

6-28-2019

## Quantitative or Qualitative: Selecting the Right Methodological Approach for Credible Evidence

Kenneth R. Jones  
*University of Kentucky, krjone3@email.uky.edu*

Eugenia P. Gwynn  
*North Carolina A&T State University, egwynn@ncsu.edu*

Allison Teeter  
*Kansas State University, amt8968@ksu.edu*

Follow this and additional works at: <https://scholarsjunction.msstate.edu/jhse>



Part of the [Social and Behavioral Sciences Commons](#)

---

### Recommended Citation

Jones, K. R., Gwynn, E. P., & Teeter, A. (2019). Quantitative or Qualitative: Selecting the Right Methodological Approach for Credible Evidence. *Journal of Human Sciences and Extension*, 7(2), 5. <https://scholarsjunction.msstate.edu/jhse/vol7/iss2/5>

This Original Research is brought to you for free and open access by Scholars Junction. It has been accepted for inclusion in *Journal of Human Sciences and Extension* by an authorized editor of Scholars Junction. For more information, please contact [scholcomm@msstate.libanswers.com](mailto:scholcomm@msstate.libanswers.com).

## Quantitative or Qualitative: Selecting the Right Methodological Approach for Credible Evidence

**Kenneth R. Jones**

*University of Kentucky*

**Eugenia P. Gwynn**

*North Carolina A&T State University*

**Allison M. Teeter**

*Kansas State University*

*This article provides insight into how an adequate approach to selecting methods can establish credible and actionable evidence. The authors offer strategies to effectively support Extension professionals, including program developers and evaluators, in being more deliberate when selecting appropriate qualitative and quantitative methods. In addition, several examples of commonly used measures are described to help in determining their applicability for evaluating Extension programs. Benefits and challenges of select methods are discussed as well as pitfalls to avoid that can derail the evaluative process. Lastly, a few cases are shared to present how Extension is aiming to establish credible evidence through state efforts and at the national level. The authors discuss the use of practical designs (e.g., common measures) that offer a more uniform way of evaluating programs. Examples are also included to highlight the effective use of Extension reporting systems that aim to streamline data collection, evaluation, and reporting as a means to ensure more credibility.*

*Keywords:* quantitative, qualitative, mixed methods, evaluation

***“If the methods we use match the purpose of the evaluation, if we employ these methods ethically with technical competence, and if our decisions and the underlying reasoning are apparent, the evaluation will meet our ultimate goal — to produce credible and actionable evidence. Therefore, we must choose our methods wisely.”***

**—Sharon F. Rallis (2015, p. 137)**

---

Direct correspondence to Kenneth R. Jones at krjone3@email.uky.edu

## Introduction

The Cooperative Extension Service (Extension) provides communities with an array of resources, mainly through programs and projects that aim to improve the lives of local citizens (Gavazzi & Gee, 2018). Despite the wealth of knowledge provided to Extension's clientele (primarily through those working at the county level), a vast majority of the organization's professional staff have limited formal training as evaluators (Lamm, Israel, & Diehl, 2013). There is no doubt that in any organization, one will be hard-pressed to find individuals with more passion than Extension professionals who aim to generate ideas and mobilize individuals to implement change. However, there is little preparation in aiding Extension staff in the process of program development and evaluation (McClure, Fuhrman, & Morgan, 2012; Rennekamp & Arnold, 2009).

We currently live in an era where reporting and accountability of funds are paramount to sustaining programs that make a difference. Therefore, the process of evaluating programs is the crux of validating program successes that benefit individuals and communities (Mullins et al., 2015). It is important to offer the preparation that helps Extension staff determine which evaluation methods are conducive to gathering and analyzing data to demonstrate the quality and effectiveness of Extension programs. Evaluation is no longer just a worthy goal, but an organization's responsibility, to serve as a medium to improve communication and the programs themselves (Franz, 2013). It is no longer acceptable to gather a minutia of data just to "check the box." Moreover, evaluation and the evidence it provides can be the key to maintaining current and securing future funding (Franz, Arnold, & Baughman, 2014; Lamm & Israel, 2013).

### Credible Evidence: Research versus Evaluation

Applying the most suitable methods can play a powerful role in program development and evaluation (Creswell, 2003). It is important to understand the needs of what is to be evaluated. However, the key is to use methods appropriately for data collection, analysis, etc. Not only is it critical to be mindful of the need for rigor and credibility, but the distinctions between methods must also be clearly understood. Hessler (1992) described methodology as the "science and art of evaluating the worthiness" of a problem that guides research design decisions (p. 26).

