Method Notes

This section includes shorter (e.g., 10–15 double-spaced manuscript pages or less) papers describing methods and techniques that can improve evaluation practice. Method notes may include reports of new evaluation tools, products, and/or services that are useful for practicing evaluators. Alternatively, they may describe new uses of existing tools. Also appropriate for this section are user-friendly guidelines for the proper use of conventional tools and methods, particularly for those that are commonly misused in practice.

The Archipelago Approach To Mixed Method Evaluation

FRANCES LAWRENZ AND DOUGLAS HUFFMAN

ABSTRACT

In this article, we propose a new metaphorical framework for understanding and using mixed methods evaluation. Our re-conceptualization of mixing methods provides important insights into the theoretical and practical challenges of using a mixed method approach. Mixed methods can create philosophical and practical dilemmas in the ways data are collected, analyzed, interpreted and reported. We used an archipelago as a metaphor for resolving these challenges. An archipelago is a set of islands that loosely forms a group. Using the archipelago as a metaphor helped to clarify and re-conceptualize the evaluation approach and its findings by allowing simultaneous consideration of different mixed methods and stances. The results have implications for those attempting to use mixed methods to evaluate programs.

INTRODUCTION

Mixing methods has become a common approach to program evaluation. Despite its popularity, mixing methods can create many unforeseen challenges. The mixing of methods can lead to challenges regarding the measurement of key variables, the analysis of data, and the

Frances Lawrenz • Departments of Educational Psychology and Curriculum and Instruction, College of Education and Human Development, University of Minnesota, Minneapolis, MN 55455, USA; Tel: (1) 612-625-2046; E-mail: lawrenz@umn.edu.

American Journal of Evaluation, Vol. 23, No. 3, 2002, pp. 331–338.
ISSN: 1098-2140

interpretation of results. In this article, we describe a new theoretical framework for thinking about mixed methods called the "archipelago approach." An archipelago is a set of islands connected such that they loosely form a group. The Aleutian peninsula in Alaska is an example of an archipelago comprised of multiple islands connected by an underwater peninsula. The islands that poke out of the surface of the water are merely small tips of the massive underwater structure that forms the foundation of the archipelago. In a sense, the islands are separate, but at the same time connected.

The archipelago concept helped us think about mixing methods by allowing us to view the different data as separate islands, while at the same time seeing all the different data as fundamentally linked by a foundation, as in an archipelago. The underwater foundation of the archipelago was viewed as the underlying "truth" of the program we evaluated, while the individual islands were viewed as ways of probing for the truth, or as data to be analyzed and interpreted. In an archipelago, the vast majority of the structure is underwater and hence out of sight. The islands are in plain sight, but they only represent a relatively small portion of the entire archipelago. This parallels the evaluation of programs. The underlying "truth" of the program is often not in plain view or easy to see. One can collect multiple types of data and use various analysis procedures, but in the end, it is difficult to completely uncover the truth. The islands represent data and our goal as evaluators is to piece together the individual islands and infer the underlying foundation of the program. Sometimes our goal is to piece together the islands to find a pattern or to explain exactly how these different islands fit with each other. Sometimes data are outliers, just as sometimes in a group of islands, one island may not fit into the archipelago. One has to determine if the island is part of the entire archipelago, an outlier that is part of a separate system, or perhaps an individual island in its own right. It is only through the examination of the individual islands that we begin to understand the foundation that holds them all together. Overall, the archipelago approach has helped us to combine and mix methods in a new way that goes beyond some of the current descriptions in the literature of mixing methods.

