

# A Lesson in Carefully Managing Resources: A Case Study from an Evaluation of The Music Education Program

Kristin A. Hobson  
*Western Michigan University*

Jason T. Burkhardt  
*Western Michigan University*

**Background:** A music education program with a goal of enhancing cognitive development of preschool-aged children enrolled in local preschools is evaluated by The Evaluation Center at Western Michigan University. The budget for the evaluation was small, and therefore presented several challenges to the evaluation team.

**Purpose:** Through a case study of a local education program, the authors explore the challenges and possible solutions of evaluating a program within time, budget, data, and political constraints.

**Setting:** The case study of a local music education program is implemented in a medium size Midwestern town.

**Intervention:** The local music education program is a participatory music program for preschool-aged children. The program provides a combination of music classes, performances, and parent education to all participants, with further musical opportunities for interested parents and children.

**Research Design:** The research and data collection portion of the evaluation was conducted using participatory strategies, and involved a management oriented approach.

**Data Collection and Analysis:** Data collection for years one and two (process) consisted of primarily qualitative methods, triangulated by exploratory quantitative data collection. Data collection for year three (impact) will consist of primarily quantitative methods, triangulated with results from the qualitative analysis done in years one and two.

**Findings:** Four factors, time, data, budget, and political, constrain the design and conduct of a local evaluation, yet, evaluators may implement multiple solutions that lessen the effect of the constraints on the evaluation.

**Keywords:** *evaluation; evaluation constraints; RealWorld Evaluation*

---

## Introduction

In their seminal 1963 work, Donald Campbell and Julian Stanley assert that validity (particularly internal validity) is the “sine qua non” consideration in research (Campbell & Stanley, 1963, as quoted in Shadish, et al, p. 97, Mathison, 2005, p. 440). By extension any evaluation that uses research methods to gather evidence of merit or value should also be concerned with issues of validity. However, the rigorous methods that are often touted as gold standards of research and evidence (e.g. metaanalysis and RCT) are not often cost effective for small programs needing evidence of value for the purposes of funding and short term decision making.

In *Evaluation: A systematic approach*, Rossi, Lipsey, and Freeman (2004) wrote, “One of the most challenging aspects of evaluation is that there is no ‘one size fits all’ approach” (p.32). They cite a number of contextual influences that surround program evaluations of all types and sizes, including administrative contexts, political structures, and conceptual and organizational factors, to name a few (Rossi, Lipsey, & Freeman, 2004). Additionally, Michael Patton (2008) states that all evaluations require a process of negotiation between the evaluator and the primary intended users. While these admonitions are useful reminders, one finds that guidance on how to make these trade offs is often lacking in the formal literature (Shadish, Cook, & Leviton, 1991). Indeed, Shadish, Cook, and Leviton (1991) suggest that a better theory of evaluation would give prescriptions about how to make these tradeoffs. This begs the question, then, ‘How *does* one conduct a methodologically rigorous and valid evaluation, while meeting the budget and time dependent needs of the client?’

In light of the absence of clear guidance on this matter, the evaluator must construct a framework for understanding each individual evaluation as they begin their work. We framed the present evaluation with a four factor approach suggested by Bamberger, Rugh, and Mabry (2006). In their book, *RealWorld evaluation: Working under time, budget, data, and political constraints*, Bamberger, Rugh, and Mabry (2006) identified four factors that could potentially decrease the rigor of the evaluation design as well as reliability and validity of the findings. These factors included time, data, budget, and political constraints. The following paper presents a case study of an educational program evaluation, including the constraints that affected the design and conduct of the evaluation, the negotiations

between the client and the evaluation team that enabled the evaluation, and the solutions implemented to decrease the impact of the multiple constraints operating on the evaluation.

## Music Education Program

In concert with a local symphony and music instruction provider, a philanthropic organization funded the Music Education<sup>1</sup> program for a period of three years, commencing in June 2010 and ending in August 2013. The Music Education program is an experiential music program for preschool aged children presumed to be at risk for educational or developmental delays. The program consists of multiple music classes, performances, and parent education, including: a nationally registered music education curriculum herein after referred to as Weekly Music, Music Books, Music Families, private music lessons, Family Concerts, and community events.

