

Measuring Outcomes for Fatherhood Programs

Types of Measures

There are a number of different types of measures used in research.

- **Qualitative measures** often include open-ended questions asking respondents to describe their experience, beliefs, or views about a topic. For example, evaluators could ask fathers to describe if, how, and to what extent their relationship with their children has changed since attending the fatherhood program.
- **Quantitative measures** are statistics driven and may include questions or items that ask fathers to rate or rank some aspect of behavior, attitudes, or beliefs. For example, fathers may be asked to rate how important their role as a father is to them compared with other roles. There are also quantitative measures that ask fathers to count the frequency of a behavior, such as the number of days per month they had contact with children. In addition, there are quantitative observational measures, where an observer may count the number of times that the father engages in a child-related behavior such as praising the child.

There are advantages and disadvantages to using each type of measure described above. **Qualitative** measures are useful for obtaining a deeper understanding of a topic such as why fathers were able to increase their involvement with children. Another advantage is that fathers are able to describe an experience in their own words without having the evaluator decide in advance which experiences are important to focus on. A disadvantage of qualitative measures is that it is more difficult to compare the answers given by different individuals. One father may say he is much happier with his access and visitation arrangement following his participation in the program. If another father says that things are better today than they were, is this an equally positive response?

An advantage of **quantitative** measures is that they allow the evaluator to calculate how much change has occurred in some target behavior, attitude, or belief over a specified period of time. For example, the evaluator may want to determine whether participation in the fatherhood program was associated with increased contact with children. The best measurement approach would be to ask fathers (or mothers or children) how often the father had contact (e.g., face to face, telephone) with the child over a specified time period. Quantitative measures may be single questions, or a series of questions, that measure varying aspects of the same general variable. For example, in measuring contact, the

FRPN Co-Director Jay Fagan, PhD | *Professor, Temple University School of Social Work*
Ritter Hall Annex, 5th Floor | Philadelphia, PA 19122 | (215) 204-1288 | jfagan@temple.edu

FRPN Co-Director Jessica Pearson, PhD | *Director, Center for Policy Research*
1570 Emerson Street | Denver, CO 80218 | (303) 837-1555 | jspearson@centerforpolicyresearch.org

FRPN Coordinator Rebecca Kaufman, MSW | *Temple University*
Ritter Hall Annex, 6th Floor | Philadelphia, PA 19122 | (215) 204-5706 | rebecca.kaufman@temple.edu

researcher may want to use a series of questions that get at the total amount of time spent with the child, the type of contact (such as overnights or supervised visits), or the regularity of contact.

Measuring the Appropriate Outcomes

Researchers and practitioners should be very thoughtful and deliberate when selecting outcome measures. One of the first steps is to decide the expected outcomes of the program. For example, are fathers expected to become more knowledgeable about parenting and child development as a result of participating in the program? Are fathers expected to have more stable and better jobs as a result of their involvement in the program? The use of a logic model will be helpful in determining expected outcomes from program participation (see FRPN brief on logic models).

Selecting Outcome Measures

After determining appropriate outcomes for the program, the next step is to consider possible measurement tools. FRPN strongly recommends that practitioners consult with researchers about this matter. Researchers have been trained to evaluate the quality of measurement tools. If multiple programs use the same outcome measures, it becomes possible to compare across programs. In addition, using instruments that have already been tested can help ensure that the questions are understood and measure what they propose to measure.

The following are criteria to consider when selecting measures:

1. Does the measure ask questions that are relevant to the outcome being measured?
2. Is the measure appropriate for the population of fathers being evaluated? For example, does the measure use words that are familiar to the population?
3. Are the items clear and easy to answer? Negatively worded items can be confusing. Sometimes respondents are given a series of statements and asked how much they agree or disagree with each. The statement, "We did not need to spend more time on any topic," is less straightforward than, "We spent the right amount of time on each topic." You should also avoid compound questions. If you ask the father respond to the item, "The child's mother does not approve of me as a father and she lets the children know how she feels," there may be confusion because he agrees that the mother does not approve of him, but he is unsure about whether she shares this with the children. It is better to split these into two questions.
4. Does the measure consist of a single item? Single item measures may lack validity. In the example above, we noted that "contact with children" may include number of days or regularity of contact. On the other hand, measures that include too many items may lead to participant fatigue.
5. Does the measure have adequate variance? That is, do all of the fathers tend to give the same response to an item? For example, if at the pre-test all the fathers respond "agree strongly" to

the statement, “I enjoy spending time with my children,” this item will be of little use in measuring change over time.

Reliability and Validity

You may hear discussions about whether a data collection instrument or measure is reliable and valid. Essentially, if a measure is valid, it means that it is measuring what it purports to measure, and if it is reliable, it means the results of using the instrument should be consistent.

For those who design measures and scales, there is a little more to it. For example:

Internal consistency reliability assesses whether the items in the scale are consistent in measuring the same construct.

