Determining Methods for Ensuring Accuracy in Evaluation

Whitney Ann Shakuri-Rad

Research Paper submitted for

CI 676 Program Evaluation

at Marshall University

in partial fulfillment of the requirements

for the degree of

Doctor in Education

in

Curriculum and Instruction

Ron Childress, Ed.D. and Brenda Tuckwiller, Ed.D.

Graduate School of Education and Professional Development

South Charleston, West Virginia

2013

Keywords: program, accuracy, development, evaluation

Copyright 2013 Whitney Ann Shakuri-Rad

# Abstract

Determining Methods for Ensuring Accuracy in Evaluation

 When searching for any form of information, the seeker always wishes for truth and accuracy. No one seeks a falsehood or misconception. This is the reason the research is undertaken so meticulously. If the researcher conducts an evaluation they are seeking some truth and the path to that truth may be through evaluation. Evaluation is a way to determine information or accuracy regarding a topic. The problem arises when the researcher wants to evaluate their own or another researcher’s evaluation (meta-evaluation); it becomes difficult to maintain perspective and must be undertaken with absolute care and focus. The following research aimed to determine the ways in which studies seek accuracy and sought examples where studies fell short in terms of accuracy and precision. There is a method to accurate meta-evaluation through decreased bias, increased time frames, and a decrease in basing individual’s personal evaluations on the results of others.

# Table of Contents

[Abstract 2](#_Toc354771513)

[Table of Contents 3](#_Toc354771514)

[Introduction 4](#_Toc354771515)

[Problem Statement 5](#_Toc354771516)

[Purpose Statement 5](#_Toc354771517)

[Questions 6](#_Toc354771518)

[Delimitations 6](#_Toc354771519)

[Significance of the Study 6](#_Toc354771520)

[Methods 6](#_Toc354771521)

[Review of the Literature 7](#_Toc354771522)

[Increased Accuracy versus Inaccurate Findings 7](#_Toc354771523)

[Educational Evaluation 9](#_Toc354771524)

[Bias in Evaluation 11](#_Toc354771525)

[It is All in the Detail 12](#_Toc354771526)

[Conclusions 12](#_Toc354771527)

[Discussion and Implications (discuss findings) 13](#_Toc354771528)

[Implications for Further Research 14](#_Toc354771529)

[References 15](#_Toc354771530)

Determining Methods for Ensuring Accuracy in Evaluation

# Introduction

Evaluation, in some form, has always existed. Humans have always sought to understand their surroundings, no matter how primitively. We would not understand our world or each other without an evaluative process. Even the cavemen must have compared themselves and their tools to one another, even if it is not what we think of when we consider evaluation today. Even though we have always sought some form of testing it was not until the 1900’s that we saw a more rigorous and systematic approach to evaluation, especially as it applied to education and social programs. The problem that arises with evaluating education is that you eventually start evaluating evaluations, which can create a never ending circle, filled with problems (Rossi, Lipsey, & Freeman, 2004).

The first step to an appropriate evaluation is acknowledging that we must consider the context and perspective of all parties (Shula & Cousins, 1997). If we are evaluating an evaluation then we must consider multiple stages and subjects that can be involved, which adds to the complexity of the task. How can there be consistency when the first evaluation may not have been created by the same person as the second evaluation (Darby, 2007)? A similar problem arises when two evaluations are written by the same person, because how can someone evaluate themselves? There does not seem to be a correct answer except that if we are aware of these deficiencies we can endeavor to make our original evaluation and subsequent evaluation as thorough and accurate as possible.

When considering what is possible we must also concede that evaluation can be highly subjective. As shown in Schilling and Applegate’s (2012) article about educational evaluation, how do we truly know another human’s capabilities? How do we accurately test human nature and ability when humanity itself is so subjective to a multitude of factors? Is assessment always precise enough to determine knowledge possessed or attained, if we are looking at education. How do we know if we are asking the right questions? Multiple varied evaluations may help bridge these gaps.

