INTEGRATED HUMAN SERVICE MODELS:
ASSESSING IMPLEMENTATION FIDELITY THROUGH THE 'LINE OF SIGHT' PERSPECTIVE

Working Draft

Tom Corbett & Jennifer L. Noyes[1]

July 23, 2006

There is an old saying that we don't really understand something if we cannot communicate it to others. In our work with state and local sites pursuing systems integration, we have found this axiom holds true. Proponents of human services reform often believe they know what they are trying to accomplish through systems integration. However, when asked to tell a coherent story about their reform model to an outsider, they often stumble in the telling or the narrative varies dramatically across local stakeholders, even when each purportedly attempts to convey the same vision.

The underlying rationale for developing a coherent narrative about what is being done and why it is being done is simple. These models are complex, dynamic, and subject to idiosyncratic interpretation in respect to what they are all about. Therefore, unlike other reform efforts, the quality of implementation is of paramount importance and in fact may compete in importance with the quality of thought that goes into the development of the program model being introduced. Ultimate success of a systems integration effort, we believe, depends very much on the fidelity with which new protocols, practices, and policies are executed on a daily basis.[2]

Given this, it is hard to imagine that substantive change can occur in a policy environment where a common vision has not been fully articulated. It is also hard to imagine that the effectiveness of a model (project 'impacts') can be rigorously evaluated with any confidence absent empirical evidence that the model has, in fact, been implemented as designed. Finally, it is unlikely that an arguably effective model can be replicated unless a detailed understanding of what is happening, or should be happening, on the front-lines is well understood and documented.

Thus, this paper pushes us to consider the importance of formative research (or, how well a model is put into practice) to the pursuit of systems integration. Of particular concern is implementation fidelity, or the faithfulness with which a concept is executed in practice. In doing so, we address two fundamental challenges: 1) getting all stakeholders to have the same understanding with respect to what a given reform model is all about and how it is supposed to function and 2) applying simple formative research techniques toward improving the quality (fidelity) with which reforms are introduced and sustained.

[2] The critical importance of implementation to policy success is well established, see for example The Implementation Game by Eugene Bardach, Cambridge MA: MIT press (1977), and Policy Implementation and Bureaucracy by Randall Ripley and Grace Franklin, Chicago I: The Dorsey press (1982).

Balancing the Formative and Summative Evaluation Agendas

Members of the knowledge consuming fraternity (policy entrepreneurs) and the knowledge producing community (evaluators and researchers) want to know whether or not systems integration, or at least their version of it, works.[3] That is, they first ask the summative evaluation question—whether integration makes a difference in customer outcomes over what they would have evidenced when exposed to a counterfactual service delivery environment.

We have been quite taken with the breadth of interest in this core evaluation question. Virtually every site we have visited around the country wants more information about the effectiveness of its model, or of integrated service delivery models more generally.[4] Few if any, however, know quite how to go about obtaining this evidence with any rigor.

This degree of interest perhaps should not be surprising. After all, pursuing systems integration is hard work that challenges traditional processes and practices. Skeptics want proof that it is worth the investments necessary and proponents want evidence to move their agendas forward. However, answering the conventional impact question is not easy, as discussed elsewhere.[5] In some cases, it may not be practical, even if feasible.

Rigorous protocols for assessing the net effects, or impacts, of integrated service models require clarity about the character of an intervention. Because the structure of system integration initiatives varies with their objectives (e.g., to support working families; to prepare vulnerable individuals for work; to remedy and/or alleviate severe disabilities such as depression, substance abuse, or mental retardation) as well as a host of idiosyncratic local factors, it is first necessary to arrive at a clear consensus about the purposes and structure of a given model.

