Frequently Un-Asked Questions

Ten Tips to Assess the Integrity of Arts Studies

Suzanne Callahan

Frequently Un-Asked Questions

Ten Tips to Assess the Integrity of Arts Studies

Suzanne Callahan

We all know the ancient fable about the blind men who gather around and touch an elephant. To the one who touches the tusk, an elephant is sharp like a spear; to the one who touches the trunk, it is round like a snake; to the one who touches the side, it is flat. Through their outstretched hands, each man's position around the elephant influences what they "see," the tactile information they receive, and, ultimately, the conclusions they draw about what an elephant is — or is not. This metaphor of the blind conclusions echoes some of the research practices within the arts field. Commonly the "hand" is an online survey, through which data from a few people — typically whoever readily responds — may be used to draw conclusions about the worth of a program or the existence of a trend. The real elephant in the room these days is when practitioners, including those in the arts field, read the words "a study showed that . . ." and readily accept its findings as fact. Limited and biased data may shape conclusions and create convincing factoids and infographics, which zip around social media at breakneck speed to influence our perceptions of trends within the arts field.

Following are ten tips to help guide arts practitioners in assessing the integrity of studies that we encounter, and, ultimately, in deciding whether we should act on or disregard their findings. Each tip begins by asking a question about what to look for in a study, then explains why that question is important for assessing the study's quality, and concludes with a generally accepted standard for the tip. Tips are drawn in part from my book *Singing Our Praises* and the books and teachings of other evaluators. Common terms in research are defined, and the sample methodology section at the end illustrates most of the ten tips in action. In line with the field's growing commitment to equity, these tips incorporate some of the principles of the Center for Culturally Responsive Evaluation and Assessment (CREA). CREA is a group of internationally renowned researchers who require that studies position culture as central to the research process and take seriously the influences of cultural norms, practices, and ethnicity.[1] Finally, because conducting research is as much art as science, no single solution exists for each tip, but they can serve as guidelines to spark thought and discussion.

The real elephant in the room is when practitioners readily accept its findings as fact. Limited and biased data may shape conclusions and create convincing factoids and infographics, which zip around social media at breakneck speed to influence our perceptions of trends.

---

1. Methodology. Does the study contain a methodology section? Does that section address what research was conducted and why methods were chosen? Studies can recount findings from prior research and other sources, generate new analysis on existing data, and/or gather and report on new data. Studies should specify whether new or existing data are used, how, and why. The methodology should be stated, should clearly address tips 2–6, and ideally allude to other tips.

2. Goals and Measurement. What are the reasons for and goals of the study? How will goals be measured? How will the information be used? Among the many possible goals, a study can be designed to better understand a community or issue, assess whether a condition has improved or worsened, inform the design of a new project, or learn about the results of a program. The study asks and answers one or more overarching questions, called *research questions*, related to its goal. Researchers must pay close attention to a concept called *operationalization*, which refers to how study goals are being measured within the methods used and questions asked. A mundane example: If you want to know the weight of an elephant, you use a scale, not a yardstick. And if you want to know the length of its tusk, you use a tape measure, rather than a yardstick; though both measure in inches, one will give you a more accurate answer. A relevant example: If a study goal is to learn about artists' employment patterns in a particular region, a research question might be, "Has the gender gap widened or

narrowed in the past five years?" Researchers would not rely singularly on an open, online survey of whoever happens to respond. Instead, they might gather (or reference) comprehensive data sources that specify location, income, gender, and job status and that are deemed to be representative of the range of artists, locations, and cultures (see tip 4). If sufficient data were not attainable, the study designers might narrow the research goals and collect data accordingly. The study should state its goal(s), research question(s), and how they will be measured, as well as the ways in which findings will be used.

3. Population and Sampling. What population is being studied? If a sample of that population was used, how was it selected, and what was the sample size? *Population* refers to the entire group of people/organizations being studied. Examples are nonprofit performing arts presenters in the Midwest, visual artists in New York, teaching artists in a particular county, and arts administrators who participate in a professional development program. Studies can focus on more than one population.

