Assessing Program Strategies and Priority Outcomes: Evaluating Both Sides of Program Models

by:

Kathryn E. H. Race
Race & Associates, Ltd.
4430 N. Winchester Avenue
Chicago, IL 60640
(773) 878-8535
www.raceassociates.com

Paper presented at the Annual Meeting of the
American Evaluation Association
Denver, CO

November 2008

Assessing Program Strategies and Priority Outcomes:
Evaluating Both Sides of Program Models

Kathryn E. H. Race
Race & Associates, Ltd.

Through exemplar case studies from evaluation practice, this paper discusses ways in which program strategies can be assessed descriptively by rubric development and other assessment tools when sample size prohibits the use of quantitative methods such as structural equation modeling. Examples are taken from education evaluations: one in an informal setting involving a 3-year multi-phased, hybrid after-school and family outreach program; and the other from a formal setting involving a math and science partnership in physical science endorsement of middle-school teachers. The gathering of empirical evidence to determine the degree to which the implemented program aligns with the designed program is highlighted. The paper focuses, however, on using program fidelity data as an integral part of an outcomes evaluation as a means of measuring the strength of intervention and program "dosage" assessment. Implications for the application of this approach in other evaluation venues are discussed as well.

Theory-driven program evaluation is a relatively new effort that serves to build upon knowledge acquired by the practice of evaluation (Donaldson, 2003). By grounding the program and evaluation in a model that is theory based, these efforts extend beyond a simple logic model. Program models or theory of change models can help program stakeholders create a concise representation of the program, use core program strategies based on program goals backed by empirical evidence and/or social science theory, and identify priority outcomes that logically link to these articulated core program strategies (Rossi, Lipsey & Freeman, 2004; Race & Brett, 2004).

A well-articulated program model is cornerstone to an outcomes-based evaluation for several reasons. One important reason is that an articulated program model can help identify core program components, clarify ambiguities in program concepts, and reduce gaps in program planning and structure. This clarity is especially meaningful when the program is complex, the logical links to program outcomes may not be obvious, and/or when multiple stakeholders through, for example, program partnerships have differing perspectives regarding program concepts, underlying assumptions, and outcomes.

Thus, the intent of a program model is to accurately represent that which is being evaluated through captured core program strategies; specify expected levels of program participation; identify priority outcomes that logical link to program components; articulate underlying assumptions and values; and, acknowledge the collaborative context in which the program may operate. The success of a program model is judged by numerous criteria. Among the most fundamental is the degree to which the model is representative of the program and the degree of buy-in of its merit and relevance by key program stakeholders. Such is often accomplished through iterative steps (starting with backward thinking from outcomes to strategies) until consensus or at least agreement is reached by program principals. To strengthen the program model, explicit statements of core strategies, core values and underlying assumptions associated with the program can permit an assessment of the degree to which the program (and its model) reflects best practices that are based on empirical evidence or social science theory literature. And once established, the program model can help guide modifications as program changes are identified.

These efforts are consistent with program models or theory of change models that link specific program strategies and activities to program and long-term outcomes (Rogers, Hasci, Petronsino & Huebner, 2000; Race & Brett, 2004; Renger, 2006). Such models serve to conceptualize a program relative to its operation, the logic that connects its activities to the intended outcomes, and the rationale for why the program does what it does (Rossi, Lipsey & Freeman, 2004).

Relative to the evaluation, a program model can help assess how well the implemented program aligns with the program as designed. Also, once established the program model can help facilitate the use of program assessment information -- as a means to determine program fidelity-- as an integral part of the planned outcomes assessment. Ultimately, the program model can serve to direct the evaluation in its assessment of how well the program meets expected, priority outcomes but does so with a thorough understanding of the services administered.

The purpose of this paper is to explore the importance and value of evaluating a program through assessing how core program strategies are implemented (relative to the designed program) and linking these assessments to subsequent analyses of outcomes. We will illustrate this with two exemplar case studies pulled from evaluation practice where stakeholders had a high level of involvement in the development of their program model, which resulted in a high level of buy-in in its use. After a brief overview of each example, we will describe the methods used – one based on a descriptive analysis of program staff generated assessments and the second based on the development and use of a rubric that was specifically created to meet assessment needs. We conclude with a discussion of the implications for the application of this approach in other evaluations.