Multiple evaluations can lead to a higher degree of accuracy as shown in various teacher evaluations. Some states are testing new programs where teachers are evaluated in person by the principal and a peer for at least 45 minutes, submit portfolios, respond to student feedback, and continually partake in professional development. The states that are requiring these measures are looking to evaluate the efficacy of the teachers and their methods in a thorough fashion. Other states simply require that the principal evaluate the teacher once a year and there is no time requirement. It seems that the longer, more thorough evaluation would produce better results, but again, an evaluator can never be too careful. (Shakman, Riordan, Sanchez, Cook, Fournier, & Brett, & the Regional Educational Laboratory Northeast & Islands, 2012)

# Problem Statement

The problem addressed is: how do researchers evaluate their evaluations? Evaluating an evaluation may create a never-ending circle without meaningful results. A researcher must be able to determine if their evaluation is appropriate and excellent. There are, however, many roadblocks that must be overcome to ensure accuracy.

# Purpose Statement

The purpose of this study was to examine the various evaluation methods and analyze their successes and failures in regards to accurate results and authenticity in addition to determining what the most accurate way of evaluating an evaluation may be.

# Questions

Specific questions addressed in this study included:

What are the most effective evaluation models?

How do researchers evaluate their evaluations?

How do researchers aim to increase their degree of accuracy and authenticity in these evaluations?

# Delimitations

This research was not limited to educational evaluation in order to maintain broad scope of information and perspective on the topic while maintaining appropriate focus. The articles chosen dealt with the topic of evaluation and approached it from an appraisal angle, ensuring inclusion of both evaluation methods and outcome results.

# Significance of the Study

 Evaluating evaluation is incredibly important because without this perspective there is no way of knowing if we are seeking the truth. The goal of any evaluation is to determine some form of truth. In order to achieve that goal one must check and re-check how and why an evaluation is performed. One must take every step necessary to make certain that truthfulness and accuracy are maintained throughout the process.

# Methods

 EBSCOhost was used to search for journal articles relevant to the topic. Selected educational journals were evaluated and then the search was broadened to include other areas of evaluation and fields of studies.

# Review of the Literature

 The following literature review examines over 25 journal articles on the topic of evaluation or the evaluation of an evaluation. This review seeks to determine their method, accuracy, and authenticity. Studies that achieved their goal as well as those with shortcomings in terms of information or accuracy were included. The following review is organized around the following subsections: Increased Accuracy versus Inaccurate Findings, Educational Evaluation, Bias in Evaluation, and It is all in the Detail.

## Increased Accuracy versus Inaccurate Findings

There are many factors that must be considered when evaluating evaluations. Factors such as location, age, and gender must be considered. In Moksnes and Espnes’ 2011 study, the stress of adolescents was taken into consideration when evaluating data. This study focused on stress in adolescents and sought a better understanding of what these stressors are. The study also detailed how little research currently exists on the topic. Their results found that stress can have a large impact on adolescent life. In view of these results, the authors’ consideration of stress of adolescents proved to be significant, which shows the importance of small details that need to be considered when evaluating evaluations It might seem like an insignificant detail, but everything matters when looking at any piece of evaluation. Evaluations can serve many purposes from assessing test scores to the effectiveness of educational programs. These evaluations can be especially useful when trying to establish which programs are effective and which are not. All information is relevant so long as it is accurate and free of bias (Young, Denny, & Donnelly, 2012). Accurate and unbiased evaluations can help people grow and develop better performances. (Derven, 2012)

Another interesting study by Farreras and Boyle (2012) looked at how self-promotion can effect evaluations. The purpose of this study was to examine the effect of faculty self-promotion on student evaluations. The study found that the educator’s self-promotion had the potential to alter the evaluations. This article exemplifies how research can be manipulated and must be thoroughly explained so that error is decreased. The evaluations could either be greatly affected by self-promotion or it may not affect it at all, but the human and social nature of the study may lead to inaccuracy if the researchers are not careful with bias. (Farreras and Boyle, 2012)

It is also beneficial for a study to include qualitative and quantitative data in order to maintain accuracy and objectivity. When assessing a program it is important to get feedback from all involved parties, instead of just one perspective (Pogrund, 2013). Pre-tests and post-tests also increase the range of information and can improve researcher perspective (Dwyer, Hogan, and Stewart, 2012).