While those with quantitative research backgrounds may objectively gravitate toward survey designs to gather data, qualitative methodologists may prefer using strategies that rely on personal feelings and meanings (or interpretations) to capture the essence of participants' experiences (Newman & Benz, 1998). At this juncture in science, it is important to acknowledge that one is no less standardized than the other. While quantitative methods have been revered and remain as a gold standard for achieving credible evidence, qualitative approaches should also be seen as an adequate way to assess programs and projects that aim to affect the lived experiences of clientele (Secrest & Sidani, 1995). Therefore, it is imperative to utilize the best methods to achieve the desired goals and results.

While evaluators should be aware of key research concepts, researchers should, in turn, have the wherewithal to package research theory into a form that is understood by consumers and lay audiences (Fink, 2015; Garbarino & Holland, 2009). Clientele are more inclined to use research results that are not scholarly rhetoric but clearly articulate how their issues can be resolved (McDavid & Hawthorn, 2006). While research aims to prove (through testing and further developing) theory, evaluation strives to improve (programs, etc.)—hence, the reason to have a solid understanding of the relational similarities and contrasts between research and evaluation. Many research fundamentals apply to evaluation practices, but the implementation of research and evaluation designs may vary due to the research agenda, evaluation purpose, or intended benefits to subjects/clientele. However, both research and evaluation offer criteria that stress the importance of credibility (and ethics), whether as a researcher or evaluation specialist. Table 1 provides a synopsis describing some differences between evaluation and research.

**Table 1. Differences Between Evaluation and Research**

	<b>Evaluation</b>	<b>Research</b>
<b>What is the purpose?</b>	<ul style="list-style-type: none"> <li>• Make value statements about merit or worth</li> <li>• Provide information for decision making</li> </ul>	<ul style="list-style-type: none"> <li>• Add to knowledge in the field</li> <li>• Develop laws and theories</li> </ul>
<b>Who determines the agenda or focus?</b>	<ul style="list-style-type: none"> <li>• Stakeholders and evaluator(s) jointly</li> <li>• Funding agencies</li> </ul>	<ul style="list-style-type: none"> <li>• Researchers</li> <li>• Academic institutions</li> <li>• Funding agencies</li> </ul>
<b>Is generalizability important?</b>	<ul style="list-style-type: none"> <li>• The focus is on the particulars of the program, policy and the context</li> <li>• Generalizability is less important</li> </ul>	<ul style="list-style-type: none"> <li>• Yes, it adds to theory</li> <li>• Contributes to the field</li> </ul>
<b>How are results utilized?</b>	<ul style="list-style-type: none"> <li>• Decision making (e.g., about the project’s activities, development of future projects)</li> </ul>	<ul style="list-style-type: none"> <li>• Knowledge sharing</li> <li>• Practice Improvement</li> </ul>
<b>What criteria determine credibility?</b>	<ul style="list-style-type: none"> <li>• Accuracy</li> <li>• Utility</li> <li>• Feasibility</li> <li>• Propriety</li> <li>• Accountability</li> </ul>	<ul style="list-style-type: none"> <li>• Internal validity</li> <li>• External validity</li> <li>• Generalizability</li> </ul>

*Note:* Adapted from Fitzpatrick, Sanders, and Worthen (2011).

Gathering adequate and accurate data is critical to the credible evidence of any evaluative procedure. However, with the advocacy of research models stressing approaches such as randomized controlled trials (RCTs), used in many scientific experiments to help minimize bias and increase validity in research, the work of many evaluators who focus primarily on community-based evaluations (and deal with a surplus of social external factors) may be perceived as less rigorous. In other words, those conducting basic research (typically in non-social science areas) may see the applied approach often adopted by evaluators as less scrupulous, when this work may be even more intense due to the social dynamics that can affect evaluation implementation processes.

So, what is credible when it comes to choosing the right evaluation methods? Is credibility a quantitative approach that aligns with RCTs for purposes of examining factors that affect members of separate comparison groups? Is credibility a qualitative approach to apply meaning to the experiences of participants? This article provides insight into how adequate, more appropriate evaluation methods can establish credible and actionable evidence of program impact. The authors confer a means to support Extension professionals, including evaluators and other program developers, to be more explicit in the selection of methods that deliver a usable, more credible Extension program. Despite the differences among evaluation methods, it is imperative to have a solid understanding of and rationale for using a variety of approaches. For instance, although the presentation of some findings can rely deeply on narratives, patterns, and themes (qualitative data), quantitative data may be expressed similarly or in the traditional format that uses numerical concepts (Secrest & Sidani, 1995). The authors will discuss a variety of approaches for conducting program evaluations, including quantitative, qualitative, and a combination of both (mixed methods). Moreover, the totality of possibilities must be considered when determining the most applicable method. It is important to note that the proper method is key, and the evaluator should be familiar with what is most suitable for a specific audience.