Broadly speaking, respondents can be selected by three methods. In cases where the population to be studied is small, it is often desirable to include all people (or organizations) involved.

To assess representativeness, the study should address who or what types of people (or organizations or groups) in actuality are most represented within the data, who might be underrepresented, and/or who might be missing entirely.

When the population is so large that all cannot be included in the study, then a commonly accepted practice is to *sample*, or select a subset of the population, either randomly or based on some agreed-upon criteria. *Random sampling* means selecting a sample at random and is one of several practices that can (depending on the type of data) help ensure that the findings from a sample reflect the broader population that is being studied. Within a random sample, researchers should ensure that respondents can only respond once and that all respondents have an equal chance of being included. (Note that randomization refers to *how* respondents are

selected, not who responded. Collecting data from whoever is walking down the street is *not* random sampling; it is called convenience sampling and is generally less reliable.) However, depending on the study design, and particularly if qualitative data are sought, researchers may opt to use *purposeful sampling*, in which people are intentionally selected because they have a set of characteristics that are of interest or relevance to the study. Since there is no single ideal sample size, the size needed depends on a number of circumstances. Statistical guidelines can help determine an acceptable sample size for quantitative data; to help determine the sample size needed, refer to one of many online sources, such as the one at Creative Research Systems' website (www.surveysystem.com/sscalc.htm). Guidance can also be provided by a researcher or statistician from a nearby university.[2] The study should state the size of the population(s) and how its participants were selected, including if and how they were sampled.

4. Representativeness. Did the study employ a process to determine its representativeness? Are the data obtained representative of the population studied? *Representativeness* means the degree to which the data that were collected reflect the entire population being studied. Determining representativeness depends not only on who or what organizations were approached for information, but on who and how many responded, as well as the completeness of responses. *Response rate* is typically defined as the percentage of people who responded out of the total who were asked for data.[3] To assess representativeness, the study should address who or what types of people (or organizations or groups) in actuality are most represented within the data, who might be underrepresented, and/or who might be missing entirely. Such a process typically involves comparing the proportions of respondents to the proportions of the field/population at large by some measurable factors.

An example: When analyzing an existing data set about the arts field's use of technology for the Andrew W. Mellon Foundation from 594 members of five national service organizations (NSOs), my firm began with a two-phase determination of the data's representativeness. We first conducted a quantitative comparison against a national source, the National Center for Charitable Statistics database, using location, budget size, discipline, and other variables, to

determine if our respondent pool sufficiently mirrored the NCCS data. We then interviewed the executive directors of all five of the NSOs, asking them to assess the respondent list for its representativeness based on their knowledge of their membership and the arts field. The Mellon Foundation rightly stipulated in advance that if the data were found to not be representative of the national field of these five disciplines, we would halt the analysis, as we did not want to risk drawing misleading conclusions.[4]

Sampling, response rate, and representativeness are critical to determining whether one can draw conclusions from the data. When data come from a small sample, and/or when the response rate is low, the data obtained may not sufficiently represent the population being studied. If data are deemed to not represent the population to be studied, that is a red flag; the researchers should halt the analysis, gather more data, or issue a caveat as to the data's quality.

If researchers aim to gather information from populations with which they do not have relationships, or which — for any reason — might hesitate to respond, extra time and steps should

People generally respond best when they are in a comfortable environment, being questioned by someone they trust, and understand why the information is being collected and how it will be used.

be built in to obtain data (see tip 5). (Sometimes such groups are oversampled to ensure that if response rates are low, sufficient data are obtained.) It is not acceptable to merely state that a given segment of a population "did not respond" (see tip 6).

Representativeness and response rates should be assessed and clearly stated within the study. The study's author(s) should specify, proportionately, what demographics, experiences, and viewpoints are represented, and what might be missing, unknown, or underrepresented.