Inter-rater reliability refers to when different data collectors use a form to record what happens during a parent-child interaction, or if two data collectors try to extract the same information from a program file, then the results they come up with should be the same.

Content validity focuses on the extent to which the measure covers the range of meanings included within the concept. For example, if a researcher creates a scale that is supposed to measure being a “good” father, the scale needs to include items to measure a number of different aspects of fatherhood.

Criterion validity assesses the degree to which the measure correlates with another indicator or measure of the same variable. Researchers may ask about the participants’ satisfaction with the program with several different questions. All of the questions should show high correlations.

Construct validity assesses the extent to which a measure relates to other variables within a system of theoretical relationships. During an exit interview, a researcher may ask respondents whether the program did a good job of explaining the importance of regular contact with children. At the follow-up interview, if those who say the program did a good job covering regularity of contact show greater regularity, report fewer missed visits, and report that contact is following the parenting plan, the item has construct validity.

Challenges

One of the greatest challenges is that researchers and practitioners may not be able to find already existing measures that seem to address the outcomes expected of fathers who participate in the program. In these instances, evaluators may have to develop new measures. It is always important, however, to use good measurement practices when developing a new measure.

Another challenge is that the existing measures may be appropriate for some fathers in your program but not for others. For example, you may serve fathers whose children range in age from birth to young adulthood. Existing fatherhood measures are typically designed to be relevant to children of a particular

developmental period (e.g., infancy). Researchers and practitioners may have to select several measures and match them to the age of children.

Another challenge is that researchers and practitioners may only find measures that are exceptionally long and time consuming to complete. Some researchers have developed short forms of a measure, and these may be just as valid as using the longer version of the scale. It is sometimes worthwhile to contact the researcher who developed the measure to determine if there is a short form.

Sample Measures of Father Involvement

The Relationship with the Child questionnaire (Coley & Morris, 2002) is a promising measure for assessing quantity of father involvement because it is one of the few measures developed for both residential and nonresidential fathers. Based on the Lamb, Pleck, Charnov, and Levine (1985) conceptualization of paternal involvement, this instrument measures three aspects of father involvement: responsibility, accessibility, and engagement. Coley and Morris (2002) developed this measure of father involvement by adapting questions from previous studies (Cabrera et al., 2004; Coley & Chase-Lansdale, 1999).

- To assess responsibility, fathers are asked: (1) *How much responsibility do you take for raising the child?* and (2) *How much does your help with financial and material support of the child help the child's mother?* These questions are rated on a 4-point Likert scale (1 = none, 4 = a lot); higher scores suggest greater paternal responsibility.
- To measure accessibility, fathers are asked: (3) *How often do you see or visit with the child?* and (4) *How often does the child see or visit with your family?* These questions are rated on a 9-point Likert scale (1 = never, 9 = every day), with higher scores indicating greater paternal accessibility.
- To measure engagement, fathers are asked: (5) *How many hours per week do you take care of the child?* and (6) *How much does your involvement make things easier for the [child's mother] or make [her] a better parent? (1 = none to 4 = a lot).* Studies have shown strong predictive validity and internal reliability for the Relationship with the Child questionnaire (Hernandez & Coley, 2007).

Resources

Additional suggestions for specific measurement instruments can be found at http://www.ibassoc.com/reports/documents/rf_evaluation_final_v2-small.pdf.

- Cabrera, N. J., R. Ryan, J. D. Shannon, J. Brooks-Gunn, C. Vogel, H. Raikes, C. S. Tamis-LeMonda, et al. (2004). Fathers in the Early Head Start National Research and Evaluation Study: How are they involved with their children? *Fathering: A Journal of Theory, Research, and Practice About Men as Fathers*, 2, 5 – 30.
- Coley, R. L. & P. L. Chase-Lansdale. (1999). Stability and change in paternal involvement among urban African American fathers. *Journal of Family Psychology*, 13, 416-435.
- Coley, R. L. & J. E. Morris. (2002). Comparing father and mother reports of father involvement among low-income minority families. *Journal of Marriage & Family*, 64, 982-997.
- Hernandez, D. C. & R. Coley. (2007). Measuring father involvement within low-income families: Who is a reliable and valid reporter? *Parenting: Science & Practice*, 7, 69-97.
- Lamb, M. E., J. H. Pleck, E. L. Charnov, & J. A. Levine. (1985). Paternal behavior in humans. *American Zoologist*, 25, 883-894.
- Rubin, A. & E. R. Babbie. (2011). *Essential research methods for social work* (7th Ed.). Belmont, CA: Brooks/Cole Cengage Learning.
- White, J. M., & D. M. Klein. (1996). *Family theories*. Thousand Oaks, CA: Sage Publications.
- Frey, B. B. (2014). **100 Questions (and Answers) About Tests and Measurement**. Thousand Oaks, CA: Sage.
- Salkind, N. J. (2013). *Tests & Measurement for People Who (Think They) Hate Tests & Measurement* (2nd Ed.). Thousand Oaks, CA.