However, several points of confusion come into play that complicate our ability to arrive at this clear understanding and, by extension, challenge our ability to measure client impacts:

> *Confusion about populations* - How do we define the population of interest when traditional client groupings are merged and we may not really know who is being served, or when those who determine eligibility and route clients to the appropriate service (the 'gatekeepers') do not necessarily have specialized training to make complex eligibility decisions across several programs, or when the policy intent is to minimize distinctions within the program and the target population?

> *Confusion about program boundaries* - Integrated systems are designed to blur the boundaries among participating agencies and professional personnel. If evaluators are unable otherwise to specify the nature of the intervention, they might reasonably define participation in a program or programs as the "intervention." Comprehensive service

---

[3] The authors have organized workshops where members of the evaluation community and policy officials pursuing systems integration were able to discuss evaluation interests and agendas.

[4] Feedback on this issue has been gathered through site visits with state and local officials who are developing and implementing integrated system models and in meetings of so-called 'lighthouse' sites sponsored by the authors, the National Governors Association's Center for Best Practices through its policy academy initiative, and by the federal government. Interest in answering the core impact evaluation question does not necessarily mean that state and local officials realize what it takes methodologically to measure outcome impacts with any rigor.

[5] See "The Service integration agenda: Political, conceptual, and methodological challenges," Thomas Corbett and Jennifer L. Noyes, in *FOCUS,* Vol. 22:2, Summer 2003m pp. 55-6.

strategies, however, often permit variable forms of "engagement" with the system where some are formally enrolled in a program while others touch upon several programs or have some kind of self-service relationship to the system.

Confusion about service technology - What, exactly, are we evaluating? How do we define and classify the intervention or set of interventions being used when services might be very specifically tailored to the circumstances of a particular family?   Consider, for example, that when asked what they do to help customers respond, front line staff in some of these models respond "whatever it takes."

Confusion about appropriate time lines for achieving an evaluable reform – For many reforms, the point at which substantive change has been introduced and when one can anticipate measurable outcomes is apparent, perhaps even self-evident. In the case of systems integration efforts, the process of introducing reform can be continuous and iterative over a long period of time. It raises questions about when the implementation process is, in fact, complete.

Confusion about appropriate time frames for expecting results - One justification for integrated systems is that they can better address the needs of families with multiple challenges. What time frames are appropriate for expecting impacts to materialize?

Confusion about outcomes - In an environment of multiple programs, diverse professional involvement, and individualized treatment modalities, it may be hard to reach consensus on outcomes.

While a rigorous impact evaluation should establish causality—the new service paradigm results in observable differences in client behavior and circumstances—the conceptual confusions just enumerated makes them challenging objects of inquiry. Given confusion about the basics (who is served, how, and to what ends), is it feasible to assess changes in client outcomes in ways that can be attributed with any confidence to an integrated service model? How can we determine causality if classic random-assignment experimental evaluations of these systems prove impracticable or politically insupportable (e.g., how does one randomize subjects to treatment and control groups when the entire local system of services is being reworked)? Are these models so idiosyncratic and unique that we cannot generalize about effectiveness?

These questions, among others, lead us to a conundrum. To address society's most problematic challenges, we need very sophisticated human service delivery models. Ideally, they should be multi-dimensional, adaptive, and dynamic. Yet, the more such models approach this ideal, the more questions are likely to be raised about whether or not they work. However, the further we push the envelope in advancing the integration agenda, thus generating an even greater need for impact studies tapping net effectiveness, the less likely it is that we can apply appropriate evaluation tools to bear upon the critical research questions.

Given this, we believe that two prior evaluation questions must be given priority as we seek solutions to the methodological challenges associated with doing rigorous impact evaluations. First, is the way services are delivered significantly different from business as usual? Second, can we plausibly argue that this new way of doing business alters the customer's service experience in a manner that is theoretically associated with the goals being pursued? Absent achieving that, little confidence that impacts on customer behaviors or circumstances will follow is warranted.