Overview of Exemplar Case Studies

Structured Program Feedback

The first example stems from a 3-year evaluation of a multi-phased, hybrid after-school and family outreach program (Race, 2007). All developed evaluation measures came directly from the program model, where a specific subset of these measures focused on program fidelity assessment. (See Figure 1 for a simplified version of this program model.) Framing evaluation as a continuous management tool rather than a discrete evaluation study (Scheirer, 2005), data were gathered as an integral part of the services provided by the program. An overview of the program operation and evaluation measurement tools used in this evaluation (see Table A-1); and the program strategies,

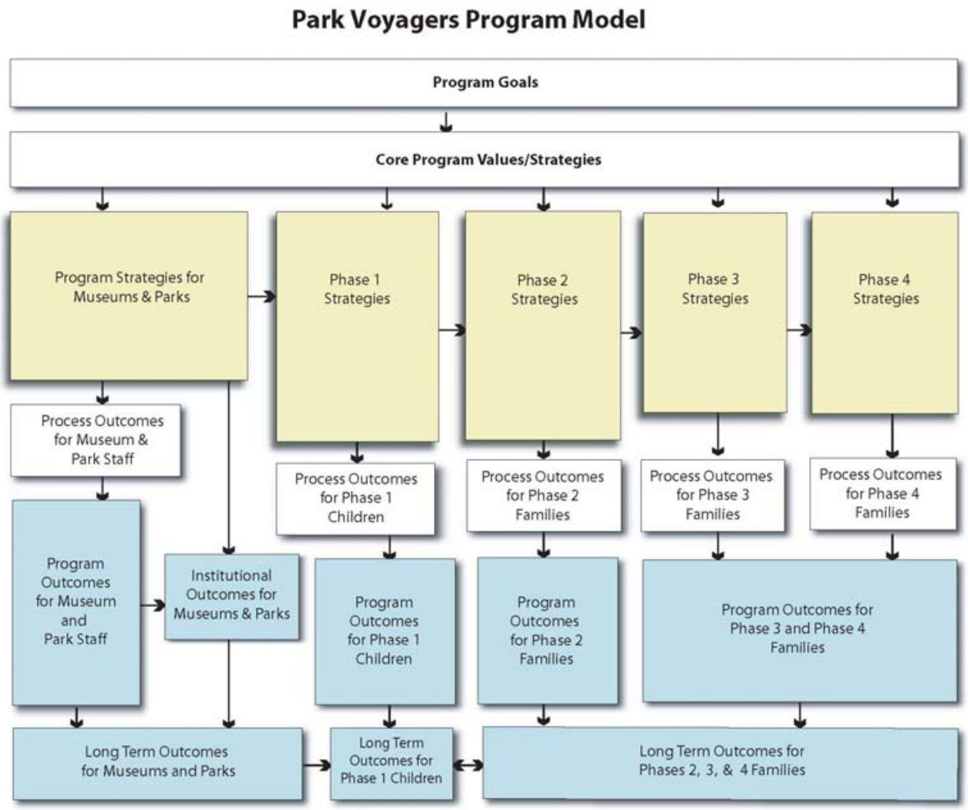

Figure 1. Abbreviated program model of a hybrid after-school, family outreach program named, *Park Voyagers.* (Race, 2007)

process outcomes, and program outcomes measured (see Table A-2) are shown in Appendix A. Most germane to the present discussion are the feedback forms, which the evaluator specifically designed with items selected from the program model, such as best practices used, level of participation, and other program attributes. The individual educator or group of educators who conducted a particular program session completed these forms. The assessments were based on separate implementations of the program across various local settings and analyses were based on the assessments of the educators as a group. In this way, core strategies implemented during phase 1, and 2 of the program were assessed. Progress outcomes were assessed for children (phase 1) and families (phase 2) also based on this methodology, supplemented with registration and attendance information. Program outcomes were assessed using an array of evaluation measures including data from these feedback forms, registration and attendance information, as well as family surveys and the shadowing of participating families during museum visits using a written protocol.

