

The Tools of the Trade: How to Find or Create the Evaluation Tools You Need

Dan A. McDonald
Donna J. Peterson
CYFERnet Evaluation
The University of Arizona

CYFERnet Evaluation Web Resources

- Designing a Program Evaluation
- Process Evaluation Tools and Instruments
- Outcome Evaluation Tools and Instruments
- Data Analysis and Reporting
- Evaluating Early Childhood Outcomes
- Evaluating School Age Outcomes
- Evaluating Teen Outcomes
- Evaluating Parent/Family Outcomes
- Evaluating Community Outcomes
- Evaluating Organizational Capacity
- Evaluating Technology Use
- Evaluating Program Sustainability
- Building Capacity for Evaluation




CYFERnet Evaluation Resources

- Early Childhood
- School Age
- Teens
- Parent/Family
- Community

- Program Support
- Evaluation
- Technology
- Online Training
- About CYFERnet

- Experts Directory
- Community Project Profiles
- Spanish Language Resources
- CYFAR Conference
- CYFAR 07 Conference Proceedings

CYFAR • CYFAR Program

 • 4-H Youth Development Program


- Site Map

Search: Title/Abstract Full Text Author

[Go](#) [Advanced Search](#)

Evaluation Tools and Instruments

- **General Information about Tools and Measures**
- **Process Evaluation**
Populations served, satisfaction, program accountability, etc.
- **Physical Activity, Nutrition and Diet**
- **Developmental and Behavioral Health**
Learning disabilities, emotional/behavioral problems, etc.
- **Risky Behaviors**
Drug and alcohol use, sexuality, violence, etc.
- **Positive Youth Development**
Academic success, leadership, youth-adult partnerships, etc.
- **Child Care**
- **Community Change**
Collaboration, community involvement, etc.
- **Program Sustainability**
- **Technology**
- **Organizational Capacity**




Evaluation

Evaluation

- Designing a Program Evaluation
- Evaluation Tools and Instruments
- Data Analysis and Reporting
- Evaluating Early Childhood Outcomes
- Evaluating School Age Outcomes
- Evaluating Teen Outcomes
- Evaluating Parent/Family Outcomes
- Evaluating Community Outcomes
- Evaluating Organizational Capacity
- Evaluating Technology
- Program Sustainability
- Building Capacity for Evaluation
- Reports and Sample Evaluations

- **Submit educational Materials to CYFERnet**

JUST for KIDS



Teens R It

Reliability and Validity

- A quick Reading Assessment
- How reliable is this measure?
- How valid is this measure?

Reliability and Validity

- Reliability: *Are things being measured consistently?*
- Validity: *Are we measuring what we think we are?*
- *Bathroom Scale example*

Why are these concepts important?

- Without the agreement of independent observers able to replicate research/evaluation procedures, or the ability to use research tools and procedures that yield consistent measurements, researchers/evaluators would be unable to satisfactorily draw conclusions, formulate theories, or make claims about the generalizability of their work.

Reliability

- Extent to which an experiment, test, or any measuring procedure yields the same result when repeated
- Refers to the precision of a measurement
- *Are things being measured consistently?*

Four Types of Reliability

- Equivalency or Parallel Forms Reliability
- Stability or Test-Retest Reliability
- Internal Consistency
- Interrater or Interobserver Reliability

Equivalency or Parallel Forms Reliability

- Extent to which two items/sets of scores measure identical concepts at an identical level of difficulty
- Two different instruments designed to measure identical constructs are developed and the degree of relationship (correlation) assessed
- The higher the correlation coefficient, statistically referred to as r , the better



Stability or Test-Retest Reliability

- Consistency of repeated measurements on the same subjects
- To determine stability, a measure or test is repeated on the same subjects at two different times and results are correlated
- Two possible drawbacks:
 1. A person may have changed between the first and second measurement
 2. The initial administration of an instrument might in itself induce a person to answer differently on the second administration (“practice effect”)

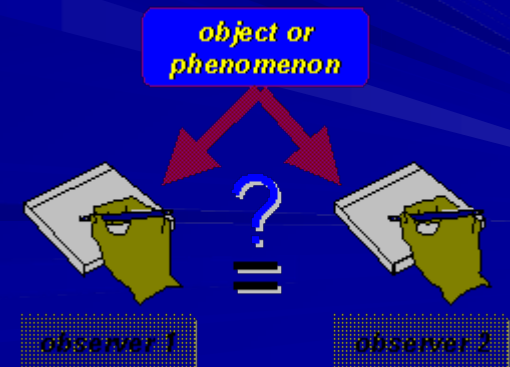


Internal Consistency Reliability

- Extent to which tests or procedures assess the same characteristic, skill or quality
- *Do the items in a measure correlate highly?*
- Cronbach's alpha is used to show how well the items complement each other in measuring different aspects of the same variable
 - alpha reliabilities above .70 are considered good
- Helps researchers interpret data and predict the value of scores and the limits of the relationship among variables