A corollary, though no less important, interest is whether the innovation bears any correspondence to what was envisioned by the reform's architects, and to any underlying theory of change. To answer these questions, the analyst needs to empirically measure how the new model functions in practice relative to theory or expectation. In effect, we argue for a balanced

approach to evaluating integrated service models where much more attention is given to formative research questions (how well is the model functioning) before attention and resources are applied to traditional summative research interests (what are the impacts).[6]

Designing and implementing a cross-system innovation should focus on one guiding question: is the policy entrepreneur introducing a model that reshapes the customers experience in ways that will plausibly lead to hypothesized or desired outcomes? This gets at the core implementation fidelity concern: whether complex service models are being introduced with sufficient care and quality.

We find that doing what we call a 'line-of-sight' exercise to be a useful way of developing a narrative about the reform and for assessing the fidelity of systems design and implementation. Our work with local and state sites across the country has revealed significant pent-up demand for assistance in this area. Local policy entrepreneurs want help in thinking through what it is they are trying to accomplish and how they can go about assessing whether or not they are moving in the right direction. While this falls short of determining whether client impacts are forthcoming, it is of primary interest to many sites and is where the current evaluation needs rest.

Below, we walk through a template for thinking about these challenges and outline a way of thinking that can have broad applicability to sites struggling with these issues.

Applying a Line-of-Site Perspective to Formative Research

All evaluative research starts with some management or theoretical uncertainty. There is something about the real world and how it works that demands investigation or something about our theory of how the world works that needs verification. Framing our initial management question in ways that focus on the fidelity of implementation, the core evaluation task typically involves some form of the empirical documentation of the client's experience. That is, we are interested in determining if we have really altered the way business is done in ways that impact the consumer experience.

Like all good research, the analyst assumes a null hypothesis: that systems integration in fact does not reshape the consumers experience over what they would get under a regime of distinct, categorical programs. That is, the analyst tries to anticipate where things might go wrong, or where the basic design might be flawed, or where theory might be deficient. Here is how the *line-of-sight* (LOS) perspective, which we briefly review below, comes into play.[7]

The *first* step in the LOS exercise is to articulate clearly the target population and the outcome desired for it. Careful delineation of the target population is an essential preliminary to articulating what is wrong with the current configuration of programs for this population. In addition, there must be a clear understanding of the intended end outcome for the population.

The *second* step is to develop a narrative that describes what the customer will ideally need to experience in order to achieve the intended end outcomes. Developing such a narrative requires

---

[6] This distinction is described well in <u>Evaluative Research,</u> by Edward Suchman, New York: Russell Sage (1967).

[7] For a more complete discussion of this perspective, see "Cross-Systems Innovations: The Line-of-Sight exercise, or getting from where you are to where you want to be", Jennifer L Noyes and Thomas Corbett, in *FOCUS*, Vol. 24, No. 1, Fall 2005, pp. 36-41.

that we recognize an implicit "life-cycle" to the relationship between participants and the system. This life-cycle can be thought of as a sequence of events, interactions, and decision points that play out over the period of an individual participant's tenure in the innovative service delivery model. In the end, the key question to be addressed is what the new system needs to look like from the participant's perspective in order to reasonably believe the end outcomes can be achieved.

Although one cannot know with certainty what a given participant or family will experience in the new system, it should be possible to map out modal scenarios for what a typical family might experience if the innovation were operating as intended. At a minimum, such a life-cycle includes: awareness of the program (signaling and outreach); the front-end experience (application, diagnostics, and routing to key services); service delivery and ongoing case-management (progress monitoring and problem resolution); and exit strategies (determining success and follow-up). For each step of the life-cycle, it should be clear how a given practice or protocol contributes to achieving overall project purposes. If one cannot plausibly defend the 'why' of any particular step, it ought to be revisited.