### **Why Methods Matter**

Why even entertain the importance of evaluation methods? It is indeed at the core of what solidifies the process of collecting credible and actionable evidence. A method is what an evaluator must understand when considering available resources for conducting an evaluation (Greene, 2007). Not only will this set the stage for what can be evaluated within reason, but it also affects how an evaluation process is designed. In addition, one must also consider methods when determining the target audience. For instance, an evaluator must have the skills to determine which method is more appropriate for young children as opposed to gathering data from older teens. This is in line with any stakeholder of a program (Creswell & Creswell, 2018), for stakeholders provide a source for answering pertinent questions which lead to answers and solutions. Specific program outcomes are often a driving force when considering proper methods as well. When in doubt, it is important to revisit the program's logic model, the purpose of the evaluation, and the evaluation questions. A mismatch of methods and questions will inevitably lead to incomplete and/or inaccurate information that will be viewed as less than credible. Evaluators should keep in mind that the "determination of what is credible is often context-dependent (i.e., varies across programs and stakeholders), and is naturally tied to the evaluation design, implementation, and standards adhered to for data collection, analysis, and interpretation" (Centers for Disease Control and Prevention [CDC], 2011, p. 25).

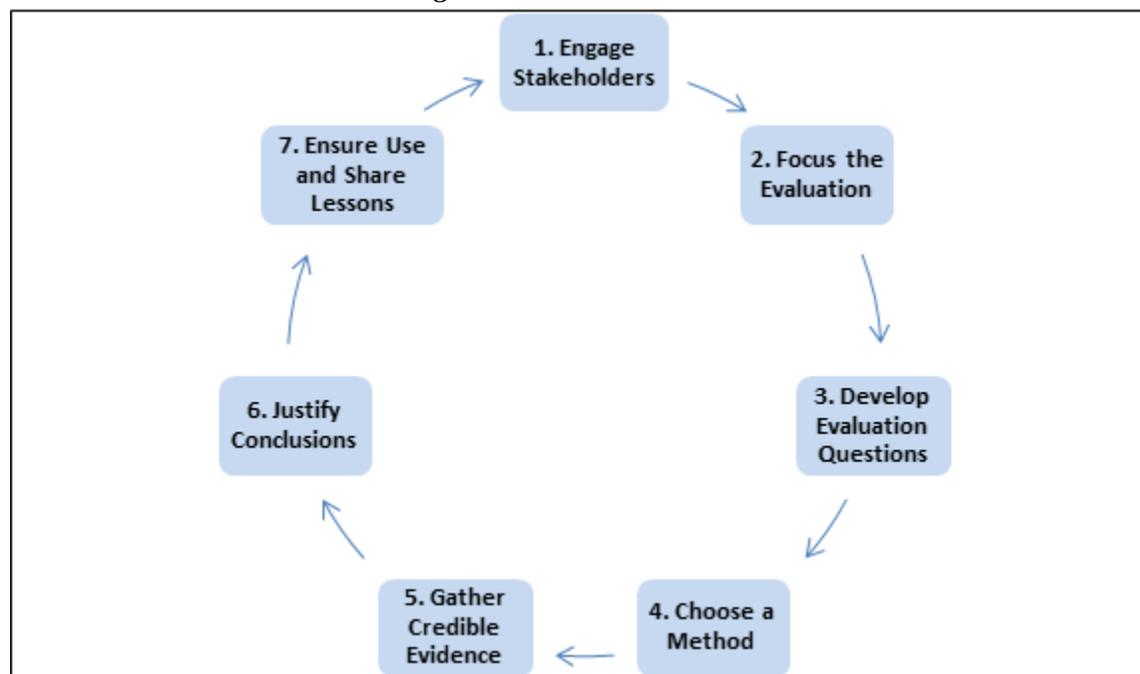
Utilizing the wrong methods could cause significant limitations on what can be interpreted from the results. Having a keen understanding of research methods allows evaluators and program participants alike, to become more astute consumers of evidence (Gooden & Berry-James, 2018).

This is, in fact, very applicable and a critical part of the decision-making process in regard to Extension program development and management.

### Choosing the Proper Evaluation Method

Producing credible and actionable evidence begins not with selecting a *preferred* method, but with selecting the most *appropriate* method (Rallis, 2015). The decision to select the best method should be informed by the evaluator's knowledge and skills, program outcomes, and stakeholder feedback as well as the evaluation's purpose and questions. In this section, the authors examine the formal evaluation process outlined in Figure 1. Emphasis will be placed on the initial four steps of the process, highlighting their importance in selecting the appropriate method (Step 4).