5. Data Collection Methods. How were data collected? Where were people asked for data? What were they asked? To what degree did the data collection methods reflect the goals of the study?

Many books have been written on the vast topic of data collection.[5] Here are some key points about how, where, and what is asked or researched.[6]

How. Data collection methods and instruments depend heavily on the goals of the study. Broadly speaking information can be gathered in four ways:[7] Researchers can talk to people (e.g., interviews), observe people (in person or on video), gather written online responses (surveys), or review existing information (e.g., journals, financial records, meeting notes, web statistics, and possibly social media content). The study should state which one or combination of these methods were used.

Where. People generally respond best when they are in a comfortable environment, being questioned by someone they trust, and understand why the information is being collected and how it will be used. If sensitive information is being gathered, people may be more likely to respond if they are guaranteed confidentiality or anonymity. Observations typically happen on-site, as programs happen, but are sometimes done in other ways such as via videos or online platforms.

What. The design of instruments themselves is key to obtaining useful information. Designing effective questions can be challenging, more so than it might appear. The best questions are short, clear, unbiased, and appropriate to the topic. Largely, questions address two broad categories: what people (or organizations) do or what they are — their behavior or attributes — and what people say they want or think is true — their attitudes, needs, and beliefs. Good questions should consider three areas: if respondents are able to answer the questions, whether the questions will produce specific and credible information, and whether respondents will be willing to provide the information. Ideally, questions asked should appear within (or as an appendix to) studies.[8] Observations are typically conducted uniformly; for example, if researchers are to assess the quality of arts classes, all teaching artists included in the research would be observed using similar measures.

In deciding on methods, it is crucial for researchers to *consider the context for the research and any factors that might affect data*

collection. Researchers must assess and address their own experience (or lack thereof) with the population(s) to be studied, by income, education level, location, gender, ethnicity and other cultural factors deemed relevant to the study. They should be competent in their own understanding of the issues to be researched, including interpreting the range of responses that they may receive. As posited by Dr. Stafford Hood, the founder of CREA, researchers should begin by assuming that they do not know a culture and/or its context. They must

---

One of the biggest threats to conducting effective studies is gathering information that does not truly or fully represent the population being studied and thus misrepresents an issue, program, or trend.

---

remain vigilant about questioning and even disregarding their own assumptions about how a given culture operates, particularly when it is not their own. This learning can be enhanced by using *mixed methods*, or carefully selecting complementary quantitative and qualitative data that work together to create a more holistic picture of a program and the culture in which it operates. In summary, it is crucial that researchers challenge themselves to learn about programs from different vantage points and perspectives, which in turn tells a more robust story about a program or trend. Using the analogy from above, *what I am asking the field to do is to take off its blinders and walk around the elephant.*

As methods are selected, researchers must remain sensitive to how respondents will feel when providing information. *Notions of comfort and trust are relative to the population.* For some respondents, this may mean being formally interviewed in their office by an "outsider," and for others, it could mean talking to a peer in a neighborhood café or church basement. Some people respond better orally than in writing, particularly if they are English language learners. Regarding surveys, those who do not have access to computers may prefer to respond using a cell phone or tablet, whereas those who are unaccustomed to using computers may prefer to respond on paper. Similarly, *the language used in questions is critical* and must

be commonly understood by the population studied. For example, in a recent training offered by CREA, Dr. Rodney Hopson led a discussion about the use of the word *related* versus *kin* to describe family relationships. These terms convey the same idea to different populations of people and could therefore generate different responses.[9] In the arts field, terms such as *safe, accessible,* and *quality* can mean many different things depending on the population and their context.[10]

Regardless of what methods are chosen, researchers should follow the simple advice of Dr. Kathryn Newcomer, at George Washington University and former president of the American Evaluation Association: "Say what you did." Newcomer means that a study should state its methods transparently, leaving few questions about how it was conducted.[11] The study should describe clearly how data were collected, including who was asked, how they were asked, and what was asked of them, and it should describe the ways in which data collection was tailored to the culture(s) of respondents.