The *third* step in developing a LOS is to place the narrative describing the participant's experience in the context of the system outputs and outcomes through the use of an outcome-sequence chart. For example, if the integrated system is supposed to deliver multiple services to at-risk families before crises develop—if it operates from a prevention perspective— then the outcome-sequence chart ought to reflect how specific protocols and procedures lead to those ends. If the driving purpose behind systems integration is to strengthen families in ways that will lead to more productive attachments to the labor market, then the systems design features should relate explicitly and plausibly to those particular outcomes. The linkages ought to be clear and persuasive to any reasonable, dispassionate, observer.

In laying out the outcome-sequence chart based on the participant's theoretical life-cycle within the new model, we typically move from left to right. On the far left are some of the activities we view as instrumental to the functioning of the model. We then move through the process changes anticipated to the intermediate outcomes they are expected to produce, and then to the final intended outcome. In following this process, it is possible to see clearly the critical junctures for movement along the sequence, determine where gaps in the line of sight may exist, and ensure that benchmarks of progress toward the end outcomes are established, and ultimately met, along the way.[8] In effect, the story line developed becomes and institutional biography, rooted in the realities of real programs and systems.

Finally, once a LOS has been established from the participant's perspective, the *fourth* step is to consider what changes are needed in institutional practice, administration, and policy to support the transformation in the participant's experience.

Rigorously completing this exercise has several benefits. If done properly, it leads to a fuller understanding of the changes embedded in the new model for organizing and delivering services. It lays out more explicitly the underlying theory of change supporting the proposed changes. It articulates more fully how the new arrangements differ from what customers would experience under the old approach, thereby better demarking the intervention from the counterfactual. And

---

[8] Ideally, one would develop a graphic that depicts what a similar client family would have experienced under the prior system, or what we call the counterfactual. Comparing the two scenarios should give us a rough, or approximate, feel for what the differences are from the customer's perspective.

not least, it provides a convenient roadmap for doing a more formal process or implementation evaluation.

In particular, the LOS allows us to identify the critical junctures in our 'outcome sequence' process and to think creatively about ways in which we can capture data to empirically assess whether or not the changes are first, being implemented as intended and second, having any effect on the participant's experience. Such an evaluation process can:

- sharpen thinking about how to arrive at the best model possible for achieving what you want to accomplish;
- enhance management's capacity for engaging in a process of continuous improvement;[9] and
- provide empirical documentation that the new system, in fact, represents a different reality for customers—an experience that might plausibly lead to expected outcomes.

Clearly, this type of evaluation encompasses both management and research concerns. Thus, it demands that stakeholders from both the knowledge producing and consuming worlds work together; both must bring their respective skills and perspectives to the table.

What Do We Want to Measure?

The LOS exercise is a critical step toward identifying what aspects of the model deserve ongoing attention and thus, helps to articulate salient research questions. The next logical step is to determine how to measure, or empirically document, each task or protocol or decision deemed important to the functioning of the model. For each management question or concern, we need to explore whether there is a way of documenting systemically what is going on and how we are doing.

There are several kinds of systems attributes that we can measure as a way of tapping the fidelity with which we are introducing our reforms. Not all of these potential attributes need to be included in any given evaluation. The following discussion should, however, help us think about those dimensions of the proposed model that might require closer monitoring. We call these input, process, and operational attributes.[10]

---

[9] In a companion paper titled "Moving Toward a Comprehensive Definition of Human Services Integration" the authors introduce the concept of a life cycle for pursuing systems integration. The point made there is that pursuing integration is not an event but a continuous cycle based upon continuous monitoring of key operational measures that feeds into processes of ongoing adaptation and model refinement.