*Figure 1. The Evaluation Process*



*Note:* Adapted from the CDC's *Guide to Developing an Effective Evaluation Plan* (2011, p. 5).

#### Step 1: Engage Stakeholders

Stakeholders, those who have a vested interest in the program being evaluated (Mertens & Wilson, 2012), can include but are not limited to funding agencies, program participants, project collaborators (e.g., partners), and the community. The extent to which these individuals are involved in the evaluation process dictates the success of the evaluation. For example, it is suggested that a meeting with stakeholders take place before developing the evaluation plan. Doing so provides an opportunity to begin laying the groundwork for the evaluation by determining stakeholder needs and interests, asking questions, and gathering feedback. This initial meeting builds trusts, encourages transparency, and ensures that the evaluation gathers

data that will be useful to the stakeholders, thereby increasing the likelihood that the evaluation results will be utilized. Given that it may not be feasible to include all stakeholder groups in the process, special consideration should be given to including those stakeholders that enrich the credibility of an evaluation (Chawla, 2015). Having the stakeholders provide specific guidance on evaluation at this stage of the evaluation process can affect the use of the recommendations, which can ultimately influence funding decisions (Pell Institute, 2019).

## **Step 2: Focus the Evaluation**

Another crucial step in selecting an appropriate method is identifying the evaluation purpose. According to Davidson (2005), “choosing the appropriate method hinges on getting the evaluation purpose and questions right” (p. 14). If a logic model for the program does not exist, this is the ideal time to develop one. The activity can be used as a chance to build capacity among stakeholders with the already existing or newly created logic model serving as a road map for the program, thus identifying goals, activities, and expected outcomes. In essence, this formulates the action of determining the reason for evaluating a program. Engaging stakeholders in conversations regarding which components of the program’s logic model should be a focus of the evaluation, determining which stakeholder groups will be participants in the evaluation, and learning how the intended users of the evaluation will use the results are equally important (BetterEvaluation, 2013).

As an evaluator begins thinking more about the evaluation questions and methods, further consideration should be given to the purpose of the evaluation. For example, the goal of the evaluation of a program receiving funding from an external stakeholder may be oversight and compliance or accountability. For oversight and compliance purposes, data illustrating that the program is following the rules and meeting expectations will be needed. If the purpose of the evaluation is accountability, the evaluator should seek to gather data that will demonstrate programmatic impacts. Documenting the program’s accomplishments is just one way to demonstrate its effectiveness to funders. An evaluation might also aim to improve or increase the knowledge base of what constitutes an effective program (i.e., determine what works, why, and in what contexts so that programs may be replicated). Finally, many evaluations focus on program improvement. When this is the purpose of the evaluation, the evaluator should gather data that will enhance the program’s quality through the identification of ways to improve program implementation and effectiveness (Mark, Henry, & Julnes, 2000). Regardless of the purpose of the evaluation, specific goals should be clearly defined before developing evaluation questions. If not, misalignment of the evaluation purpose and corresponding questions may decrease the credibility and usefulness of evidence.

### **Step 3: Develop Evaluation Questions**

Evaluation questions assist in identifying the types of data to be collected, selecting the appropriate data collection methods to be used, and/or finding the appropriate evaluation instrument(s) to be utilized. It is important to formulate questions that, when answered, will highlight the connections between program activities and short-, medium-, and long-term outcomes (Corporation for National & Community Service, 2019). Some sample evaluation questions include:

- To what extent have program activities been implemented? Were planned program activities/outputs completed on schedule?
- What is the program doing well, and what are potential areas for improvement?
- What impact is the program having on participants? What changes in knowledge, skills, attitudes, or behaviors have occurred due to participants' participation?
- How will program activities, outputs, and outcomes be sustained beyond the funding cycle?

According to the CDC (2011), when developing evaluation questions, it is also important to keep in mind the program's stage of development as well as the program's information needs, such as what will be most useful to stakeholders. Other considerations include the feasibility of answering each question and how much time, effort, and resources will be needed to answer the questions effectively and efficiently.

### **Step 4: Choose a Method**

Once stakeholders have been engaged and a purpose and evaluation questions are in place, it is now time to select a method. In selecting a method, an evaluator must carefully consider the resources available, the type of data (i.e., quantitative or qualitative) that will need to be gathered, and how the results will be presented and used. Each method comes with specific advantages and limitations that should be taken into consideration. For example, when gathering data from program participants regarding a sensitive subject, anonymous surveys or one-on-one interviews might be most appropriate. Participants may be more willing to respond openly and honestly as opposed to potentially censoring themselves in a focus group interview. When in doubt, it may help by returning to stakeholders for support to provide additional insight into which method may yield the best results (given their own knowledge of the program and its participants).