6. Bias and Limitations. What types of bias and limitations exist within the study? Were they stated? How were they addressed or controlled for in the data collection process? To say that data are *biased* means responses are skewed in some manner merely because of who is asked, how questions are asked, and/or who responded (or did not). It is important to note that *all studies have limitations and some degree of bias.* But an overly flawed research process can reap biased data. One of the biggest threats to conducting effective studies is gathering information that does not truly or fully represent the population being studied and thus misrepresents an issue, program, or trend. Biased data may lead to false conclusions, possibly resulting in recommendations or changes to programs that are inappropriate. Some of the most common forms of bias are as follows:

- **Nonresponse bias** means that the study lacks data from those who chose not to respond or were never contacted, or relies on data from those who readily respond without following up to reach those who do not. It is key to achieve an adequate response rate and attempt to control for nonresponse bias because researchers cannot assume that those who did not respond would answer in the same way

as those who did. Standards for response rates, however, are changing. Particularly for small studies, guidance provided by Pricilla Salant and Don Dillman is useful: "A low response rate serves as a warning that non-response error might be a problem. . . . If the flag goes up, get out your magnifying glass . . . find out whether the people who did not respond are different from those who did in ways that matter to the study."[12]

However, while there is no single standard, the prolific use of online surveys has generally prompted researchers to lower their expectations for response rates, particularly for larger studies. In a recent book Don A. Dillman and colleagues, experts on survey administration, address interpreting response rates for Internet populations, which is "much more challenging"

Data have historically been used to harm some communities, by cutting programs or funding or reinforcing stereotypes, which, in turn, can reinforce inequity. Using limited or biased data risks perpetuating such harm.

---

than in the past, when respondents were contacted primarily by phone or paper. Avoiding such bias is an area where guidance from a statistician or other expert is strongly advised.[13] Regardless of the mode of communication, studies typically request data multiple times, rather than relying on responses received from one request. Providing incentives can be valuable in encouraging responses and avoiding non-response bias. Incentives can take many forms and need not be expensive, such as offering modest gift cards or having a well-known person (sometimes a community leader or gatekeeper) request the response. Newcomer's advice again comes in handy: methods should be described, response rates shared, and limitations stated. Susan Morton and colleagues advise researchers to follow the three Ds, which in turn allow readers to assess the validity of studies: *Disclose* response rates and other factors that influence the data's validity; share the

denominator, or how the response rate was calculated; and *detail* what is known about nonrespondents as well as attempts to improve participation.[14]

- **Self-reporting bias** can happen if individuals and organizations are reporting on how well they themselves are doing within a program or grant (or maybe with their finances). For example, an organization that is behind schedule or over budget on a project may feel fearful and hesitate to reveal its full status, defaulting instead to comments that are superficial or reveal only positive results.

- **Social desirability bias** occurs when people feel pressured to respond in particular ways due to their relationship with the asker. A mundane example: regardless of their true opinions, people typically compliment others' haircuts or cooking. A more relevant example: people who wish to be booked, funded, or presented may offer positive feedback, regardless of what they are asked. We are all human, and people who have a vested interest in the organization conducting the study might be influenced in what they say or what information they provide.

Though biases cannot be avoided altogether, one powerful strategy to increase the validity and reliability of a study is to use *multiple methods*, or a combination of research strategies to gather data. Using multiple methods can decrease the likelihood of drawing conclusions that are either misleading or that hold true for some parts of the population(s) studied but not others.[15]

Researchers should closely monitor and address response rates as well as representativeness from communities who may be less likely to respond. See tips 4 and 5 for suggestions. They must also give time and thought to identifying and monitoring biases, either real or potential. Data have historically been used to harm some communities, by cutting programs or funding or reinforcing stereotypes, which, in turn, can reinforce inequity. Using limited or biased data risks perpetuating such harm. Researchers should state biases as well as, ideally, if and how they were addressed within the study.