[10] What we do not discuss in this paper are the outcomes associated with more conventional evaluations such as the following: 1) S*horter-term outcomes,* which are typically measures of participant behavior or circumstances, although they might include measures of community functioning. What characterizes outcomes from outputs is that they are rooted in changes in the target population of interest, not in changes in how the system operates. Theoretically, we can differentiate these outcomes into short-term and longer-term measures. Short-term measures typically tap behaviors and circumstances that can plausibly be captured while people are participating in the program or within a reasonably short time after exit. 2) *Longer-term outcomes,* which include some "sleeper" measures that may not be evidenced for some time after participation in the program (e.g., the return on an investment in early childhood development might not be fully realized for several years). A few may be intergenerational (e.g., building stronger families is expected to pay dividends as children mature into adulthood). Although these longer-term goals may not

- *Input attributes* cover the essential investments (money, facilities, skills, etc) required to make the model a feasible, realistic undertaking. This set of benchmarks is intended only to determine whether the resources, technologies, and supports thought to be pivotal to the functioning of the new model are in place. That is, are the necessary resources available, and structured in a way that a dispassionate observer might plausibly argue that a typical customer's service experience will be transformed?

- *Process attributes* cover systems qualities that get at how well a complex service network is functioning as a system—how well the various parts are synchronized. That is, can one empirically measure those points where the implicit narrative embedded in the outcome-sequence chart diagramming the anticipated client experience can go wrong? Assessments (triage), decision points, referrals, key collaborative treatment modalities, cross-system communications, the employment of sanctions, and so forth, are points where complex processes can break down or operate in ways not envisioned by systems designers and managers. For the evaluator, the challenge is to envision creative ways to capture data in anticipation of these potential break points so that problems can be detected early (the canary in the mineshaft' function) and remedial responses designed.

- *Operational attributes* tap qualities of the new system that more directly indicate that the customer's experience has been changed. These attributes are largely rooted in changes in the way the integrated system does business and serve as reasonable proxies for those changes in the quality and character of the participant's experience that lead us to hypothesize differences in the behaviors and circumstances of the target population. Are participants more actively involved in developing service plans? Are service plans more comprehensive, dealing with multiple issues simultaneously? Are services delivered more coherently, or with less duplication of effort? Are plans individualized to the circumstances of the families? Do we have evidence of improved operational efficiencies? Is there evidence of higher quality services?

For each of these dimensions, it ought to be possible to develop quantitative measures that give us some idea of whether or not we are moving in the right direction.

Possible Input Measures. We are interested in these measures because we want to see if there is some realistic chance of obtaining expected results given what we are putting in place. Do we have the right resources, in the correct quantity, and in the proper arrangements? Some illustrative inputs include the following:

> *Staff*—Are there sufficient staff resources to carry out all anticipated tasks and activities?

> *Expertise*—Are the necessary technical and professional skills available and in the proper places?

> *Financial Support*—Has the necessary fiscal support been obtained or is there a plan for using existing resources in new ways?

---

lend themselves to shorter-term assessments of effectiveness, they are useful in shaping how the system ought to be designed and managed. They provide an ending for the narrative development exercise.

Technical Support—If sophisticated information technology or communication systems are necessary to make the new system work, are they in place?

Inter-agency Agreements—If agreements, from contractual to memoranda of understandings, are necessary, have they been secured?

Physical Capacity—Are certain physical plant resources necessary, such as a one-stop services shopping center? If so, are they available or is there a reasonable expectation that they will be available?

A full implementation analysis might well lead to an assessment of the investment requirements demanded of the new system: the service technologies, the communication requirements, the physical plant requirements, the needed staff skills and so forth. For example, if you need people at the front end who can make relatively sophisticated decisions about where families should be referred, they probably need good diagnostic skills and a range of knowledge about a number of programs. Or, what if the new model demanded a real time capacity for communicating across programs located in different buildings about given cases? Can this level of interaction be supported by existing communications infrastructure?

There are a number of these 'feasibility' questions. For each, the analysts can assess whether expectations are realistic or whether more thought needs to be applied to the supportive resource question.

Possible Process Measures. Each of the potential problem areas listed below emanates from the inherent complexity of these models. It is one thing if you basically deliver one service or benefit to whomever walks thought he door and is deemed eligible. It is another thing if you are trying to serve many types of families, yet individualize the services being provided. These are illustrative problem areas and should not be considered exhaustive.