A comprehensive treatment of the use of mixed methods was provided in the special issue of *New Directions for Evaluation* on "Advances in Mixed Method Evaluation: The Challenges and Benefits of Integrating Diverse Paradigms" (Greene & Caracelli, 1997). Greene and Caracelli suggest three stances to mixing paradigms: The purist stance, the pragmatic stance, and the dialectical stance. Caracelli and Greene (1997) also distinguish between different kinds of mixed methods designs, which will be addressed later in the paper. Proponents of the purist stance argue that different inquiry frameworks or paradigms embody fundamentally different and incompatible assumptions and that they cannot be combined into one inquiry. Therefore, they gather and analyze data in only one way. In our archipelago model, this stance could lead to analyzing different islands, but the islands would all have the same categories of descriptive elements. Proponents of the pragmatic stance assume that the different paradigms are logically independent and therefore, can be mixed and matched to achieve the combination most appropriate for a given inquiry problem. In our archipelago approach, this stance would be similar to drawing the line of best fit between the islands without considering the possibility of curvature or differential importance or other systems or outliers among the islands. Each island is used to develop an overall view of the group. The dialectical position argues that the differences between philosophical paradigms not only exist but are important and can be used both within and across studies toward a dialectical discovery of enhanced understandings of new and revisioned perspectives and meanings. In our model, this would be similar to what a biogeologist would do with disparate data about island groups. What are all the different data

sources, what are the characteristics most likely to be related to the inference of an underlying foundation, what possible models of an underlying foundation are there, what data pose contradictions to the possible models, etc.? Overall, we contend that the archipelago metaphor helps to move these three stances forward by providing a way of considering the differences in each approach, and extending the possibilities for interpretation and combination of methods. The metaphor allows simultaneous consideration of different mixed method stances and designs thereby providing a framework for integrating the types of designs which Greene and Caracelli (1997) suggest are distinct.

THE EVALUATION

Teacher enhancement and curriculum development projects have been funded by the National Science Foundation for many years and there is strong evidence that science teachers are in need of this continuing, supplementary education (Weiss, 1997). Although most people believe that the teacher enhancement programs are beneficial, there have been increasing calls for more careful accountability of the outcomes (NSF, 1993). Very few evaluations have documented the actual effect of enhancement on teachers' classroom behavior or student outcomes (NSF, 1996). Even when outcome data are gathered, usually only one method is employed in spite of the advantages in using of a variety of data collection methods (Patton, 1990; Shadish, 1993). We had the opportunity to conduct a long-term evaluation of a science education reform, which employed mixed methods. We believe that the approach and the description we provide here of that evaluation raise valuable issues regarding mixed methods evaluation.

We evaluated a national project called Scope, Sequence and Coordination (SS&C), which was developed and implemented by the National Science Teachers Association (NSTA). The ultimate goal of the NSTA project was to help students achieve the National Research Council (NRC) Science Standards (NRC, 1996). It was one of the first comprehensive efforts designed to translate the Science Education Standards into actual classroom practice. The SS&C project was based upon the belief that *all* students should study every science each year in a rigorous, inquiry oriented, hands-on, connected, and spiraling manner. The funded portion of the enhancement effort lasted almost 3 years and involved science teachers in 13 high schools across the nation. The project consisted of two summer workshops followed by the implementation of provided curricular materials matched to the philosophy of SS&C. The curricular materials were centrally published by the project and shipped to all sites for implementation. Curricular materials are available on the NSTA website at www.nsta.org.

Our evaluation of the project was designed to fit into a politically charged environment. When we were contracted to conduct the comprehensive evaluation of SS&C, the NRC Science Standards were just being published. The NSTA and the Association for the Advancement of Science (AAAS) had both been involved in the development of guidance for science education reform prior to the NRC receiving funding from the National Science Foundation to develop the Standards as a neutral party. It was crucial that our evaluation provide information that was persuasive and rigorous enough to convince parties on both sides of the fence. The NSF and AAAS are both traditional scientific communities with long histories of using more "logical-positivistic" research approaches. On the other hand, we knew we needed rich descriptive information to fully describe the impact of a program of this magnitude. Consequently, mixed methods were an obvious choice. The question was, exactly how should we mix the methods?

THE EVALUATION APPROACHES

The evaluation of the SS&C program was comprehensive and included three different approaches. Furthermore, within two of the approaches different methods were also employed. The use of the archipelago metaphor will be discussed in more detail below, but overall the three approaches would be analogous to having three different broad types of island characterizations and distinctions of the island descriptions within two of the island types. In terms of the archipelago, we were trying to find different clusters of islands and different information about the islands in the clusters so that we would be able to better understand the underlying foundation or "truth" about the SS&C program. Using only one set of islands or only one type of information about the islands would have limited our ability to understand the complete structure.