What we learned from the assessment of core program strategies was that, in general and with some exceptions, the program was implemented with high fidelity as compared to the designed program. However, the assessment of process outcomes suggested that program exposure varied considerably for participating children and families within and across local settings. Importantly based on actual participation, the pairing of a session or workshop (structured activity before a visit) with a museum visit (opportunity to see and explore concepts presented in these sessions/workshops) was less evident across participating families. This suggested that the program as experienced was quite different than the designed program; and, for many families the connections between a session or workshop and a follow-up museum visit may not have been made. This had direct implications for our outcomes assessment as these session/workshop-museum visit connections were seen as integral to the program and only a subset of families experienced the program in this way. Although overall sample size did not permit analyses that teased out these experiential differences, in other evaluations this may be a possibility. Thus, it is likely that different outcomes can be expected from differential program exposure and should be tested when possible.

The phase 1 after-school portion of the program targeted toward children was envisioned by stakeholders as a vehicle to draw families into phase 2 (although not a prerequisite). What we learned from analysis of attendance data was that results were mixed with only about a third to one half of phase 1 families participating in phase 2, with these findings varying across local settings. Although we learned about phase 1-phase 2 program participation levels, we were unable to explore if such participation levels linked to any antecedent conditions such as family income level, demographics, or prior history of positive museum visits with their families.

Had we had sufficient sample size (and sufficient information at the individual family level), we could have divided families into groups by program fidelity (e.g., high versus low) with comparisons made across performance measures taking into account variations in overall program attendance. Thus, we would have been able to include a measure of program quality (session/workshop-museum visit pairings) and a measure of program exposure (program participation by number of hours and/or phases) in our assessment of outcomes. Although these comparisons are not without limitations, such analyses might shed light on what level of intervention is sufficient to have an effect on priority outcomes (Race, 1990).

Rubric Development and Use

The second example comes from a math and science partnership involving an endorsement program for middle-school teachers in physical science culminating in a masters degree for participants. Members of the planning team and the evaluator developed a detailed, complex program model, a simplified version of which is shown in Figure 2. To develop the rubric (which is still under development but in good working order), members of the planning team selected a subset of key program strategies (3 to 5) thought to be essential for each of four main program areas: specialized content knowledge, pedagogical content knowledge development, authentic research experience, and social learning. We reasoned that including all program strategies within the rubric at any given time would result in a task that would be quite time consuming to complete. We also reasoned that across the program, core program strategies could be rotated out of the rubric and new ones rotated in if a given strategy demonstrated sufficient stability