Saturation—Develop measures tapping the extent to which the target population is being reached. In new, complex service models, it is possible that signaling to the target community may be unclear or marketing may be insufficient.

Misdirection—These complex service models often encompass decision points where client needs are assessed using sophisticated diagnostics that cut across several traditional agencies. Every time, choice becomes part of the model, it is possible that assessments will be crudely done and families (or individuals) sent on an inappropriate service trajectory.

Drift—Drift is a condition where an individual or family needing help must wait to get a particular service or benefit. As the complexity of the model increases, there are more opportunities for clients to experience periods of inactivity (or down time). Since several agencies or service providers may be involved in any one family, the probability of queuing increases. Assessing how quickly clients move through the system (what we have called celerity) may be an important attribute to measure.

Leakage–Another possible consequence of complexity is 'leakage" where clients simply get lost. This possibility increases dramatically when there are many opportunities for individuals or families to be referred to other providers or programs. Undoubtedly, some features of an integrated service model (e.g., collocation or team case management) are

specifically designed to minimize this issue but their success in doing so remains an empirical question.

Dosage/Quality –Typically, in programs that provide a single service, or limited number of services, the amount of help provided is more likely to be a known factor. In very individualized service delivery environments that characterize cross-system models, there may be many unknowns, particularly whether some particular services are systematically being omitted or poorly delivered. Or, it might be that some types of clients do not fair well in the new environment.

Such process attributes primarily deal with issues of flow through the linear outcome-sequence graphic that depicts the customer's service trajectory in the new system. They tap key dimensions of whether or not the new system is reaching the intended target population and whether or the target population is being directed along the correct service paths. Finally, they are designed to tap where delays or mistakes or other problems might arise in the efficient functioning of a complex system.

Possible Operational Measures. These measures tap systems outcomes that are believed to be closely associated with typical integrated service models and give us feedback on whether a system is operating as intended if we have any plausible expectation to transforming the client experience and thus achieving desired outcomes. Again, the following enumeration is for illustrative purposes only. Any real set of measures would depend on the characteristics of the model under review and the interests and preferences of responsible officials. A well-functioning integrated service model might well evidence some of the following systems attributes:

Coherence—There would be fewer case plans for the family than under previous regimes and, where there are more than one, there would be less redundancy and contradictions across plans. Redundancy occurs when there is duplication of like or very similar tasks. Contradictions happen when provisions included in separate plans impose time conflicts on the client (be in two places at the same time) or impose expectations on customers that appear inconsistent with one another.

Comprehensiveness—At the same time, the single plan (or at least fewer plans) would be broader in scope, that is, they would deal with a broader range of substantive issues than would have been dealt with under the old service delivery system.

Early Identification and Intervention—These models are often designed to identify issues and problems early and thus, enable a set of services to be delivered when the intervention might be more effective and less costly.

Family Coverage—In many integrated models, the intent is to encompass the family and not just the individual in a case plan. Thus, it might be possible to compare the extent to which service interventions focus on more family members under the new regime compared to the old.

Family Empowerment—Some models make a point that the customer (a 'family' in many instances) should fully participate in the development of any treatment plan. There are, of course, mitigating situations where access to benefits is premised on individuals or families behaving in certain ways. Still, the architects of the new system may wish to document the extent to which families are actively involved in what they get and how.

9

Individualized Treatment—Another attribute of advanced integrated service systems is the capacity to individualize treatment plans and to deliver on those plans. The underlying premise is that 'one size does not fit all.' With some imagination, it may be possible to empirically document whether customers with substantially different presenting circumstances are, in fact, treated differently, and in ways that reflect their uniqueness.

Cooperation/Communication—Another key dimension or attribute of integrated systems touches on the dimensions of cooperation and communication. Relative to the counterfactual regime, the newer systems should evidence more interaction among systems partners around given cases. There ought to be paper or computer trails indicating discussions and cooperative decision making focused on agreed upon family goals.