Interrater Reliability

- Extent to which two or more individuals (coders, raters, observers) agree
- Addresses the consistency of the implementation of a rating system
- Interrater reliability is dependent upon the ability of two or more individuals to be consistent



Validity

- Extent to which the measurement procedure actually measures the concept that it is intended to measure
- Refers to whether a measurement actually taps into some underlying 'reality'
- *Are we measuring what we think we are?*

Internal and External Validity

- Internal validity:
 - evidence that what you did in the study (i.e., the program) caused what you observed (i.e., the outcome) to happen
- External validity:
 - extent to which the results of a study are generalizable or transferable to other persons in other places and at other times

Types of Internal Validity

- Face Validity
- Criterion Related Validity
- Construct Validity
- Content Validity

Face Validity

- Does it seem that we are measuring what we claim?
- Does the measure seem like a reasonable way to gain the information we are attempting to obtain?
- A subjective measure of validity

Content Validity

- Extent to which items in the instrument reflect the purpose of the data collection effort
- Does the content of the measuring instrument reflect the specific intended domain of the concept?

Criterion Related Validity

- Demonstrates the accuracy of a measure or procedure by correlating it with another measure or procedure which has been demonstrated to be valid (called the criterion)
- *Concurrent* criterion validity:
 - are results of a new questionnaire consistent with results of established measures, e.g., a "gold standard"
- *Predictive* criterion validity:
 - assesses the ability of a survey to predict future phenomena

Construct Validity

- Seeks agreement between a theoretical concept and a specific measuring device or procedure
- Does the measured concept relate empirically to other measured variables in ways that are theoretically expected?
- Convergent validity:
 - the agreement among ratings (gathered independently) where measures should be theoretically related
 - extent to which different data collection approaches produce similar results
- Discriminant validity:
 - the lack of a relationship among measures which theoretically should not be related
 - extent to which results do not correlate with similar but distinct concepts

Enhancing Internal Validity

- Think through your constructs carefully before developing an instrument.
- Have experts critique your instruments.
- Implement multiple versions of an instrument to see if they produce the same results.
- Implement multiple measures of key constructs and show that they behave as you theoretically expect.

Enhancing Internal Validity (contd.)

- Keep the population with which the measure was normed in mind.
- Be aware that participants may try to guess what the purpose of the study is and answer questions based on their guess.
- Make participants feel comfortable.
- Do not consciously or unconsciously bias participant responses.

Enhancing External Validity

- Be thoughtful when selecting a sample.
- Once selected, try to keep dropout rates low.
- Thoroughly document your sampling methods, participant characteristics, and study procedures.
- Do your study in a variety of places, with different people and at different times.

Qualitative Research Quality Standards

- Confirmability (equivalent to objectivity)
- Dependability (equivalent to reliability)
- Credibility (equivalent to internal validity)
- Transferability (equivalent to external validity or generalizability)

Confirmability

- Equivalent to objectivity
- Are the data and interpretations grounded in the participants' actual descriptions and not in the researcher's imagination?
- Questions to assess this standard include:
 - 1) Are the study's methods described explicitly and in detail?
 - 2) Can the sequence of data collection, processing, transforming, and conclusion drawing be followed?
 - 3) Are the conclusions linked to displayed data?
 - 4) Has the researcher been explicit about personal assumptions, values, biases, and affective states and how they may have played a role in the study?
 - 5) Are the data available for reanalysis by other individuals?

Dependability

- Equivalent to reliability
- Was the data collection process consistent and stable over time and across researchers and methods?
- Questions to assess this standard include:
 - 1) Are the research questions clear and the study procedures congruent with them?
 - 2) Were data collected across the full range of appropriate settings, times, respondents, etc., suggested by the research questions?
 - 3) Are findings meaningfully parallel across data sources?
 - 4) Were any forms of peer or colleague review in place?

Credibility

- Equivalent to internal validity
- Is there a match between the realities described by participants and those represented by the researcher? In other words, are the findings credible and do they present an accurate description of the topic being examined?
- Questions to assess this standard include:
 - 1) How context-rich and meaningful are the descriptions?
 - 2) Does the account make sense and seem convincing or plausible to the reader?
 - 3) Is the account comprehensive?
 - 4) Did the original informants feel the conclusions were accurate?