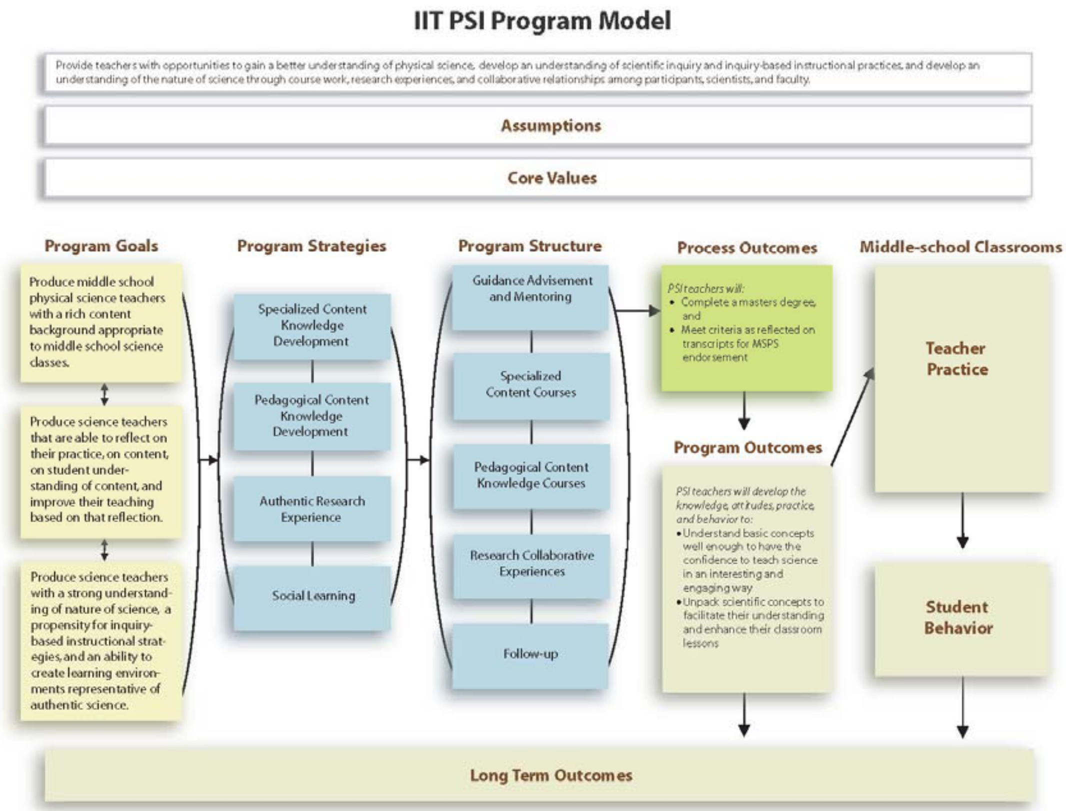

Figure 2. Abbreviated program model for Illinois Institute of Technology's Physical Science Initiative. Used with permission (Meyer, Fluet, Race, Zawojewski & Newman, 2007).

over time. For many of these benchmarks (i.e., program components or performance markers by teachers), a 3-point rating scale was used to make the assessment (3= Exceptional, 2 = Accomplished; and 1 = Not Accomplished). Others were assessed more simply using an "either or" dichotomized choice. In addition, a brief statement of what method used (or could be used) to make these judgments was noted as well. And in its final stage, the educational research and evaluation literature will be cited in support of each core program strategy. This rubric is shown in Appendix B and is used with permission.

The overall goal was to develop a rubric that was relevant and meaningful, reasonable in its completion requirements, and ultimately useful in guiding decisions on program changes. We see the rubric as an assessment of the "overall quality" of the implemented program. This is indeed consistent with other efforts used to enhance the quality of science and mathematics programs (Kaser, Bourexis, Loucks-Horsley & Raizen, 1999).

It is a bit challenging to note lessons learned at this early stage as the first use of the rubric is expected to occur late in the first semester in which the program has been implemented. One notable lesson, however, is that this effort illustrates the active use of the program model by members of the planning team. Importantly it will serve to guide program decisions made in the future. How will we use the results of this rubric review in the outcomes-based evaluation? The simple answer at this stage is that it should guide outcomes assessment too, in that when program components have been executed well, concomitant performance by participating teachers should be evident. If performance problems arise in these situations, it might suggest that the target audience and program are out of alignment and may need to be re-evaluated. If a program component is not executed well, then related outcomes should not be expected. And finally, analyses may provide the opportunity to explore performance differences based on differential program exposure and program execution.

Implications

As a final point of discussion, we compared examples across several attributes to help underscore the usefulness of this approach (or similar approaches) in future evaluation opportunities. This comparison is summarized in Table 1.

One strength of either one of these or similar methods is that it encourages thinking about "unpacking" the exposure to the program variable. Program variations are real phenomenon and there is merit in viewing program exposure as a dynamic process. Programs are not static monolithic interventions administered equally to all participants based on quality or frequency. And too, although not discussed here program impact may be moderated by other factors (such as, prior history of family visits to museums that were positive, or prior exposure with reform-based pedagogical instructional strategies). Our ability to analyze findings in a context that reflects the complexity of the program will in the long run help us make better judgments about program impact.

One challenge of either one of these or similar methods is determining how to use such information in subsequent outcomes analyses. It is this exploration that this paper directed its focus. Can program fidelity data be incorporated explicitly in the outcomes analysis or is one left to speculate on the implications of the program in discussion of conclusions? Using such data explicitly in analysis can be especially problematic when data are largely descriptively in nature, say where a large number of data points are evident. Which data are relevant? Can this information be useful if aggregated in some way? What happens when the summaries are more qualitative in nature and do not lend themselves well to quantification? Is even the ability to speculate about differential program effects due to differential exposure worth the effort in an attempt to understand what services are delivered, in what ways, and to whom? This paper has argued in favor of a "yes" to this question.