7. Analysis. Did the methods and analysis make sense relative to the topic and subjects of the research? Were they stated?

Both quantitative and qualitative data can be interpreted in many ways. Analysis is too lengthy to cover in this article.[16] The good news is that we live in a time when computers have simplified the process of analyzing data. What researchers used to do on paper and index cards, a computer and data analysis software can tabulate in seconds. However, computers only know what researchers tell them, so the challenge for researchers is to combine insight and experience with technology in order to summarize large amounts of data in a manner that helps the reader grasp the main points and make well-considered decisions.[17] Whether using new or existing data, researchers should be crystal clear about which questions or variables are being analyzed and how they are interpreting the data. To help the reader, researchers must summarize their findings but not to the extent that they lose nuances within the data, which might reveal important cultural observations.

That said, from a culturally responsive perspective, researchers should not overuse infographics, factoids, and short summaries. Infographics can hide culturally specific nuances and insights that represent and matter to the population(s) studied. Disaggregating the data can reveal important truths that otherwise would be missed. The process used for analysis should be stated, logical, and appropriate to the data, and it should strive to reveal the truths and complexities of the population(s) being studied.

8. Conclusions and Recommendations. How confident are you regarding any conclusions drawn by the study? How were they drawn? If there are recommendations, is it clear whether they come from the data or the authors' own viewpoints? Drawing conclusions should be done carefully and (particularly for large studies) by people who understand research methods and statistics. *Conclusions should be traceable to the data such that another trained researcher would be likely to draw similar conclusions.* The most common problem that occurs in studies, particularly if conducted with research novices, is drawing hasty conclusions about trends and recommending changes that are based on limited — or inaccurate — information. Once a bit of research has been done, those involved in the study can tend to feel insightful. This can be a vulnerable place because researchers risk drawing conclusions and recommending changes that held true for a limited sample of subjects but not for the

entire population of those served by a program or affected by a trend. The bottom line is, if we base decisions and policy changes on incorrect conclusions from a study, we must always ask

---

As the arts field commits to addressing equity and uses research to identify needs and measure progress, it is crucial that our methodologies and findings incorporate cultural responsiveness.

---

ourselves, What might it cost the arts field, particularly the community(ies) we aim to study and serve? Sometimes the best result from a study is to state that data are inconclusive or additional research is needed. The connection between data and conclusions should be clear, traceable, and defensible.

9. Clarity. Is the study understandable? Will it make sense to the part(s) of the arts field it seeks to serve? For the study to be used, the field generally needs to be able to access and understand it. Visuals such as graphs and charts should be clear and accurately reflect the data. Language should be understandable and easy to follow. As cautioned in tip 7, do not confuse clarity for reductive factoids alone.

10. Cultural Responsivity. Is cultural responsivity addressed at all phases of the study, from its design to data collection to conclusions and uses? Although this point is incorporated into some of the tips above, it is vital that researchers remain vigilant about how culture influences research. As the arts field commits to addressing equity and uses research to identify needs and measure progress, it is crucial that our methodologies and findings incorporate cultural responsiveness. Cultural inequity in the United States began when our country was founded, has persisted ever since, and has influenced the ways we define and measure success and conduct research. There is a long and sometimes painful history of reports being written by outside researchers who came in to "study" communities and ultimately caused harm. One-size-fits-all data collection efforts conducted by outsiders rarely — if ever — incorporate the cultural sensitivity required to learn about and represent a culture. The onus is on researchers, *working closely with people*

from the cultures being studied, to devise methods that better capture the range of viewpoints that their studies claim to represent, and disclose weaknesses. In their design, data collection, and conclusions, studies should address the culture(s) of focus, stating who directed their design, how culturally specific factors were addressed, and what shortcomings arose.