Transferability

- Equivalent to external validity or generalizability.
- Are the conclusions of a study transferable to other contexts?
- Questions to assess this standard include:
 - 1) Are the characteristics of the original sample, settings, and methods described in enough detail to permit comparisons to other samples?
 - 2) Do the findings include enough thick description for readers to assess the potential transferability or application to their own situations?
 - 3) Does a range of readers report the findings to be consistent with their own experience?
 - 4) Are the processes and outcomes generic enough to apply to other settings?

Critiquing Measures

- What are the positive and negative aspects of these measures for the sample program?
- What can you tell about the reliability and validity of these instruments?

Critiquing Measures

- Program 1: Community Enhancement
 - Ages 4-13—community, literacy and life skills
- Program 2: Seeds to Success
 - High school age—work force prep, life skills, service learning
- Program 3: Parent Education
 - Adults—parenting and life skills

Electronic Survey Formats

- Distributed as e-mail messages
- Posted as web forms online
- Distributed via publicly available computers in high-traffic areas such as libraries and shopping malls
- Placed on a laptop and completed there rather than on paper

Strengths of Electronic Surveys

- Cost savings
- Ease of editing/analysis
- Faster transmission time
- Easy use of pre-letters
- Higher response rate
- More candid responses
- Potentially quicker response time with wider magnitude coverage

Weaknesses of Electronic Surveys

- Sample demographic limitations
- Lower levels of confidentiality
- Layout and presentation issues
- Computer equipment
- Additional orientation/instructions
- Potential technical problems with hardware and software
- Response rate

Dillman Method

- Prenotice – 3 days before questionnaire is sent
- Questionnaire
- Thank You/Reminder/Replacement Questionnaire – 1 week after first questionnaire
- Additional contacts do not significantly increase response rate

A Few Online Survey Sites

- SurveyMonkey
- QuestionPro
- Zoomerang

Factors to Consider

When selecting an online survey source, consider:

1. Number of responses
2. Number of surveys
3. Number of pages and/or questions
4. Question format/styles
5. Ease of survey design
6. Customizability
7. Download data for analysis in SPSS or Excel
8. Reporting features
9. Price

References

- Barribeau, P., Butler, B., Corney, J., Doney, M., Gault, J., Gordon, J., Fetzer, R., Klein, A., Ackerson Rogers, C., Stein, I. F., Steiner, C., Urschel, H., Waggoner, T., & Palmquist, M. (2005). *Survey research*. Writing@CSU. Colorado State University Department of English. Retrieved 4/28/08 from <http://writing.colostate.edu/guides/research/survey/>.
- Dillman, D. A. (2006). *Mail and internet surveys: The tailored design method 2007 update (2nd Ed.)*. Wiley Publications.
- Fisher, S., Andersen, R., & Heath, A. (no date). *Survey research methods: Reliability, validity and scale construction*. Retrieved 4/28/08 from <http://malroy.econ.ox.ac.uk/fisher/survey/week2.ppt>
- Guba, E. G., & Lincoln, Y. S. (1985). Fourth generation evaluation. Newbury Park, CA: Sage Publications.
- Howell, J., Miller, P., Hee Park, H., Sattler, D., Schack, T., Sperry, E., Widhalm, S., & Palmquist, M. (2005). *Reliability and validity*. Writing@CSU. Colorado State University Department of English. Retrieved 4/28/08 from <http://writing.colostate.edu/guides/research/relval/>.

References (contd.)

- Kitchenham, B. & Lawrence, S. (2002). Principles of survey research: Part 4 – questionnaire evaluation. *ACM SIGSOFT Software Engineering Notes*, 27(3), 20-23. Retrieved 4/28/08 from <http://delivery.acm.org/10.1145/640000/638580/p20-kitchenham.pdf?key1=638580&key2=1905207021&coll=GUIDE&dl=GUIDE&CFID=22360817&CFTOKEN=37007042>
- Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Russ-Eft, Darlene F. (1980). Validity and reliability in survey research (Technical report No. 15). Washington, DC: National Center for Education Statistics. Retrieved 4/28/08 from: http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/34/a3/4a.pdf
- Trochim, W. K. (2006). Research methods knowledge base. Retrieved 4/29/08 from <http://www.socialresearchmethods.net/kb/index.php>
- Walonick, D. (no date). Survival statistics: Validity and reliability. Retrieved 4/28/08 from <http://www.statpac.com/statistics-book/basics.htm#ValidityandReliability>

Contact Information

- Dan McDonald –
mcdonald@ag.arizona.edu
- Donna Peterson –
pdonna@ag.arizona.edu

Thank You!

Questions?