It seems critical that careful thought be given to measures of program fidelity to ensure that there is a true core standard with which the implemented program can be described and measured. Extending this thinking to outcomes assessment seems to have special relevance. And although alternatives that rely on sophisticated statistical approaches such as structural equation modeling or mixed-level models have merit, Donaldson

Table 1
Comparison of Examples by Select Attributes

Attribute	Structured Program Feedback	Rubric Development and Use
Methodology	Information based on multiple program staff members across multiple implementations.	Core program strategies are systematically reviewed by those most familiar with the program.
Use of Program Model	Feedback forms created from program model.	Rubric created from program model.
Role played by Program Staff/Planning Team	▪ Program staff responsible for completing feedback forms. ▪ Program Director had oversight responsibility to insure compliance. ▪ Program staff reviewed forms and recommended changes.	▪ Provided rubric structure input. ▪ Created rubric. ▪ Development guided by criteria of relevance, reasonableness, and usefulness.
Role of Evaluator	▪ Responsible for the development of structured feedback forms. ▪ Used information in analyses (see below).	▪ Served as coach (e.g., gave team idea; provided input on rubric structure; helped create guideline criteria).
Use/Anticipated Use of Information by Program Stakeholders	▪ Process helped clarify needed program changes (e.g., connecting sessions/ workshops to strength of the individual museums).	▪ Will provide empirical evidence of program quality. ▪ Will help guide program decisions and changes.
Use/Anticipated Use of Information by Evaluator	Used information in program fidelity and priority outcome assessments.	Will review information and may guide recommendations for program outcome assessment.

(2001) asserts that gathering some data, through focus groups, theme studies, interviews, surveys and other data sources can be useful and that even limited data are better than no data.

References

Donaldson, S. I. (2001). Mediator and moderator analysis in program development. In, S. Sussman, (Ed)., *Handbook of program development for health behavior research and practice* (pp. 470-496). Newbury Park, CA: Sage.

Donaldson, S. I. (2003). Theory-driven program evaluation in the new millennium. In S. I. Donaldson & M. Scriven, (Eds.), *Evaluating social programs and problems: Visions for the new millennium (*pp. 109-141). Mahwah, NJ: Lawrence Erlbaum.

Fluet, K., Meyer, D., Zawojewski, J., & Newman, W. (2007). I*llinois Institute of Technology's Physical Science Initiative program model evaluation rubric.* Chicago: Illinois Institute of Technology, Mathematics and Science Education Department.

Meyer, D., Fluet, K., Race, K. Zawojewski, J., & Newman, W. (2007). I*llinois Institute of Technology's Physical Science Initiative program model.* Chicago: Illinois Institute of Technology, Mathematics and Science Education Department.

Kaser, J. S., Bourexis, P. S., Loucks-Horsley, S., & Raizen, S. A. (1999). Enhancing program quality in science and mathematics. Thousands Oaks, CA: Corwin.

Race, K. E., H. (1990). Treatment fidelity among evaluated and non-evaluated workplace Safety-belt program. *Evaluation Review, 14*(3), 290-296

Race, K., E., H., & Brett, B. (2007, August). An outcomes-based evaluation of Museums In the Park, *Park Voyagers* program: Final report. Chicago: Race & Associates, Ltd.

Race, K., & Brett, B. (2004, July). *Evaluation of the Cristo Rey network schools.* Chicago: Race & Associates, Ltd.

Renger, R. (2006). Consequences to federal programs when the logic-modeling process is not followed with fidelity. *American Journal of Evaluation, 27,* 452- 463.

Rogers, P. J., Petronsino, A., Huebner, T.A., & Hasci, T. A. (2000). Program theory: Practice, Promise and problems. In, P.J. Rogers, T.A. Hasci, A. Petronsino. *Program Theory in Evaluation: Challenges and Opportunities. New Directions for Evaluation, 87,* 5-13.

Rossi, P H., Lispey, M. W., & Freeman, H. E. (2004) *Evaluation: A systematic approach.* (7th ed.). Thousand Oaks, CA: Sage.

Scheirer, M.A. (2005, November). *Improving the use of evaluation in human service agencies.* Panel presentation at the meeting of the American Evaluation Association, Toronto.