Weekly Music is a weekly, 30- or 45- minute music class delivered in preschool classrooms and to two parent-child classes within a medium size Midwestern town. The music class includes group singing, moving to music, finger plays, echoing rhythmic and melodic patterns, playing simple rhythm instruments, and saying nursery rhymes. A certified Weekly Music specialist leads the class, assisted by the classroom teacher and aides as well as parents/guardians (in the parent-child classes). Each teacher receives a set of the music materials for use in their classrooms outside of the weekly music class sessions, and each child receives a music book and two copies of a CD with songs, which the children sing in music class.

In the preschool classrooms where Weekly Music is delivered, the children also participate in Music Books. In Music Books, students listen to a story about music in class, and then they make or decorate a simple prop related to the story, such as a violin. After students have listened to the story, the local symphony musicians and a narrator visit the classrooms and perform the story with music, while students play along with their props. The musicians also show students their instruments, as well as demonstrate and explain basic music concepts like high and low pitch, fast and slow tempo, and loud and soft dynamics. Each child takes home a copy of the story, a CD of orchestral music, and an activity book.

---

<sup>1</sup> The program's name and other identifying characteristics have been changed to protect the confidentiality of the participants in the evaluation.

Parents of students who participate in Weekly Music and Music Books are invited to attend three annual informational events. These events, called Musical Families, inform parents about program components, provide ideas for parental involvement in music with their child at home, and educate parents about the importance of music, literacy, and parent involvement in education.

Ten students who participate in Weekly Music and Music Books also receive free, weekly, 30-minute private violin lessons at a local music instruction provider. Five additional students will receive the lessons in year three. The same students continue with lessons as long as they meet attendance and advancement requirements.

Students and families also have the opportunity to attend Family Concerts, which are one-hour symphony performances presented biannually. The concerts center around plots and theatrical elements geared for young audiences. Each preschool classroom teacher receives six Family Concert tickets annually to distribute to interested families at his or her discretion. In addition, each of the ten private violin lesson recipients may request four tickets per year.

The local symphony also delivers free public presentations of Music Books biannually at a local public library. Preschool classes from the community, including classes that receive the Music Education program, are invited to attend the public presentations of Music Books.

## The Evaluation

The administrators of Music Education requested that the evaluation team deliver an evaluation aimed at supporting program improvement, as well as indications of specific program outcomes. Thus, the evaluation team responded by designing an evaluation that could be used both formatively and summatively by the Music Education team. In year one, the evaluation focused on gathering information related to the process of executing Music Education. To do this, a team of evaluators from the Evaluation Center conducted site visits, which were comprised of observations of Weekly Music and Music Books. Observations occurred at each of the preschool classrooms that received Music Education. The team also interviewed parents of children who received Music Education in their classroom and interviewed the music specialist.

In year two, the evaluation will continue to focus on collecting data deemed usable by program managers and stakeholders for informing program improvement decisions. The data also

will be used to assess the extent to which objectives are met. The evaluation team will administer a parent survey in the winter and will observe Music Education and Music Books in the spring. Year three will focus on evaluating program outcomes, undertaken with the intent to provide information for accountability to the funders.

## Challenges

In *RealWorld Evaluation*, Bamberger, Rugh, and Mabry (2006) outlined four major factors that constrain the design and conduct of an evaluation: (1) time, (2) data, (3) budget, and (4) political. The authors make the point that the evaluator must work to realistically deal with these constraints, while doing their best to adhere to the accuracy, propriety, feasibility, and utility standards as set forth by the Joint Committee on Standards for Educational Evaluation (The Joint Committee on Standards for Educational Evaluation, 1994). The constraints that impacted this evaluation and their solutions are presented below.