Quasi-Experimental Design Approach

The first evaluation approach included a quasi-experimental design with a matched comparison group and both quantitative and qualitative assessments of achievement. About half of the evaluation effort in terms of time and money was committed to this approach, producing in archipelago terms a cluster of about half of those islands used to understand the program. In this comparison design, the students at each site during the baseline year before the reform effort were used as a comparison group for the students in succeeding years. Using the same school as a comparison for itself helped to control for many contextual variables. We hoped that considering the data in a comparative way would allow us to combine data across sites, because we were not intending to discuss absolute level of achievement (which would vary by site) but change from year to year. Student achievement was measured using multiple choice and open-ended science content items, a five-station hands-on laboratory skills test and a hands-on test requiring students to design and conduct an experiment. Together these different assessments provided a comprehensive view of student achievement.

In Greene and Caracelli (1997) terms, this part of the evaluation was pragmatic because it was explicitly included to meet the needs of our stakeholders. The field of science education is mixed in terms of what type of assessment information is perceived as valid. This arises from fundamental differences in the perceptions of what should be taught and measured in science as well as differences in attitudes toward assessment devices. Many believe improved problem solving ability or inquiry skills are the appropriate goal of science education. Others believe a broad knowledge base and ability to understand science related social issues should be the goal (NRC, 1996). The controversy over practical or open-ended science assessment devices versus multiple choice tests is well documented (Lawrenz, Huffman, & Welch, 2001). In its use of multiple assessment measures, the evaluation represented what Caracelli and Greene (1997) call a holistic, integrated design because each measure was seen as necessary in order to understand the complex phenomena of achievement and because the assessments were both qualitative and quantitative.

The archipelago concept helped us to combine the results from the different assessment methods in a more complex integrative way than suggested by Caracelli and Greene (1997). The achievement data collection methods provided different islands, but all of the islands had some descriptive similarity. We considered each type of assessment as a separate island. The islands then formed a cluster of achievement information regarding the underlying "truth" of the

program we were evaluating. The islands all had information about the level of attainment of the students, or metaphorically the topography (underlying form) of the island. The various types of assessment measures, however, provided additional details about the islands, similar to the sort of variation that occurs in terms of rock type or shorelines or other characteristics on individual islands of a cluster. The assessments ranged from the general to the specific. Metaphorically, the broad domain of the multiple choice item island could be considered as providing information about the different rock strata across an entire island (it had many items for each of the science standards), the lab skills test could be considered as providing information about minerals, and the full investigation test information about specific metamorphic rocks.

Social Interactionism Approach

The second approach used in the evaluation was social interactionism (Bogdan & Biklen, 1998). Approximately, a third of the evaluation effort focused on this approach, which in archipelago terms produced a different cluster of islands with information about different characteristics than the cluster of islands related to achievement. Data for this part of our evaluation were gathered through site visits to the schools and classrooms, and surveys. School observations were gathered using open-ended field notes, where the researcher wrote down salient impressions and observations. Classroom observation forms contained field note sections, open-ended category items and quantitative portions to describe, among other things, questions used by teacher, the rapport between the teachers and students, and the amount of time spent in various activities. Interviews with teachers, principals, and students were semi-structured, with broad questions covering topics such as descriptions of a typical lesson and supportiveness of the school. Teacher interviews focused on beliefs about how science should be taught, and what facilitated or inhibited teaching. The student and teacher surveys included items about the classroom learning environment, the frequency of class activities, and student attitude and motivation toward studying science. All of these data were used to provide a narrative description of each site.