A researcher must also be aware of the purpose of their evaluation. Reliability and lack of bias toward the topic and sources used must be explored thoroughly. As with some teacher or administrator evaluations, the data were based on student achievement, which can involve many factors that may skew the results (Tennessee Department of Education, 2012). Some studies seek to use multiple forms of information to evaluate a teacher or administrator, which leads to a higher level of accuracy, but still may not be as accurate as needed. It is difficult to evaluate one person’s ability based on another person’s performance (Institute of Education Sciences, 2012). This observation was made in Clifford, Hansen, and Wraight’s (2012) study of principal evaluation. They suggested a need for a set system of evaluation with various types of information that was not based entirely on the performance of others. As Xu (2012) points out, there is a need for proper educational evaluation to “ensure democracy, fairness, and effectiveness.” These studies seek to address accountability, whether with teachers, administrators, or the school psychologists, but are they seeking truth or someone to blame when things do not go as planned (Morrison, 2013)?

## Educational Evaluation

Evaluations must also be thorough. As one study of Chicago’s teacher’s evaluations found “the system identified 93% of the teachers as either Superior or Excellent—at the same time that 66% of CPS schools were failing to meet state standards, suggesting a major disconnect between classroom results and classroom evaluations.” It appears that there are two very distinct sides that need to be considered. On one side teachers and administrators are held accountable for things they cannot always change, such as student ability or effort. On the other side, they are not held sufficiently accountable for what they are able to control, for example their behavior towards students or their choices in regard to the curriculum. (Sartain, Stoelinga, Brown, & Consortium on Chicago School, 2011).

Limiting data to one school or institution may also lead to inaccurate and biased results. Bowllan (2011) sought to examine a school bullying program, but only observed one school. If a researcher seeks accurate results they need to address the topic in as many ways or locations as possible, otherwise the data could be incredibly skewed. For example, Pack, White, Raczyn, and Wang (2011) conducted a similar study on bullying, but it was carried out over the course of an entire state for two years. That study produced more data and had a higher degree of accuracy and authenticity by including a larger range of schools over a longer time period. Compared to Bowllan’s study Pack, et al were able to implement a program to reduce bullying and, because of their prolonged time frame, they were able to observe the results of their study in a more detailed manner.

Rubrics are a useful tool in establishing evaluation criteria (Lipscomb, 2012). Using various types of assessment may also make the information more accurate. For example, a study that uses surveys alone will not provide as much information as a study that uses surveys, home visits, medical records, and reviews (McCabe, 2012). Service learning is another component that can be incredibly useful in evaluation if it is assessed appropriately (Zhang, Zeller, Griffith, Metcalf, Williams, Shea, & Misulis, 2011). Looking at evaluation in reverse and monitoring how an assessment might affect performance, and in turn, evaluation, is also important. For example, the type of assessment an educator uses can greatly alter a student’s actual and perceived performance (Savickiene, 2011).

When reviewing evaluations it is important to note researcher training levels and consider how that might affect the data (Davis, 2013). If the trainer does not feel comfortable conducting a study or does not thoroughly understand the topic then it could adversely affect the research. The same must be said for the participants. At the very least they must understand their role in the process in order to minimize errors.

