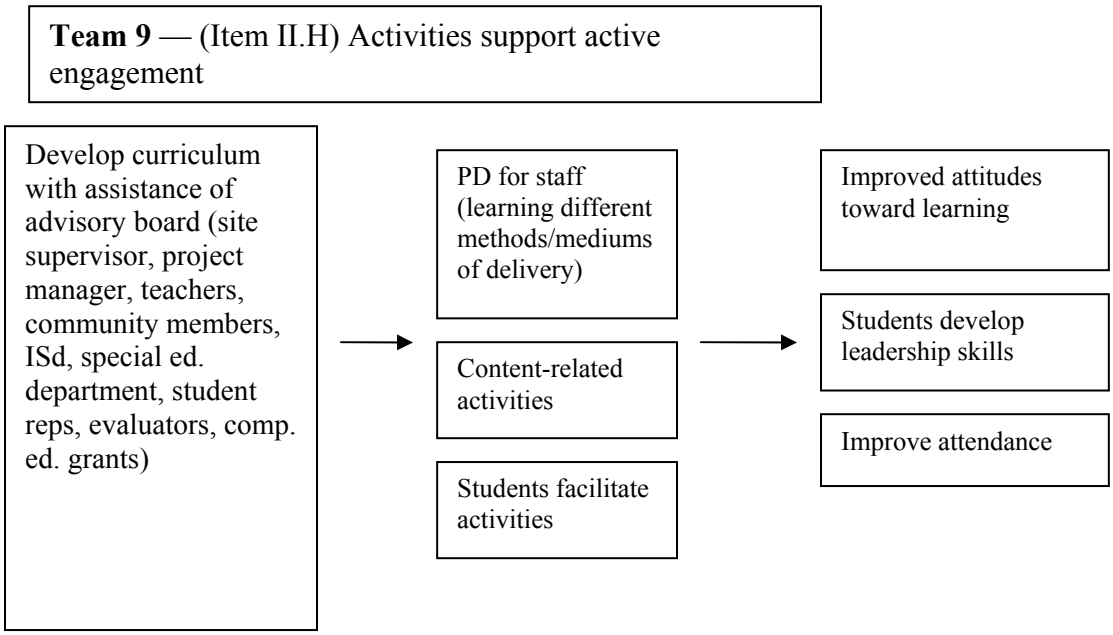
© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org