According to Rallis (2015):

If the methods we use match the purpose of the evaluation, if we employ these methods ethically with technical competence, and if our decisions and the underlying reasoning

are apparent, the evaluation will meet our ultimate goal—to produce credible and actionable evidence. (p. 137)

That is to say, the journey toward obtaining credible evidence does not end once the stakeholders have been engaged, an evaluation purpose and questions identified, and the appropriate method selected. For this article, the authors have chosen only to focus on the initial steps of the evaluation process. To continue to ensure the data gathered will be viewed as credible, the evaluator must then go forth to find or create a credible instrument or protocol, implement the evaluation with fidelity and integrity (i.e., gather credible evidence), perform the appropriate analyses, and report the findings. Adequately justifying the findings and ensuring the results are used will also add to the credibility and success of the evaluation.

### **Types of Methods**

The use of certain methods has been at the core of the debate over credibility for some time. With quantitative methods being seen as the more common among experimental designs that infer to populations (Creswell, 2003), many evaluation approaches are heavily aligned with the expectations adorned by those who hold RCTs as a premier standard. Over the years, a plethora of research and evaluation approaches have emerged, with researchers and evaluators offering philosophical and technical reasons for the most credible methods (Creswell, 2003; Davidson, 2005; Fink, 2015; Newman & Benz, 1998). Creswell (2003) argued that to fully understand the best approach to establishing a research design, general procedures of data collection (methods) should be first and foremost. In today's society, there is less demand on quantitative over qualitative or vice versa but more foci on how the stronger, thus more credible, research studies rely on efforts lying within a continuum—providing a balance between quantitative and qualitative methods (Newman & Benz, 1998; Rallis, 2015). In essence, it is critical to use the method or methods that can render the most meaningful results.

Quantitative approaches seem to be eagerly utilized, perhaps due to the misconception that surveys are quick and easy to craft. In fact, most evaluations and research methodologists will argue that just because one is using a survey or questionnaire does not make the data credible (Fowler, 2002; Kelley, Clark, Brown, & Sitzia, 2003). There are “assessments of the assessment” that should occur before anyone uses an instrument for data collection. From a conventional pilot testing to more sophisticated statistical modeling, intentional steps should be taken before any instrument is put forth and trusted by unassuming novice evaluators. Hence, there are tried and true ways of determining whether data collection tools have credibility and can promote adequate use for the evaluator and responders providing data (Presser et al., 2004). This is often seen as the responsibility of the evaluator, particularly those serving as Extension evaluation specialists. For example, having a clear understanding of how item response theory (IRT) models can determine the degree to which certain items differ in meaning among respondents is a rudimentary, but handy skill to attain (see Carlson & Davier, 2013). Gaining the

capacity to identify and determine the worth of specific items can help in designing instruments that more adequately measure meaningful constructs.

Both quantitative and qualitative methods can pose valuable questions that result in the creation of credible and actionable evidence. Answers to these questions, whether from clientele or other stakeholders, aid in providing information that is taken as accurate or truth (Donaldson, Christie, & Mark, 2015). Both quantitative and qualitative methods can be used in conducting formative and summative evaluations. Formative evaluations would be the assessment of components through the process (e.g., determining if the time of the program is adequate or if individuals are actively participating), while summative evaluations would focus on the end results (e.g., Did participants experience a change in behavior based on what they learned?). Both methods can also help to provide meaning to evaluations in determining the effectiveness of a program.

### **Quantitative Methods**

Quantitative data deal with variables that can be measured with numeric values. They are best used to answer questions such as “How many?”, “What were the outcomes?”, and “How much did it cost?”. Within an Extension context, quantitative data might be used to answer questions such as “How many youth with no familial history of 4-H participation joined 4-H in a given year?” or “What was the net gain in monthly income for limited-resource farmers who used high tunnels?” Quantitative data may involve statistical analyses that range from basic frequencies to more complex group differences, relationships, and causal estimates and projections. Credible evidence, however, does not always require complex statistical analyses, and it may be best to present simplified analyses (e.g., simple counts) that are practical and actionable for those working at the grassroots level. For example, county commissioners might primarily be interested in knowing how many of the county’s residents have participated in a particular program and whether or not they felt the program was beneficial.

Quantitative data are often collected through methods such as surveys or questionnaires (which includes pre- and post-program tests or surveys), reviews of existing documents and databases, or by gathering clinical data.