In the language of Greene and Caracelli (1997), this social interactionism approach would be considered pragmatic in that it fit the desires of our stakeholders to understand the effect of the reform on the classrooms. Furthermore, this approach, when combined with the quasi-experimental approach, could be considered as a component, expansion design. According to Caracelli and Greene (1997), a component expansion design occurs when different methods are used to assess different aspects of a program but the methods are not integrated with each other. In our evaluation, the social interaction approach focused on the outcome of classroom behavior while the quasi-experimental approach focused on the outcome of student achievement in a side-by-side fashion. The social interaction approach of the evaluation was also mixed within itself using different methods to assess the same phenomenon in order to provide convergence of opinion and increased validity. In Caracelli and Greene's (1997) terms, this would be a triangulated design. As Patton (1997) suggests, triangulated views can corroborate or provide the opportunity to examine differences in detail. Triangulation as a concept was borrowed from surveying and referred to the process of locating a particular point by locating it from three directions. This is a very linear approach and equally weights all three (or more) perspectives. The archipelago approach extends the simple linear idea of triangulation in a complex, unequally weighted, and multi-dimensional manner.

The archipelago metaphor allows us to consider these social interactionist data in a much more interactive and integrative manner. We considered each of the data sources as different

islands with its own unique characteristics, but also with some characteristics in common with the other islands. For example, all of the islands could provide information about climatic conditions but the measures of climate on each of the islands such as amount of rainfall, temperature etc. would be different. We viewed the island of observer comments as the one most indicative of the underlying "truth" about the program, with the other islands providing more detail about the relationship (Lawrenz, Huffman, & Robey, in press). However, others might argue that observers don't sufficiently understand the classroom context and that the observer island would be less indicative than the student perception island because this provides a long-term view from involved individuals. The archipelago metaphor allowed a more in-depth, interactive consideration of these different perspectives and what they meant in terms of relationships between the islands of data.

Phenomenological Approach

The third approach used in the evaluation consisted of a phenomenological study of six of the teachers as they implemented the curriculum. This study involved continuing telephone and in-person, in-depth interviews with the teachers. This approach utilized the rest of the evaluation effort and in archipelago terms produced another unique cluster of islands, one for each of the teachers. These islands had different characteristics and were positioned differently than the clusters of islands formed from the quasi-experimental and social interactionism approaches. The descriptions of these islands were very rich but the descriptions provided about each island were not necessarily the same.

In the language of Caracelli and Greene (1997), this approach within itself was purist; however, but in conjunction with the other approaches it modeled an integrated, transformative design with the potential for dialectical growth. We wanted to integrate the quasi-experimental and social interactionism approaches with the phenomenological approach in the hope that this combination would help to transform the evaluation through valuing pluralistic action. Mixing the methods in this way at least provided the potential for dialogue between the approaches and ultimately enhanced perspectives. Issues raised by the phenomenological study were incorporated into the other data gathering methods in subsequent years, most often into the teacher interviews and surveys. For example, the phenomenological study identified issues of philosophical conflict within the teachers. Generally, the more social interactionist interviews dealt with how the teacher felt about the curriculum and its implementation. These interviews had not delved deeply into what made the teachers feel that way. Therefore, the teacher interviews were redesigned to include more questions about these aspects so the final evaluation was stronger. In terms of the archipelago, the cluster of phenomenological islands identified characteristics we then looked for in the other islands to help understand their relationships and ultimately the underlying "truth."

Combining the Different Approaches

The ultimate goal of the evaluation was to make statements about the value of the SS&C program. To accomplish that we used the archipelago metaphor to help us gather and make sense out of all of the different data. The focusing of different amounts of evaluation effort on the different approaches reflected our beliefs about which approach would provide the most important information regarding the underlying "truth" about the effectiveness of the program. Achievement was viewed as likely to be the strongest indicator of the "truth" so most effort was

applied to finding out about those islands. The learning environment was viewed as the next strongest indicator of "truth" and the teachers' experiences as the least strongest indicator. The islands formed by one approach can be thought of as closer to each other than to the islands formed by another approach—perhaps as clusters on different sides or in different relationship to the underlying foundational structure. In this way, our mixed methods approach allowed us to view the broad picture produced by the different clusters and to use this to spatially combine the data.