In Closing

Hopefully these tips will help arts practitioners to assess studies with our eyes wide open. By asking good questions and designing studies effectively, we can ultimately gather and use credible information in better ways. Practicing cultural responsivity is not only a research tip — it is an ethical and moral imperative if we are committed to enacting meaningful change in the arts field. In the parable I began with, only integrating the conclusions of all of the ancient

---

One-size-fits-all data collection efforts conducted by outsiders rarely — if ever — incorporate the cultural sensitivity required to learn about and represent a culture.

"researchers" conveyed the full story of the elephant. It remains prescient today. Let's invite one another into the conversation to question, expand, and, ultimately, improve the information on which we base our important decisions.

On the following pages I offer sample text illustrating the ways in which the 10 tips might be appear within a methodology section of a credible study. Call-out boxes reference the number of the tip illustrated in that portion of the text. Please note, in the print edition of this issue, this sample text was incorrectly introduced. This introduction has been corrected as of April 1, 2019.

Suzanne Callahan founded Callahan Consulting for the Arts in 1996 to help funders and arts organizations realize their vision. For over twenty years, her firm has conducted evaluations and studies and administered national funding programs. Her book Singing Our Praises: Case Studies in the Art of Evaluation *was awarded Outstanding Publication from the American Evaluation Association for its contribution to theory and practice in evaluation.*

NOTES

1. These standards were adapted from the Center for Culturally Responsive Evaluation and Assessment (CREA), which was founded by Dr. Stafford Hood and is based at the University of Illinois at Urbana-Champaign.

2. Suzanne Callahan, *Singing Our Praises: Case Studies in the Art of Evaluation* (Washington, D.C.: Association of Performing Arts Presenters, 2005), 128–29.

3. There are other definitions for this term. Refer to American Association of Public Opinion Research, *Response Rates: An Overview*, https://www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx.

4. Callahan Consulting for the Arts, *Choreography in the United States: A Comparative Study of Training and Support Systems*, commissioned by the Joyce Theater (2012) and published by the Andrew W. Mellon Foundation (2014).

5. Refer to the resources section in Callahan, *Singing Our Praises*, 148–51. See also the Public eLibrary on the American Evaluation website at www.eval.org.

6. Adapted from Callahan, *Singing Our Praises*, 124–32.

7. The "four ways to gather information" were originally adapted from the work of Innovation Network. See also Callahan, *Singing Our Praises*, 125.

8. Callahan, *Singing Our Praises*, 128–31.

9. Discussion at Center for Culturally Responsive Evaluation and Assessment conference, Chicago, September 26, 2017.

10. Foundations of Culturally Responsive Evaluation, workshop by Rodney K. Hopson and Karen E. Kirkhart at the Center for Culturally Responsive Evaluation and Assessment, Chicago, September 26, 2017.

11. Kathryn Newcomer, director, Trachtenberg School of Public Policy and Public Administration and Professor of Public Policy and Public Administration, George Washington University.

12. Pricilla Salant and Don A. Dillman, *How to Conduct Your Own Survey* (New York: John Wiley and Sons, 1994), 70.

13. Don A. Dillman, Jolene D. Smyth, and Leah Melani Christina, *Mail and Internet Surveys: The Tailored Design Method* (New Jersey: John Wiley and Sons, 2007), 70, 92. This topic is covered throughout the book, which is recommended for researchers who conduct online surveys. There is enormous variation and duplication in email addresses, as well as "legal and cultural barriers" to contacting people online. Ideally, researchers should verify that "individuals contacted for research by email have a reasonable expectation that they will receive the email," which usually requires a "preexisting relationship with respondents." Dillman warns of response error that results when researchers rely on large, unknown respondent Internet pools; even a large response size can be extremely skewed, leading to false conclusions. Because such samples are nonrandomized, the standard randomization principles usually do not apply, and researchers should make appropriate adjustments.