**Surveys/questionnaires.** One of the most commonly used quantitative data collection methods is surveys or questionnaires. Surveys can be self-administered or administered by someone else and conducted either face-to-face, by mail, by telephone, or online. When using surveys and questionnaires, the reliability (i.e., the extent to which the instrument produces consistent outcomes) and validity (i.e., the extent to which the instrument measures what it is intended to measure) must be taken into account before its use. Pre- and post-surveys and/or tests are examples of survey tools that are used to document changes in knowledge, attitudes, skills, motivations, and behaviors. Typically, program participants take a pre-test or survey, receive the intervention or program, and then are tested or surveyed after the fact. The difference between the pre- and post-measurements represent change attributed to the program or activity.

Variations of the pre-posttest design include posttest only, retrospective pre- and posttest, pre- and posttest with comparison group, pre- and posttest with follow-up, and intermediate testing and posttest (Bennett, 1984). It should be noted, however, that surveys/questionnaires can also contain open-ended responses (e.g., “What did you learn about healthy food choices?”) that can make them qualitative in nature.

**Surveillance data.** The Centers for Disease Control and Prevention defines public health surveillance as “ongoing, systematic collection, analysis, interpretation, and dissemination of data regarding a health-related event for use in public health action to reduce morbidity and mortality and to improve health” (Rolka et al., 2012). Evaluating health surveillance systems ensures the effective and efficient monitoring of public health problems (Teutsch & Churchill, 2000). Because each surveillance system is unique, evaluating them is not an easy process and requires flexibility (Calba et al., 2015). Typically, surveillance information is analyzed by time, place, and person, using tables and graphs to summarize and present data (Nsubuga et al., 2006). Examples of surveillance systems at the local, state, and national levels include vital statistics such as deaths and births, disease reporting including HIV prevalence, and surveys (e.g., Youth Risk Behavior Survey; YBRS). While it is not a typical Extension practice, Extension programs are often centered on public health topics such as chronic disease prevention, obesity, or the opioid epidemic. Therefore, it is worth considering a partnership with other organizations such as health departments to learn about gathering credible and actionable evidence through the use of surveillance systems.

**Record reviews.** There may not always be a need to collect new data to evaluate a program. In this case, document or record reviews can be used. Documents can be internal (e.g., attendance sheets for a workshop on financial management) or external (e.g., government agency report) and allow one to evaluate a program with minimal disruption (National Minority Aids Council, n.d.). Information from documents are useful for gathering background information, determining the success of program implementation, assisting in formulating questions for surveys or focus group protocols, and answering “what” and “how many,” such as the number and types of program participants and program costs. It should be noted that while record reviews can provide numerical data, they can also provide qualitative information as well. Data can be gathered from a variety of sources such as exit reports, meeting minutes, newsletters, and/or marketing materials.

The chosen quantitative method will depend on a variety of factors, such as cost and the amount of time one has to conduct data collection and analysis. In addition, due to Extension’s limited capacity to have dedicated staff to evaluate each individual program in a meaningful and appropriate way, the required resources of the methods used must be considered. Table 2 provides a more detailed comparison of some of these factors.

**Table 2. Required Resources, Advantages, and Challenges of Quantitative Methods**

<b>Method (Required Resources)</b>	<b>Advantages</b>	<b>Challenges</b>
<b>Surveys/ Questionnaires (Low)</b>	<ul style="list-style-type: none"> <li>• Inexpensive to administer</li> <li>• Can be used to gather large amounts of data</li> <li>• Easy to analyze quantitative data statistically</li> <li>• Reliable and valid measures may already exist</li> <li>• Pre- and posttests offer better evidence of program effectiveness compared to other methods</li> <li>• Posttests with follow up provide valuable information about medium and long-term impacts</li> </ul>	<ul style="list-style-type: none"> <li>• Inability to capture the full story</li> <li>• Question wording can bias respondent's answers</li> <li>• Prone to error</li> <li>• Tracking and contacting participants for follow-up can be time-consuming</li> <li>• For retrospective pre- and posttest designs, may be difficult for participants to remember how they thought/behaved before the program</li> </ul>
<b>Surveillance data (Moderate-High)</b>	<ul style="list-style-type: none"> <li>• Greater awareness of potential threats</li> <li>• Ability to collect detailed information</li> <li>• May be more representative</li> </ul>	<ul style="list-style-type: none"> <li>• Expensive</li> <li>• Labor intensive</li> <li>• Difficult to sustain over time</li> <li>• Possibility of misuse of information</li> </ul>
<b>Record reviews (Moderate)</b>	<ul style="list-style-type: none"> <li>• Provides comprehensive and historical information</li> <li>• Minimal disruption to program or activity</li> <li>• Information is readily available</li> </ul>	<ul style="list-style-type: none"> <li>• Time-consuming</li> <li>• Information may be incomplete</li> <li>• Data are restricted to what already exists</li> <li>• Need to be clear beforehand about what data are being sought</li> </ul>

*Note:* Adapted from Siebold (2011).