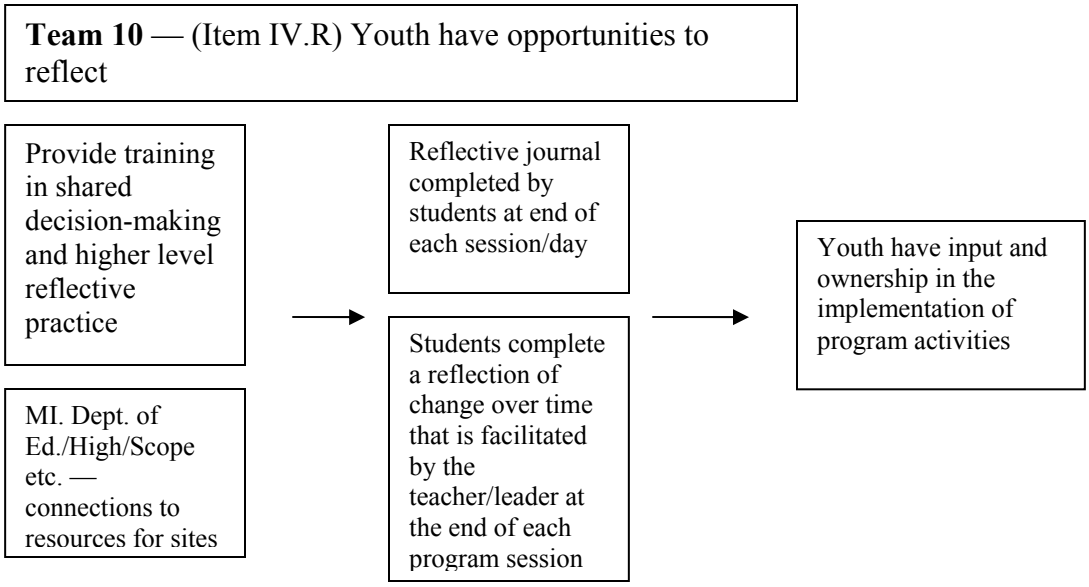


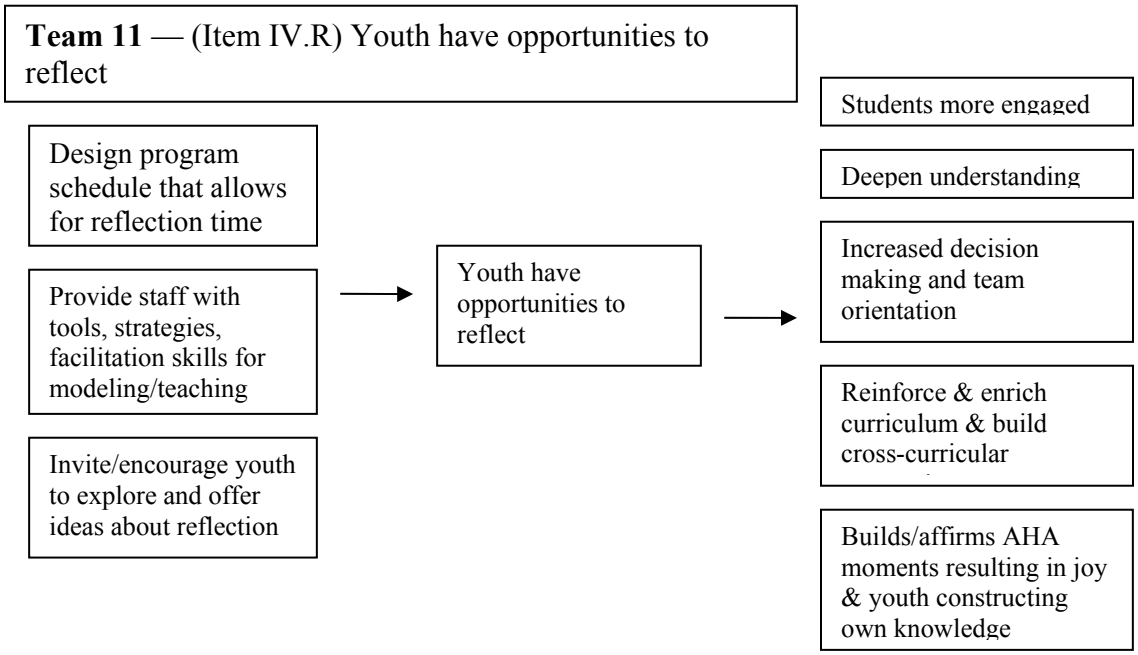
© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org



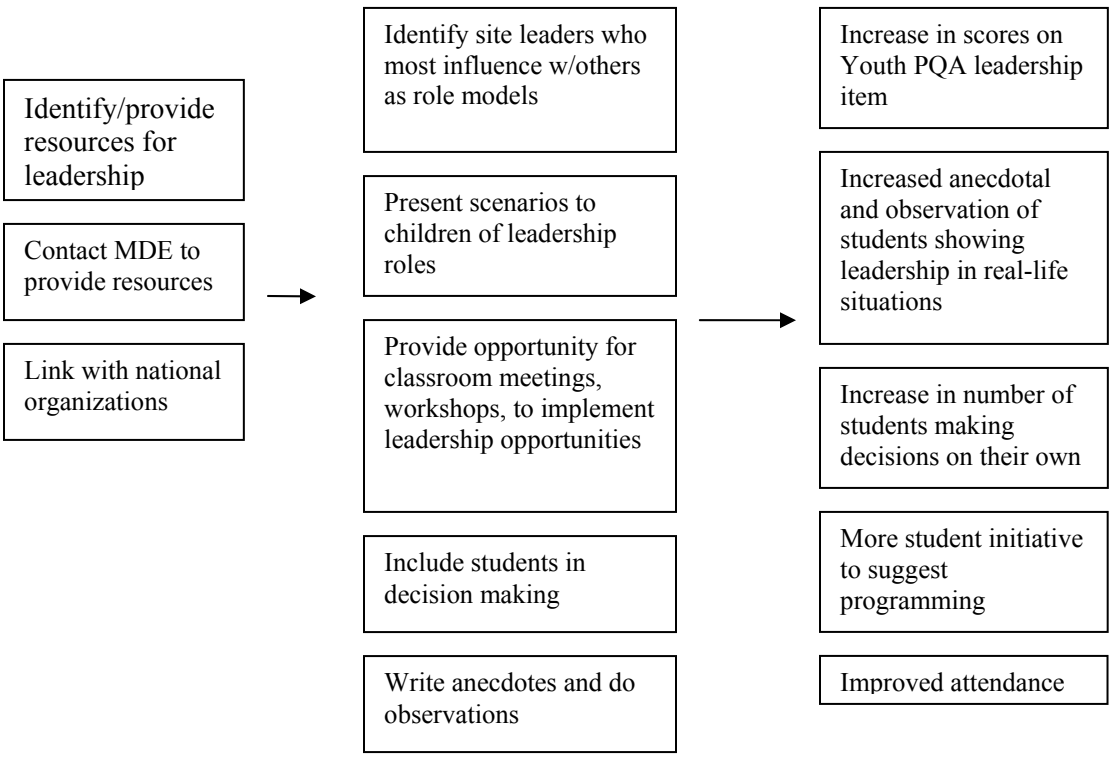


© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org


Team 10 — (Item IV.R) Youth have opportunities to reflect
Provide training in shared decision-making and higher level reflective practice
MI. Dept. of Ed./High/Scope etc. — connections to resources for sites
Reflective journal completed by students at end of each session/day
Students complete a reflection of change over time that is facilitated by the teacher/leader at the end of each program session
Youth have input and ownership in the implementation of program activities


Team 11 — (Item IV.R) Youth have opportunities to reflect
Design program schedule that allows for reflection time
Provide staff with tools, strategies, facilitation skills for modeling/teaching
Invite/encourage youth to explore and offer ideas about reflection
Youth have opportunities to reflect
Students more engaged
Deepen understanding
Increased decision making and team orientation
Reinforce & enrich curriculum & build cross-curricular
Builds/affirms AHA moments resulting in joy & youth constructing own knowledge

Team 12 — (Item II.H) Youth have opportunities to act as group facilitators and mentors



© 2005 High/Scope Educational Research Foundation ▪ youth.highscope.org