Adams and Umbach (2012) conducted a study to evaluate response data of online surveys and examine why someone may or may not respond to an evaluation. The results were compared to traditional paper survey models. The purpose of the study was to understand why students do not respond to online surveys about courses or professors at the same rate that they would a paper survey. They found that the reasons were poor grades, apathy, disconnect from the program, and forgetfulness. The lack of responses can lead to a very high level of bias and will therefore not provide accurate data.

## Bias in Evaluation

Human emotion is another area that can lead to bias about sensitive topics. A nine year study was conducted on reducing childhood obesity through diet and exercise. (Phillips, Raczynski, West, Pulley, Bursac, & Leviton, 2013). The study aimed to evaluate how successful the program was, and acknowledged that this topic must be addressed carefully, but did not do so themselves. The researchers interviewed parents about their child’s weight and about their knowledge of the program, but the role of human emotions were not considered in the interview process. Childhood obesity may be a very sensitive topic for many individuals and when seeking to evaluate the efficacy of any program the researcher must be aware of how emotions may alter the data.

In order to alleviate the deficiencies in data on sensitive topics the researcher may alter how the project is presented to the subjects. In the example above (Phillips et al., 2013) one could present the study as an improvement program without expressing that someone may be to blame for the issue of childhood obesity. For instance, if a study seeks to end childhood obesity through diet and exercise and makes parents feel as if they are to blame, parents may resent the study or build walls against giving any useful information for fear of how others may view them. An alternative to this situation would be to develop the program, but address it more as increasing fruit and vegetable intake versus reducing obesity. Studies like this may find more accurate results if bias and emotions of the subjects, or the subject’s children are taken into consideration (Potter, Schneider, Coyle, May, Robin, & Seymour, 2011).

Another viable option for the researcher is to take a programmatic rather than individual approach. Such approach was taken by Krukowski, Perez, Bursac, Goodell, Raczynski West, & Phillips (2011) who designed a study on the dietary habits of children and decided to focus on what is provided for them while they are on school property, instead of focusing on what parents could do differently. This study was able to produce valid data about the school program, without the same degree of bias as previously mentioned studies.

## It is All in the Detail

When evaluating an outside evaluation it is important to describe the evaluation in great detail. Doabler, Fien, Walker, and Baker (2012) write about evaluating several math classes and using NAEP (National Assessment of Educational Progress) scores to support their idea that the classes need better math textbooks to help raise the scores. They do not, however, adequately describe the test itself, which may lead the reader to dismiss the findings, or be skeptical of the results. The authors needed to provide more details about the assessment they were using to make their study more credible.

 Another way that researchers can ensure that their evaluations are more accurate is to search for the same information while conducting two studies. The researcher can use two separate processes to collect data or evaluate research, using dual processes may lead to a better understanding of the topic (Froggatt and Hockley, 2011). Self-evaluation is also a possibility when conducting research, but it must be done very carefully. People may know themselves best, but they do not always see themselves best (Hewitt, 2011).

# Conclusions

 The most important aspect of evaluation is knowing your subject, knowing what you want to accomplish, being thorough, and then re-evaluating your evaluation so that you achieve the highest level of accuracy. The most effective evaluations models are the ones the take the time and effort to thoroughly search for the information and then check and re-check that the study is as accurate and free of bias as possible. Researchers that wish to evaluate their own evaluation must try even harder to address their own bias and then seek to avoid that bias if they wish to have an accurate meta-evaluation. If a researcher truly wants to increase their accuracy and authenticity they should seek outside individuals to evaluate their evaluation. This is the only way to avoid their bias. Evaluation is simply a search for information, and when we search for information we want it to be thorough and accurate. The best way to do this is to check and check again. A researcher may also use multiple types of evaluation to increase accuracy. There are many ways to conduct an evaluation and many possible subjects, the most important thing is to be mindful at all times of the research and avoid any bias possible.

# Discussion and Implications

A background in education may automatically draw a researcher to all education articles, and that constituted the biggest section of this literature review, but it was also prudent to examine how evaluation is carried out in other countries and other subjects. Broadening the search appears to increase understanding in regard to where evaluation is going, how to conduct a proper evaluation, and how to evaluate that evaluation. It came to be viewed as a never ending circle where researchers are searching for the truth, with the truth being ever-changing.