Satisfaction—We should be able to assess higher levels of satisfaction on the part of service consumers and service providers, again relative to their experience under alternative ways to organizing and delivering services.

These measures enable us to compare and contrast attributes of the customer's experience and/or systems performance on along what are believed to be key dimensions associated with an integrated services model. Though one step removed from the customer outcomes of interest, they can be viewed as important markers for assessing whether progress is being made toward transforming the customer experience in the intended direction.

Let us now introduce one small nuance in terminology to the measurement issue. Let us think about calling these critical attributes in the model *benchmarks*. Why do we call them benchmarks and not outputs or outcomes? To us, the term benchmark suggests status at a point in time that is expected to be compared against some theoretical standard. Form our perspective, benchmarks can best be thought of as pre-selected values that, by policy consensus or expert opinion, represent standards to be achieved if the integrated system is to be successful.

We might well hypothesize our measurements on a given attribute will differ from what we would expect to see under the old policy environment—the counterfactual. But what we might be most interested in measuring is progress toward some *consensually* determined performance level of performance or some agreed upon *idealized* performance level. A consensus standard is one that key stakeholders decide is a level of performance they are shooting for and which might be adjusted (usually upward) periodically. An idealized standard might be one that is set by external authorities or which is very difficult to achieve but stands as a goal nevertheless. Establishing these standards, or benchmarks, is clearly a management function, further solidifying our contention that doing formative research is a joint management-technical undertaking. Let us now turn our attention to a brief discussion of selected issues in doing quality formative research.

Pursuing Formative Research

Pursuing formative research is part craft and a good deal of art. It requires both skill and imagination. In effect, for each event, interaction, and decision point in the narrative describing the participant's experience, the process of doing an outcome-sequence exercise inexorably leads us to the following questions: Why are you pursuing that particular strategy? How does it contribute to the outcomes being pursued? At the same time, this exercise implicitly considers several process questions: What can go wrong? How would you know if things were going in the

wrong direction? What kind of feedback do you need to help remedy any problems? Who needs the information? How should they think about this information?

While this discussion suggests a number of systems' attributes deserving attention, it leaves much left untreated. Which attributes should be addressed in a given situation? Which can be ignored? How do you operationalize some of these concepts? How do you collect data related to these concepts? How do you use these data?

These are challenges for both the practitioner and evaluator communities as we move forward. If both communities work together, we can advance our understanding and our techniques for doing this kind of work. A seminal methodological challenge is to develop an accepted set of strategies for collecting and interpreting empirical evidence upon which to assess the fidelity of implementation as well as ongoing operations. Each site should not have to reinvent the wheel.

Exploring this question in any detail goes beyond the scope of this article. As discussed in a volume by Lennon and Corbett, implementation studies have traditionally been neglected by the evaluation community.[11] The importance of such studies, however, has increased in recent years, as have sophistication of the methods available to evaluators. Werner, for example, suggests numerous strategies for collecting data to be used in implementation studies.[12]

Beyond the science of doing a good implementation, or process, study is the craft involved in doing systems integration in the first place. One cannot divorce the quest for understanding from the full appreciation underlying the quest for change in the first instance. Such studies are part empirical and in part professional judgment drawn from professional experience. Each study is necessarily idiosyncratic and contexture driven. Yet, there is a rhythm and logic to this work perhaps best captured by the following tasks

> *Visioning*—Thinking through the implementation challenge starts with developing a story line about how the new system is expected to function from the customer's experience. As discussed above, this involves developing a narrative about how the integrated system should function. That story helps us identify the points where things can go wrong, the vulnerable decision points and complex tasks. Thus, we can consider which tasks and decision points might require management attention.