### Time

A major constraint on this evaluation was the limited period within which the evaluation was to be conducted. The preschool students were available for direct observation during scheduled Weekly Music and Musical Storybook classes. Additionally, parents were available on a limited basis, often due to the competing demands of work and family life. Thus, in order to observe all classes and interview as many parents as possible, the design of this evaluation had to include ways in which to maximize the efficiency of the evaluation team's data collection efforts (Bamberger, Rugh, & Mabry, 2006).

In addition, the evaluation team consisted primarily of students in the Interdisciplinary Ph.D. in Evaluation program that had their own time constraints. Students' academic schedule demands hampered their participation in interviewing parents and observing preschool classes.

### Data

Due to time constraints on the evaluation, limitations in the availability of the data arose that had potential ramifications for the validity of the evaluation. (Bamberger, Rugh, & Mabry, 2006). The limited access to parents, students, and teachers required a reduction in the amount of data that could be collected, how that data could

be collected, and the amount of objectives that could be measured. This impacted the overall ability of the evaluation team to control threats to validity, particularly selection and maturation (Bamberger, Rugh, & Mabry, 2006).

The data collection was also constrained to a predefined proprietary instrument that was provided by the designers of the Weekly Music program. This instrument was created by the designer of the Music Education curriculum, and was therefore not aimed at specific site related implementation. Thus, some of the items on the observation instrument did not apply to the program sites that were observed, or were not directly measureable (for example, interactions between classroom teachers and students between program class times, which was assumed to be measurable by the designers of the instrument). This created issues with construct validity within the evaluation (Shadish, Cook, & Campbell, 2002).

### *Budget*

The limited budget of this evaluation, which underlies the aforementioned constraints, was a major factor in the design and implementation of the evaluation. Bamberger, Rugh, and Mabry (2006) highlight the fact that budget constraints limit the types of designs that can be used; in this case, the three most powerful types of designs (experimental, interrupted time series, and regression discontinuity) had to be excluded. Additionally, the ability to monitor a comparison group was also not an option. Thus, the final design for the evaluation has become a non-experimental single group (treatment) design. This design requires intensive controls at the data analysis and interpretation stages in order to support impact assessments (Bamberger, Rugh, & Mabry, 2006). Also, as noted in the time constraints section, the budgetary limitations precipitated the need for using donated student hours, which further limited the design considerations.

### *Political*

Rossi, Lipsey, and Freeman (2004) noted, "A statement of the purposes [for the evaluation] generally accompanies the initial request for an evaluation, but these announced purposes rarely tell the whole story and sometimes are only rhetorical" (p. 34). This was the case with this evaluation. The collaborative stance of this evaluation provided the opportunity for the evaluation team to be responsive to the dynamic

and changing needs of the stakeholders. However, this approach also reduced the static clarity of the evaluation process as the design had to be changed at several points throughout the project life cycle.

The first year was planned to be a formative assessment of program processes, in order to prepare for the outcome evaluation in years two and three. In preparation for the year two evaluation, several meetings occurred between the program stakeholders and the evaluation team during which lessons learned from year one and plans for years two and three were integrated.

At the onset of the meetings, the evaluation team and the stakeholders agreed that the purpose for year two would be summative, using a success case method. However, at the following meeting, the director of the music program informed us that she would like to continue with the formative assessment, thus the evaluation process in year two would be identical to year one, with the addition of mail-based parent and teacher surveys.

Additionally, the Weekly Music program consisted of a predefined curriculum and evaluation structure, which was not supposed to be modified according to the national rules of Weekly Music; however, the observation tool failed to meet the needs of the evaluation defined by the key stakeholders. Finally, the program is implemented in local preschool programs; thus creating issues with research on sensitive populations (preschool aged children).

### *Additional Challenges*

In the evaluation of the Music Education program, some of the solutions created additional challenges, many specifically related to validity issues. Several potential threats to the validity of the conclusions derived from the outcome evaluation in year three emerged, partly as a result of lessons learned in years one and two. Three common threats to validity that could potentially impact evaluation findings are identified here; issues with internal validity (selection bias and maturation), regression artifacts, and issues with construct validity (confounding constructs with levels of constructs and treatment sensitive factorial structure) (Shadish, Cook, & Campbell, 2002).