 Based on the literature it appears that the most accurate way for a research to have their evaluations evaluated is to seek an outside reviewer. Meta-evaluation is possible and can be done satisfactorily, but it is difficult to work past the bias of the research to see if the original evaluation is useful. Research and evaluation are difficult because of their transformative aspects and possibilities, add to that human error and there is a greater chance for problem or inaccurate data. Not all evaluations have to be evaluated, but when it is necessary the original researcher must make certain that all bias is eradicated, no matter how that is achieved.

# Implications for Further Research

If this research were to be continued it might be beneficial to keep searching other fields of study and other countries. In terms of further research it appears that the study of evaluation will always grow and expand because we cannot have research without evaluation. There may not be a big change or discovery in evaluation, rather evaluation will reveal something new in a certain field or study.

#

# References

Adams, M., & Umbach, P. (2012). Nonresponse and online student evaluations of teaching:

 Understanding the influence of salience, fatigue, and academic environments. *Research*

 *In Higher Education*, 53(5), 576-591.

Bowllan, N. M. (2011). Implementation and evaluation of a comprehensive, school-wide

 bullying prevention program in an urban/suburban middle school. *Journal of School*

 *Health*, 81(4), 167-173.

Clifford, M., Hansen, U., Wraight, S., & National Comprehensive Center for Teacher, Q. (2012).

 A practical guide to designing comprehensive principal evaluation systems: A tool to

 assist in the development of principal evaluation systems. *National Comprehensive*

 *Center For Teacher Quality*.

Darby, J. A. (2007). Evaluating course evaluations: the need to establish what is being

 measured. *Assessment & Evaluation in Higher Education*, 32(4), 441-455.

Davis, S., Gervin, D., White, G., Williams, A., Taylor, A., & McGriff, E. (2013). Bridging the

gap between research, evaluation, and evidence-based practice. *Journal Of Social*

*Work Education*, 49(1), 16-29.

Derven, M. (2012). Building a strategic approach to learning evaluation. *T+D*, 66(11), 54-57.

Doabler, C. T., Fien, H., Nelson-Walker, N. J., & Baker, S. K. (2012). Evaluating three

 elementary mathematics programs for presence of eight research-based instructional

design principles. *Learning Disability Quarterly*, 35(4), 200-211.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2012). An evaluation of argument mapping as a

 method of enhancing critical thinking performance in e-learning

 environments. *Metacognition And Learning*, 7(3), 219-244.

Farreras, I. G., & Boyle, R. W. (2012). The effect of faculty self-promotion on student

 evaluations of teaching. *College Student Journal*, 46(2), 314-322.

Froggatt, K., & Hockley, J. (2011). Action research in palliative care: Defining an evaluation

 methodology. *Palliative Medicine*, 25(8), 782-787.

Hewitt, M. P. (2011). The impact of self-evaluation instruction on student self-evaluation,

music performance, and self-evaluation accuracy. *Journal Of Research In Music*

 *Education*, 59(1), 6-20.

Krukowski, R., Perez, A., Bursac, Z., Goodell, M., Raczynski J., West, D., & Phillips, M.

 (2011). Development and evaluation of the school cafeteria nutrition assessment

 measures. *Journal Of School Health*, 81(8), 431-436.

Institute of Education Sciences, (2012). Learning from recent advances in measuring teacher

effectiveness. Meeting Summary (Washington, DC, August 9, 2012). *Institute Of*

*Education Sciences*.

Lipscomb, S., Chiang, H., Gill, B., & Mathematica Policy Research, I. c. (2012). Value-added

 estimates for phase 1 of the Pennsylvania teacher and principal evaluation pilot. Full

 Report. *Mathematica Policy Research, Inc*.