Table A-1
Program Operation and Evaluation Measurement Tools
by *Park Voyagers* Program Phase

<table>
<tr><th rowspan="2">Program Phase</th><th colspan="2">Program Operation</th><th colspan="2">Evaluation</th></tr>
<tr><th>Form/Tool</th><th>Description</th><th>Form/Tool</th><th>Description</th></tr>
<tr><td>Phase 1</td><td>• Registration Form
• Attendance Record
• Demographic Survey
• Field Notes</td><td>Registration form and demographic survey available in English and Spanish. Demographic survey completed by one adult per family. Field notes completed by museum educators and park staff.</td><td>• Phase 1 Feedback Form including Museum Visit</td><td>Completed by a museum educator for each lesson and accompanying museum visit as appropriate. Supported with Instructions and Operational Definitions.</td></tr>
<tr><td>Phase 2</td><td>• Registration Form
• Attendance Record
• Demographic Survey
• Field Notes</td><td>Registration form and demographic survey available in English and Spanish.

Demographic survey completed by one adult per family.

Field notes completed by museum educators and park staff.</td><td>• Phase 2 Feedback Form including Museum Visit

• Family Exhibits Assessment

• Phase 2 Family Survey</td><td>Completed by museum team for each family workshop and accompanying museum visit as appropriate. Supported with Instructions and Operational Definitions.

Ratings of exhibits by families based on five criteria; rated by program staff/evaluator.

Survey completed by one adult per family. Available in English and Spanish.</td></tr>
<tr><td>Phase 3</td><td>• Attendance Record</td><td rowspan="2">Families sign in at Park Voyagers table at main entrance to museum.

Based on successful completion of the program.</td><td rowspan="2">• Phase 3 Family Survey

• Phase 3 Museum Visits</td><td rowspan="2">Survey completed by one adult per family. Available in English and Spanish.

Shadow participating families during museum visit based on observation by evaluator(s) using written protocol.</td></tr>
<tr><td>Phase 4</td><td>• Number of passes distributed to families</td></tr>
</table>

Table A-2
Core Strategies, Process and Priority Program Outcomes

Outcomes	Museum & Park Staff	Phase 1 Children	Phase 2 Families	Phase 3 Families
Core Strategy	▪ Not measured.	▪ Familiar park setting is intended to maximize children's participation. ▪ Children are encouraged and given freedom to explore. ▪ Activities provide hands-on, multi-sensory experiences. ▪ Museum artifacts and visits intended to help children connect the conceptual and the tangible.	▪ Familiar and convenient park setting is intended to maximize family participation. ▪ All family members invited. ▪ Families learn together in new settings, including museums. ▪ Dinner and transportation provided. ▪ Workshops provide planning and navigation tools and strategies for making museum visits instructive and fun. ▪ Museum staff model inquiry and museum going behavior.	▪ Not measured.
Process	▪ Program runs efficiently. ▪ Implement program as intended.	▪ Are engaged and challenged. ▪ Are having fun. ▪ Attend sessions regularly. ▪ Participates in field trips.	▪ Are engaged and working together. ▪ Are exposed to a range of museum resources. ▪ Are having fun. ▪ Attend sessions regularly. ▪ Participates in field trips.	▪ Families participate in program sponsored museum events.
Program Priority	▪ Form meaning ful professional network outside of their institutions and see each other as resources.[a]	▪ Seek out museum resources. ▪ Show curiosity through greater questioning, observation, and exploration. ▪ Share excitement of exploration with their families. ▪ Draw families into Phase 2.	▪ Recognize value of going to museums. ▪ Are aware of what museums have to offer. ▪ Demonstrate skills, confidence, and interest in using museums as educational and recreational resources and in facilitating children's visits.	▪ Make greater and more engaged use of museums as educational/recreational resources. ▪ Demonstrate increased independence when inquiring about and using museum resources.

[a]The emphasis of this evaluation was directed toward participant outcomes. Therefore, this outcome was not measured.