14. Susan M. B. Morton, Dinusha K. Bandara, Elizabeth M. Robinson, and Polly E. Atatoa Carr, "In the 21st Century, What Is an Acceptable Response Rate?," *Australian and New Zealand Journal of Public Health* 36, no. 2 (2012): 106–8.

15. Callahan, *Singing Our Praises*, 132–33.

16. Callahan, *Singing Our Praises*, 135–39.

17. Callahan, *Singing Our Praises*, 133–39.

1. Methodology Section

Sample Methodology Section

ABC Arts organization program
for children

Goal and Research Questions. A study was conducted to learn about the degree to which ABC Arts was:

2. Study Goals and Uses of Information

making a difference in the lives of its students and their families; opening students' awareness to their own abilities; enhancing students' ability to relate to others; and broadening its students' worldview to careers other than the arts. Findings would be shared with the board and select funders and used to make program improvements in the next season. Instruments would be used by staff to evaluate the program in subsequent years.

Data Collection Methods. Using outcomes and indicators as points of reference, ABC Arts

5. Populations

5. Multiple Methods

staff and Callahan developed instruments to measure the results of its programs on three target groups: teachers, students and parents. Multiple data collection methods were used.

1) Teacher Surveys on Individual Students. One round of surveys was done at the beginning of the term, and another at the end, to compare teachers' perceptions of progress for individual students. An instrument was designed to take about 15 minutes per student, but some teachers found that it took more time. For the second survey, in order to increase the response rate, teachers were asked to enter information only if their responses to the original survey had changed through the course of the semester. A variety of methods were used to increase the response rate. A letter was sent in advance from ABC Arts staff, informing teachers of the survey. The survey was made available online and on paper. The survey deadline was extended. Staff was instrumental in increasing the response rate.

6. Efforts to Increase Response Rates

Out of the 28 teachers who were invited to participate and were able to participate, 27 submitted surveys — a response rate of 96 percent.

2) Student Surveys and Discussion Groups. The second review was targeted at ABC Arts students aged 11-15, using two methods: a survey with a corresponding group interview. Students were interviewed in groups of three or four for 30 minutes. Interviews were tape recorded and evaluators facilitated and documented the sessions (student quotes documented during these discussions are also included in the report). Researchers tried to ensure that students would be engaged and interested in filling out the survey by using the following incentives: The survey design used pictures and formatting to appeal to a younger audience. Students were reassured by evaluators that this was "not a test" and they were free to answer or not answer any questions. Facilitators introduced each section of questions on the survey and encouraged students to verbally express their answers before writing. Finally, students were given refreshments and encouraged to talk with others in the group. Again,

6. Efforts to Increase Response Rates

staff was instrumental in achieving a high response rate. A total of 41 students participated out of the 45 that had been selected, or a 91 percent response rate.

3) Parent Focus Groups. Two parental focus groups were held in evening hours, at a time when parents, some of whom worked multiple jobs, could be onsite. Dinner was provided, and the length of groups

6. Efforts to Increase Response Rates

6. Rate Stated

was kept to 90 minutes. Eighteen parents, all women, participated, representing about 40 percent of the student group.

5. Questions Shared

All surveys and scripts, as well as lists of respondents, appear in the appendices.

Selection of Respondents. In setting the sample size staff considered the trade-off

2. Goals and Measurement

3. Population and Sampling

between *breadth*, or having limited information (such as one survey only) on a greater number of students versus *depth*, or employing the multiple methods described above to examine a limited number of students' experience in a deeper manner. There were limitations on the amount of data that could be collected and analyzed, based on the budget and timeline. ABC Arts decided to prioritize the experiences of students who face income and transportation barriers in attending classes so that it might better understand and address those barriers. Therefore, the study targeted 11-15-year-olds, who had taken class for more than one semester. The sample targeted low-income students, who comprise about 80 percent of ABC Arts' student body. Artistic disciplines were considered equally in the sample. Having multiple measures of this group of students, would help to limit potential bias and increase the validity of the study, so that conclusions could be drawn with more accuracy. In

8. Relationship of Methods to Conclusions

subsequent years, and based on what was learned in year one, ABC Arts intends to evaluate its program more broadly with different student groups.