In relation to selection bias, the children were all selected into the program based primarily on their attendance in the Head Start program. This non-probability sampling method of selection and assignment meant that it was possible that the children shared unaccounted for common characteristics that could potentially confound

study findings (Shadish, Cook, and Campbell, 2002). Also, rapid intellectual and physical changes over time tend to occur in children in the two to five year age category. These natural maturation patterns would be difficult to account for without design or analytical controls, such as the addition of a non-equivalent dependent variable (Coryn & Hobson, 2011; Shadish, Cook, & Campbell, 2002). Due to the time and budget constraints mentioned earlier, these design controls were not readily available to the evaluation team.

Additionally, since the children were primarily entered into the Head Start program due to perceived developmental issues, regression artifacts also represented a plausible alternative explanation that could not be easily ruled out. More specifically, it would be expected that children who are admitted into a compensatory education program based on risk of low academic/developmental performance would naturally increase their score on subsequent observations, primarily due to chance (Shadish, Cook, & Campbell, 2002).

Another possible confound related to data and budget constraints involves the predefined measurement instruments that were employed in the initial phases of the evaluation. These instruments did not always completely relate to specifics of program implementation in this evaluation. However, construct validity in the measurement structure of this instrument was not tested in this implementation (Shadish, Cook, & Campbell, 2002). Therefore, it is possible that changes in the implementation setting may produce systematic measurement errors that can negatively impact the construct validity of this evaluation.

## Solutions

The shifting purposes of this evaluation reflect Patton's discussion of the "ready, fire, aim" method – i.e. evaluation design is an iterative process (Patton, 2011) where plans are adjusted to meet the needs of the intended users. For example, time constraints affect the evaluation budget, and the evaluation budget affects solutions to time constraints. Therefore when attempting to solve one constraint, another limitation manifests. Even though the scope of the evaluation is considered small, the interconnectedness of constraining factors created a complex evaluation environment.

## Time

The evaluation team relied on students in different phases of their doctoral programs to collect data; this improved the flexibility of the evaluation team to meet the scheduling needs of program staff by covering all the predefined times for observing the program and interviewing parents.

Limitations on the accessibility of parents due to demanding home and work schedules were ameliorated through development of a shortened parent interview that included the least number of items that would answer the evaluation questions. The interview took approximately 10 minutes to complete and therefore enabled the student evaluators to complete all the interviews in a timely manner while being sensitive to the parents' demanding schedules. The interviews were also scheduled to coincide with the timing of already scheduled parent-teacher conferences.

The involvement of the teachers and the music specialist also helped to alleviate time pressures on the evaluation. The classroom teachers and music specialist were asked to provide observational information that could fill gaps in the evaluation team's availability. Capitalizing on available relationships and resources helped to ameliorate these and other time-related issues.

## Data

Several methods were employed to deal with data constraints. One of the team members, who had extensive experience in observing children and classroom behaviors, adapted the observation instrument for this evaluation. Additionally, several of the team members with backgrounds in psychology were able to provide input into the measurement process. In order to calibrate the protocols, the project manager met with each doctoral student individually and the students compared their observation notes with each other. Thus, to whatever degree possible, inter-rater calibration and multiple observations were used to help triangulate and validate observational findings. Finally, teacher observations also provided additional evidence to corroborate plausible explanations for the observed patterns of data collected in the evaluation (Donaldson, Christie, & Mark, 2009).

## Budget

Solutions to the budgetary challenges involved the use of student workers and adaption of the evaluation design. Students donated their time in

return for experience with conducting interviews, observing programs, analyzing actual evaluation data, and writing evaluation reports. Some students also received internship credit hours, which is a graduation requirement of the students' doctoral programs.

The evaluation team also had to modify the evaluation plan to stay within the allocated budget for the evaluation. One modification included the evaluation team's reliance on the local symphony for administrative assistance, for example, distribution of the surveys with envelopes and instructions to classroom teachers, who then placed them in students' backpacks. Classroom teachers collected the surveys and returned them to the local symphony, who then gave the data to the evaluation team.