Additionally, although the individual islands within each approach cluster had relationships to each other, they could also be considered as having relationships to the islands in clusters formed by the other approaches. For example, the achievement data showed more positive effects for laboratory skills tests while the student ratings of what went on in their classrooms showed that laboratory activities were common. The archipelago metaphor helped us by allowing us to envision the student ratings of classroom activities island as more closely aligned to the island of laboratory achievement than to the island of multiple choice achievement. As another example, the observer comments were more highly correlated with student achievement on multiple choice assessments than the teachers' comments were. Therefore, we considered the observer island as closer to the achievement cluster of islands than the teacher comments island.

We were able to create an average score for all schools on the various achievement tests and to compare achievement test scores over time. However, the social interactionism and phenomenological data showed strong inherent differences within each school and within each classroom. Unfortunately, one simply cannot aggregate and keep things separate at the same time. The archipelago concept helped us to view the achievement data as three islands of types of data (multiple choice test, laboratory skills test, and full investigation) which were enriched by the data from the different schools, for example, school results could be thought of as parts of each achievement measure island. This approach created a way for us to preserve some unique aspects of each school, while at the same time considering that the schools were linked in some fundamental way. We presented the data separately but on charts together so that the uniqueness as well as the trends across the sites were immediately visible. In short, we found the archipelago concept useful in presenting a more coherent and understandable vision of the mixing of evaluation methods. We were able to make more sense out of how the program operated and ultimately its value.

CONCLUDING REMARKS

Mixing methods has wide appeal today as we try to better understand programs from multiple points of view. Additionally, the mixing of methods occurs not only across the evaluation overall but also within approaches. This mixing at various levels often creates challenges, as views cannot always be reconciled philosophically or in practice. The archipelago concept was a particularly helpful way of viewing the various types of mixing because it allowed us to link the different methods in a way that respected the uniqueness of different data and analysis methods while at the same time helping to form inferences about the "truth" of the program we were evaluating. This metaphor goes beyond more linear ideas of triangulation and bracketing by viewing the interrelationships in multiple dimensions. It provides a framework for integrating mixed methods stances and designs that Greene and Caracelli (1997) seem to suggest are distinct. Hopefully, it can lead evaluators to think less in either–or ways about mixing methods

and more in complex, integrative ways. The archipelago approach can re-conceptualize the complexities of mixed method evaluation in a way that can both help evaluators design and conduct evaluations and help explain and report results to stakeholders.

REFERENCES

Bogdan, R., & Biklen, S. (1998). *Qualitative research for education: An introduction to theory and methods* (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Caracelli, V., & Greene, J. (1997). Crafting mixed-method evaluation designs. In J. Greene & V. Caracelli (Eds.), *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms*. *New directions for evaluation* (No. 74, pp. 19–32). San Francisco: Jossey Bass.

Greene, J., & Caracelli, V. (1997). Defining and describing the paradigm issue in mixed-method evaluation. In J. Greene & V. Caracelli (Eds.), *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms. New Directions for Evaluation* (No. 74, pp. 5–17). San Francisco: Jossey Bass.

Lawrenz, F., Huffman, D., & Welch, W. (2001). The science achievement of various subgroups on alternative assessment formats. *Science Education*, *85*, 279–290.

Lawrenz, F., Huffman, D., & Robey, J. (in press). Relationships among student, teacher and observer perceptions of science classrooms and student achievement. *International Journal of Science Education*.

National Research Council. (1996). *National science education standards*. Washington, DC: Author.

National Science Foundation. (1993). *User-friendly handbook for project evaluation: Science, mathematics, engineering and technology education*. Washington, DC: Author.

National Science Foundation. (1996). CREATE/Evaluation Conference, Introductory Session. Washington, DC: Author.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.

Patton, M. Q. (1997). *Utilization focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.

Shadish, W. R. (1993). Critical multiplism: A research strategy and its attendant tactics. In L. Sechrest (Ed.), *Program evaluation: A pluralistic enterprise. New directions for program evaluation* (No. 60, pp. 13–58). San Francisco: Jossey Bass.

Weiss, I. R. (1997). *The status of science and mathematics teaching in the United States: Comparing teacher views and classroom practice to national standards.* Chapel Hill, NC: Horizon Research, Inc.