Motivation to Participate. Teachers work on a volunteer basis and may not be motivated to complete lengthy instruments. Students were young and could not be expected to complete long surveys during the time of their art class. And, parents have busy lives, many with multiple jobs and other children that keep them fully occupied during the hours when their children are at ABC Arts. The methods above, and the protocol described below, attempt to compensate for all of these factors.[1]

Protocol. The tone of all communication conveyed that students were not being evaluated, but instead that ABC Arts was undertaking a study to learn about how its arts training affects the young people that they are dedicated to serving. Because student minors would be involved in data collection informed consent was addressed at the time of registration, parents granted written permission for their children to participate. Finally, staff feedback was obtained throughout the process so that the study could benefit from their insight.

Representativeness. An assessment of the students' representativeness concluded an acceptable range according to the following factors:

Factor	Student Population	Student Respondents
Income	80% low, 15% mid, 5% high	70% low, 20% mid, 10% high
Gender	60% female, 30% male, 10% unknown	60% female, 40% male
Discipline	33% each dance, music and visual	33% each dance, music and visual
Location	95% reside in 3 nearest zip codes	80% reside in 3 nearest zip codes

Parent respondents were 100 percent female; ABC Arts does not have access full information on the gender of parents for each household but confirms that the majority of them are women.

4. Representativeness

Nonetheless, it can be assumed that parental respondents are not fully representative of parents, according to gender.

Limitations and Controls

Nature of change. ABC Arts' stakeholders were most interested in tracking growth in specific outcome areas. However, in order for the evaluation to be credible, the research design allowed for the possibility of decrease in all outcome areas as well. This is reflected in the manner in which questions were asked, and the language used in letters of invitation.

5, 6 & 10. Bias, Limitations, and Context

Degree and timing of change. Due to the limitations of the study, data collection relied chiefly on self-reporting and direct observation during a set time period of several months. Additionally, since the sample consisted of students who have been attending ABC Arts for some extended time period, it is not possible to determine the exact degree of change that can be attributed only to that semester. In fact, some questions intentionally sought

information about students' degree of change over the course of time they have studied at ABC Arts.

Other influences. Since students spend several hours per week at ABC Arts, and receive instruction and guidance in other locations (such as school and home) and from other people (such as parents, other family members, teachers, friends, and possibly other mentors), it is not possible to isolate with 100 percent certainty the influence of ABC Arts alone on students' growth.

For these reasons, multiple methods of data collection, including interviews, direct observation and surveys, were combined and used to attempt to discern ABC Arts' contribution toward young people's growth in the areas above. This allows for drawing conclusions with more confidence.

Potential Biases and Controls

Self-Reporting. When reporting on how students are progressing in their own classes, teachers may tend to give positive responses, or could perceived to be more positive than if they were evaluating other teachers' classes. Therefore, the researchers made clear that they were not being evaluated

| 6. Potential Biases and Controls |

as teachers. Multiple measures of data collection from parents, students and staff helped to control for this bias, and similar questions were also asked of the students themselves.

Social desirability. Students could feel they are being tested during the survey, or feel uncomfortable making less-than-positive statements about ABC Arts since its staff is perceived by them as authority figures. Therefore, as discussed above, data collection was made as fun and personable as possible, by keeping instruments short and visually appealing; providing food; and encouraging friendly discussion, in both what was asked and the way it was asked.

Confidentiality. Parents may be unwilling to share in a group their misgivings about ABC Arts, or personal information about their child or family life. In response, staff was not present in the room. In the end, the consultants felt that parents were forthright and the data show a range of impressions, both positive and constructive.

NOTE

1. Callahan, *Singing Our Praises*, 140.