Quantitative approaches, regardless of the instrument used, should measure the depth and breadth of implementation (e.g., the number of people who participated, the number of people who completed the program, attitudinal constructs, knowledge and behavior changes). The strengths of quantitative data for evaluation purposes include their representativeness (if the sample represents the population), the ease of analysis, and their consistency and precision (if collected reliably). However, the limitations of using quantitative data for evaluation can include poor response rates from surveys, difficulty obtaining documents, and difficulties in valid measurement (Driscoll, 2011). In addition, quantitative data do not provide an understanding of the program's context and may not be robust enough to explain complex issues or interactions (Garbarino & Holland, 2009; Holland & Campbell, 2005).

Presser and colleagues (2004) noted that questionnaire design and statistical modeling are usually perceived as priorities on separate ends of the spectrum. In other words, problems occur when either individuals with questionnaire design expertise do not have adequate knowledge of the appropriate use of statistical analyses or individuals with statistical analysis expertise do not have some understanding or familiarity with survey question/item design. The result of either

situation can be invalid and/or unreliable results. Thus, the case for understanding when to use qualitative methods and when to use quantitative approaches is warranted.

### **Qualitative Methods**

In contrast to quantitative methods, qualitative data cannot easily be converted into numbers or used for aggregating data. More specifically, a user of qualitative methods must be adept in forming interpretations to not only provide rich descriptions of complex phenomena but in constructing themes or conceptual frameworks as well as generating hypotheses (Bickman & Reich, 2015; Foley & Timonen, 2014). Qualitative data are helpful for understanding how participants felt about a program, what they experienced, or why a program was useful. This method is best suited for probing open-ended questions such as “What was the value added?”, “Who was responsible?”, and “When did something happen?”. Within an Extension framework, questions such as “How have your dietary habits changed as a result of your participation in the SNAP-Ed (Supplemental Nutrition Assistance Program Education) program?” would be best answered through the use of qualitative methods.

Qualitative data analysis will likely include the identification of themes, coding, clustering similar response data, and reducing data to meaningful and important points, such as in grounded theory-building or other approaches to qualitative analysis (Patton, 2002). Grounded-theory, developed by Glaser and Strauss (1967), is “a systematic method for constructing a theoretical analysis from data, with explicit analytic strategies and implicit guidelines for data collection” (Charmaz & Belgrave, 2012, p. 347). Using a grounded-theory approach creates meaning from data that are coded using categories and subcategories.

The most common qualitative data collection methods include focus groups, observation, interviews, and case studies.

**Focus groups.** Focus groups are comprised of small groups of people (usually 8-12) who share some characteristics or relevant experience and ideally do not know each other (Kreuger & Casey, 2015). Focus group participants discuss ideas and insights in response to open-ended questions from a facilitator. Group dynamics are also used to help generate data through themes.

**Observation.** Marshall and Rossman (1989) define observation as “the systematic description of events, behaviors, and artifacts in the social setting chosen for study” (p. 79). Observations may help explain behaviors as well as social context and meanings because the evaluator sees what is actually happening. Observations can include watching a participant or program, videotaping an intervention, or even recording people who have been asked to “think aloud” while they work (Ericsson, Krampe, & Tesch-Römer, 1993). The types of observations range from the *complete observer* who is neither seen nor noticed by the participants to the *complete participant* who is fully engaged with those who are under observation (Creswell & Creswell, 2018).

**Interviews.** Interviews are useful for complex or sensitive subjects. Conducted at the individual level, they often provide rich data, details, and perspectives from program participants and stakeholders regarding their experiences, behaviors, and opinions. Interviews may be structured and conducted under controlled conditions, or they may be conducted with a loose set of questions asked in an open-ended manner. When gathering demographic data, such as age, interview questions can also be quantitative in nature.

**Case study.** According to the U.S. General Accounting Office (U.S. GAO, 1987), a case study is a “method for learning about a complex instance, based on a comprehensive understanding of that instance obtained through extensive description and analysis of that instance taken as a whole and in its context” (p. 14). The purpose of a case study is to intently examine a particular unit (person, site, project) as a distinct whole. Case studies can be helpful for understanding how different elements (implementation, context, and other factors) fit together and produce the observed impacts. The U.S. GAO (1987) has identified six types of case studies: illustrative (intended to add realism about a program or policy), exploratory (aimed at generating hypotheses), critical instance (examines a single instance of unique interest), program implementation (investigates operations, often at several sites), program effects (examines causality), and cumulative (brings together findings from many case studies to answer an evaluation question).