Appendix B

IIT PSI Program Model Evaluation Rubric

Core Program Strategies	Benchmark	Rating			Method
		3: Exceptional	2: Accomplished	1: Not Accomplished	
Specialized Content Knowledge Development ▪ Specialized content courses address relevant state and national science teaching standards ▪ Content course pedagogy aligns with reformed-based pedagogical instructional strategies (e.g., concept-based focus; connections among concepts rather than collections of isolated facts) □ Investigations are inquiry-based rather than confirmatory laboratory exercises □ Focus on science knowledge beyond fact learning techniques/skill building or symbolic manipulation • Concept re-sequencing designed to highlight and strengthen connections among concepts	Expert identified relevant teaching standards are reflected in courses	At 100% level.	Greater than or equal to 80% of the time.	Less than 80% of the time.	Project staff identifies the appropriate standards and maps them onto the courses in the program.
	Summary Rating based on Horizon's Classroom Obser-vation Protocol	Level of 5 (accompanied by ongoing formative feedback)	Level of 4 (accompanied by ongoing formative feedback)	Level of 3 or less (accompanied by ongoing formative feedback)	Science educators observe content course delivery 3 times per semester. Summary rating based on Horizon's tool.
	X% of identified re-sequenced curriculum is for the purpose of strengthening connections	70%	50%	Less than 50%	Project staff identifies what specifically was re-sequenced and provides rationale.
Pedagogical Content Knowledge Development • Pedagogical discussion of nature of science through reflection on content topics and research collaborative experience • Examination and critique of middle-school pedagogical content knowledge designed to help: □ Teachers critically revisit how they teach science (e.g., content thinking about students, typical misconceptions, what understanding looks like) □ Teachers to be students of science • Teachers have time, space, and unstructured engagement with science and science ideas • Development of heuristics for teaching unfamiliar topics	For each of the goals, X% of the relevant course experiences are accomplished	90%	80%	Less than the 80%	Project staff identifies those experiences in the courses that address each of these two objectives. Review each of the two lists for a count each semester.

IIT PSI Program Model Evaluation Rubric (con't.)

Core Program Strategies	Benchmark	Ratings			Method
		3: Exceptional	**2: Accomplished**	**1: Not Accomplished**	
Authentic Research Experience 1. Immersive laboratory experiences directed toward application/use in the classroom 2. Teachers to participate in the social milieu of the laboratory (e.g., building collaborative relationships, participation in laboratory teams) 3. Teachers to engage in laboratory activities with scientists with opportunities for reflection, especially on nature of science issues 4. Laboratory experience extending over the course of program participation 5. Experience culminating in an instructional plan for the middle-school classroom.	X% of students have journals reflecting inclusion into the activity of the lab	90%	80%	Less than 80%	Reflective journals
	X% of students have journals reflecting interest in transferring lab experience into classroom	90%	80%	Less than 80%	Project records of teacher/research placements
		--	100 % placement	Less than 100% placement	Teacher time logs
	X% of the teachers attend 66 hours per year of each session	25% attend *more than* 66 hours	90%- 100% participate 66 hours; and, *up to* 25% of the teachers attend more than 66 hours	Less than 90% of teachers attend 66 hours	
		--	Research Practicum experiences are distributed over 2 years	Research Practicum experiences are not distributed over 2 years	Course catalog, syllabus Passing course grade
	X% of the teachers complete classroom-based projects	100%	90%	Less than 90%	Project staff identifies those experiences in the courses that address each objective. Review each of the lists for a count each semester.
	X% of planned experiences that require reflection on research practicum are implemented.	90%	80%	Less than 80%	

IIT PSI Program Model Evaluation Rubric (con't.)

Core Program Strategies	Benchmark	Ratings			Method
		3: Exceptional	**2: Accomplished**	**1: Not Accomplished**	
Social Learning • Teachers to have an MSED advisor and an IIT research mentor • Development of ample and sufficient communication among science educators, research faculty and teachers • Teachers to support one another and facilitate the exchange of innovative ideas • Peer Observation and Action Research are designed to help teachers to: ☐ Create lessons with appropriately articulated objectives, especially as to the expectations of student understanding And support opportunities for: ☐ Peers to share, critique, and reflect upon implemented lessons	Teachers are assigned an advisor and mentor	--	All	Not all	Project Research
	A record of meetings and interpersonal communication …	--	is organized and available	is not organized or is not available	Agendas, email file minutes of meetings
	X% of the planned experiences that support teacher interactions for the purposes of classroom planning	90%	80%	Less than 80%	Project staff identifies those experiences (e.g., peer feedback/observation, group work) in the courses that address each objective. Review each of the lists for a count each semester