### *Political*

Strategies employed to deal with political constraints included providing more information about the process of an evaluation in general, including what to expect from an evaluation, how to interpret the results of the evaluation and how the evaluation results could be properly used. These constraints were also addressed through close coordination with the management staff of the local symphony, the county Head Start program staff, and the principals at the program sites.

Design and implementation decisions were primarily made by the evaluation team, with input and discussion from the stakeholders. Changes in the observation instrument were coordinated with the originator of the Music Education program, the Music Education specialist, the local symphony, and the local music instruction provider. Teachers were an especially important ingredient to the solutions, because they helped coordinate observation schedules, facilitated access to parents, and provided additional observational data for periods between the classroom observations.

Finally, monthly in-person meetings and weekly e-mail/telephone communications with important stakeholders, in addition to monthly yearly and midterm reports to the staff members (e.g. a rapid-feedback, management oriented approach) (Shadish, Cook, & Leviton, 1991) provided the program staff with the necessary access to decision and information processes to allow them to effectively synthesize the results and utilize the evaluative findings.

## Conclusion

Despite its apparently small scope, this evaluation provided a set of complex challenges that the evaluators needed to overcome. The primary challenge was how to effectively answer the evaluation questions given the constrained environment in which the evaluation occurred.

Typically, randomized control trials (RCT) are considered the gold standards for evidence; however, implementation of an RCT was not feasible given the context. Consequently, the evaluation design had to use other strategies that accomplish what an RCT does, which is to rule out plausible alternative explanations for observed effects (Donaldson, Christie, & Mark, 2009). The aforementioned constraints reduced the ability of the evaluation to rule out those alternative explanations for observed effects.

In addressing the political constraints, the primary goal was to help the stakeholders create an environment in which the most effective and efficient evaluation designs could be employed. Educating the client stakeholders about the evaluation process empowered them to guide the evaluation team towards meeting their needs for the evaluation while concurrently allowing the evaluation team to design the evaluation according to accepted practices. Working collaboratively with the stakeholders enabled the evaluation team to provide them with a useful product.

Solving the political constraints made it possible to solve the other constraints. Addressing the time, data, and budget constraints facilitated the exclusion of alternative explanations for observed program effects. Through calibration of the instruments and observations, the possibility that observed impacts were due to unsound measurement strategies was reduced. By reducing the time burdens on parents, parental involvement increased and the likelihood that observed effects were due to only interviewing the most motivated parents was decreased. By strategically addressing budget constraints, the team was able to ensure the use of the most rigorous methods in light of the available resources.

Future evaluations of similar programs could incorporate an evaluation process similar to the one used here, in that they might be participatory in nature, give voice to a wide group of stakeholders, make effective use of available allies (such as teachers) to collect data, and involve rapid feedback that is sensitive to the needs of those who are responsible for the implementation of the program.

## References

- Bamberger, M., Rugh, J., & Mabry, L. (2006). *RealWorld evaluation: Working under budget, time, data, and political constraints*. Thousand Oaks, CA: Sage.
- Coryn, C. L., & Hobson, K. A. (2011). Using nonequivalent dependent variables to reduce internal validity threats in quasi-experiments: Rationale, history, and examples from practice. In S. Mathison (Ed.), *Really new directions in evaluation: Young evaluators' perspectives. New Directions for Evaluation* (Vol. 131, pp. 31-39). San Francisco, CA: Jossey-Bass.
- Donaldson, S. I., Christie, C. A., & Mark, M. M. (Eds.). (2009). *What counts as credible evidence in applied research and evaluation practice?* Thousand Oaks, CA: Sage.
- Patton, M. Q. (2008). *Utilization focused evaluation* (4<sup>th</sup> ed.). Thousand Oaks, CA: SAGE Publications.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York: Guilford.
- Rossi, P., Lipsey, M., & Freeman, H. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: SAGE Publications, Inc.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.
- The Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards* (2nd ed.). Thousand Oaks, CA: Sage.