> *Setting Expectations*—Typically, an evaluation requires a counterfactual against which to compare a new treatment or policy. This is the strength of an experimental design. Determining what will constitute 'success' in implementation analyses can be more subtle and complicated. In some instances, one might e able to compare selected aspects of the new system with what happened in the old. It might be possible, for example, to compare the number of case managers involved with a family in an integrated model with what happened before the change. In many instances, the comparisons probably will be based on expectations, or standards, developed by management. In essence, some consensus regarding key benchmarks will be the standard on which success is based.

> *Data Gathering*—These methods include administrative data, case record reviews, interviews, staff and customer questionnaires, on-site observations of critical process

---

[11]   Lennon, Mary Clare and Thomas Corbett, editors. 2003. Policy Into Action: Implementation Research and Welfare Reform. Washington DC: Urban Institute Press.
[12]   Alan Werner, 2004. A Guide to Implementation Research. Washington DC: Urban Institute Press.

points, focus groups, and so forth. The mix of methods employed, and how they are employed, depend on salient research questions emerging out of the prior stages.

Interpreting Data—Implementation studies can generate a great deal of data. Translating that data into useable information is not always straightforward. This is clearly where the craft aspect of doing this works comes into play. Quantitative evidence must be assessed in light of the 'feel' about how a system is working in light of observation or reading between the lines. Institutions, after all, have their own cultures that, in many ways, may be much more important than formal organizational charts and empirical evidence.

Providing Feedback—Rather than using results as a way of determining success in some up or down fashion, we use the evidence developed as input toward improving performance, or bringing a systems functioning in line with expectations. Given this, policy entrepreneurs must be integrally involved in all parts of this evaluation process, from developing the research questions through determining what should be measured to thinking through how data will be used to refine and improve the service models and how they are managed.

The art of doing this kind of formative research emanates from the rich diversity of integrated service models out there. Only our lack of imagination limits our ability to generate interesting and useful research questions or to arrive at methodological solutions for exploring them.

Let us take one simple example. Management and evaluators determine that the operational attributes of coherence and comprehensiveness are critical benchmarks to be monitored. They must sit down and figure out how they will measure these attributes, what sources of data to use, and against what standards they will assess performance (and using what time frames). None of these are easy questions. Just take the data or measurement issue. Should one simply review case plans, employ surveys of participating families, or try focus groups of families and staff? Imagination, resources, and the precise nature of the underlying research question come into play.

In the end, doing systems integration and understanding what you have done are interrelated and intertwined processes. They are part of the cycle of steps in the dynamic, iterative, process that characterizes the doing of systems integration.

In Conclusion

Ultimately, this implementation evaluation exercise is far more than an exercise in assessing success. It is an integral part of the very process of doing systems integration. If we do it right, integrating the Line-Of-Sight exercise with good formative research techniques will push us to think vertically (up and down the layers of an organization) by considering how micro-processes or client flow issues are shaped by the way a system is structured and administered. At each point in this hypothetical customer life cycle, there are higher-level (macro-level) institutional functions that presumably support that system-customer interaction.

These macro-level functions include things like management style, information technology, communications, professional training, legal and accounting functions, and so forth. How do each of these shape what happens when the family interacts with the system at various points in this life cycle noted above. Just think about this for a moment. If policies are not aligned, or intake staff not trained well enough, or no one can figure out how to blend resources with imagination,

the dream of an integrated system designed to provide comprehensive services will remain just that—a dream.

Too often, the evaluation task is left to experts and technicians. But evaluation is an integral part of the planning and management process. Determining what you want to measure, how you want to measure it, and how you will use what you measure cannot be divorced from the challenge of doing systems integration. Thus, as we think through what needs to be done to advance the systems integration agenda nationally, we must include a focus on doing good implementation studies and on building up a body of knowledge about strategies and tactics that can broadly be applied.

We have seen extraordinary interest across sites in doing an honest and dispassionate job of learning from these experiments. This should be a high priority for all stakeholders as the systems integration agenda moves forward.