Mccabe, B., Potash, D., Omohundro, E., & Taylor, C. (2012). Design and implementation of an

integrated, continuous evaluation, and quality improvement system for a state-based

home-visiting program. *Maternal & Child Health Journal*,16(7), 1385-1400.

Moksnes, U. K., & Espnes, G. A. (2011). Evaluation of the Norwegian version of the

 Adolescent Stress Questionnaire (ASQ-N): Factorial validity across

samples*. Scandinavian Journal Of Psychology*, 52(6), 601-608.

Morrison, J. Q. (2013). Performance evaluation and accountability for school psychologists:

 Challenges and opportunities. *Psychology in The Schools*, 50(3), 314-324.

Pack, C., White, A., Raczynski, K., & Wang, A. (2011). Evaluation of the safe school

 ambassador’s program: A student-led approach to reducing mistreatment and bullying

 in schools. *Clearing House*, 84(4), 127-133.

Phillips, M., Raczynski, J., West, D., Pulley, L., Bursac, Z., & Leviton, L. (2013). The

 evaluation of Arkansas act 1220 of 2003 to reduce childhood obesity:

 Conceptualization, design, and special challenges. *American Journal Of Community*

 *Psychology*, 51(1/2), 289-298.

Pogrund, R. L., Darst, S., & Boland, T. (2013). Evaluation study of short-term programs at a

residential school for students who are blind and visually impaired. *Journal of Visual*

*Impairment & Blindness*, 107(1), 30-42.

Potter, S. C., Schneider, D., Coyle, K. K., May G., Robin, L., & Seymour, J. (2011). What

 works? Process evaluation of a school-based fruit and vegetable distribution program in

 Mississippi. *Journal Of School Health*, 81(4)

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: a systematic approach* (7th

 ed.). Thousand Oaks, CA: Sage.

Sartain, L., Stoelinga, S., Brown, E. R., & Consortium on Chicago School, R. (2011). Rethinking

 teacher evaluation in Chicago: Lessons learned from classroom observations,

 principal-teacher conferences, and district implementation. *Research*

 *Report*. Consortium on Chicago School Research.

Savickiene, I. (2011). Designing of student learning achievement evaluation. *Quality Of*

 *Higher Education*, 874-93.

Schilling, K., & Applegate, R. (2012). Best methods for evaluating educational impact: a

 comparison of the efficacy of commonly used measures of library instruction. *Journal Of*

 *The Medical Library Association*, 100(4), 258-269.

Shakman, K., Riordan, J., Sanchez, M., Cook, K., Fournier, R., Brett, J., & Regional Educational

 Laboratory Northeast & Islands, (. (2012). An examination of performance-based

 teacher evaluation systems in five states. Summary. *Issues & Answers*. REL 2012-No.

 129. Regional Educational Laboratory Northeast & Islands,

Shulha, L. M., & Cousins, J. (1997, Fall97). Evaluation use: Theory, research, and practice since

 1986. *Evaluation Practice*. p. 195.

Tennessee Department of, E. (2012). Teacher evaluation in Tennessee: A report on year 1

implementation. *Tennessee Department Of Education*.

Xu, Y. (2012). Developing a comprehensive teaching evaluation system for foundation courses

 with enhanced validity and reliability. *Educational Technology Research &*

 *Development*, 60(5), 821-837.

Young, M., Denny, G., & Donnelly, J. (2012). Lessons from the trenches: Meeting

 evaluation challenges in school health education*. Journal of School Health*, 82(11),

 528-535.

Zhang, G., Zeller, N., Griffith, R., Metcalf, D., Williams, J., Shea, C., & Misulis, K. (2011).

 Using the context, input, process, and product evaluation model (CIPP) as a

 comprehensive framework to guide the planning, implementation, and assessment of

 service-learning programs. *Journal Of Higher Education Outreach And*

 *Engagement*, 15(4), 57-84.