Table 3 highlights the advantages, challenges, and required capacity for using these qualitative methods.

**Table 3. Required Resources, Advantages, and Challenges of Qualitative Methods**

<b>Method (Required Resources)</b>	<b>Advantages</b>	<b>Challenges</b>
<b>Focus groups (Moderate)</b>	<ul style="list-style-type: none"> <li>• Participants define what is important</li> <li>• Opportunity to clarify responses through probes</li> <li>• Less expensive and more efficient than interviews</li> <li>• Provides immediate sharing and syntheses</li> </ul>	<ul style="list-style-type: none"> <li>• Requires skilled facilitators</li> <li>• Lack of confidentiality</li> <li>• May be difficult to analyze</li> <li>• Group members and facilitators can bias responses</li> <li>• Time-consuming to conduct and analyze data</li> </ul>
<b>Observation (High)</b>	<ul style="list-style-type: none"> <li>• Can adapt to events as they occur</li> <li>• Setting is natural, flexible, and unstructured</li> <li>• Ability for researcher/evaluator to choose participation level (i.e., actively participate vs. passively observe)</li> </ul>	<ul style="list-style-type: none"> <li>• Can be difficult to interpret seen behaviors</li> <li>• Presence of researcher/evaluator may influence behaviors of participants</li> <li>• Difficult to generalize findings to entire population</li> <li>• Not realistic for large groups</li> <li>• Time-consuming to reliably train observers</li> </ul>

<b>Method (Required Resources)</b>	<b>Advantages</b>	<b>Challenges</b>
<b>Interviews (Moderate- High)</b>	<ul style="list-style-type: none"> <li>• May be easier to reach specific individuals (e.g., homeless persons, victims of domestic violence)</li> <li>• More personalized approach</li> <li>• Easier to ask open-ended questions, use probes, and pick up on nonverbal cues</li> <li>• Line of questioning can be tailored to the individual</li> </ul>	<ul style="list-style-type: none"> <li>• Can be time-consuming and/or difficult to schedule</li> <li>• Requires skilled/trained interviewer(s)</li> <li>• May be difficult to analyze and summarize findings</li> </ul>
<b>Case Study (High)</b>	<ul style="list-style-type: none"> <li>• Allows for the collection of rich details</li> <li>• Helps detect unexpected outcomes</li> <li>• Can help produce novel hypotheses for later testing</li> </ul>	<ul style="list-style-type: none"> <li>• Time-consuming to collect, organize and describe</li> <li>• Represents depth of information, rather than breadth</li> <li>• Data cannot necessarily be generalized to the wider population</li> <li>• Difficult to draw cause/effect</li> </ul>

*Note:* Adapted from Siebold (2011).

Due to the dominance of surveys as a means to gather data, practical interview approaches, documented observations, and valuable focus group discussions often go overlooked as applicable sources of data that can demonstrate high-level impact. In fact, many of these techniques can offer insight that quantitative survey data do not reveal. One of the strengths of qualitative methods is that individuals can provide ideas and stimulate memories with topics cascading as the qualitative discussion occurs (Morgan, 1997). Moreover, these methods (observations, for example) may help explain behaviors as well as social contexts and meanings witnessed by the evaluator as they actually happen (Ericsson et al., 1993). Participants are given a chance to discuss their ideas and insights in response to open-ended questions from the facilitator rather than being limited to the choices on a survey instrument. Indeed, flexibility is a key difference between quantitative and qualitative methods, with qualitative methods allowing for greater spontaneity between the evaluator and program participants (Creswell & Creswell, 2018).

However, despite their advantages, there are limitations to qualitative methods. Data collected through qualitative processes are challenging for inexperienced evaluators to analyze. Not fitting into standard categories, qualitative data collection and analyses are generally time-consuming and costly. The data associated with qualitative methods are also very subjective, thus eliminating the more concrete interpretations that are often afforded through quantitative analyses (Patton, 2002).

## Mixed Methods

There are instances, however, when quantitative and qualitative methods are combined to produce a richer and more comprehensive understanding of a program or project's impact. This technique, known as mixed methods, allows the evaluator to bring quantitative and qualitative approaches together in a study.

Creswell and Plano-Clark (2011) have identified three basic mixed methods approaches in evaluation: convergent parallel design, explanatory sequential design, and exploratory sequential design.

The convergent parallel design is primarily used for validity; that is, determining if the results are similar when using both quantitative and qualitative methods. This design would call for using a survey (quantitative) and a focus group (qualitative) concurrently with similar participants. Results are merged for comparison and then interpreted to explain convergence/divergence.

In the explanatory sequential design, a qualitative method is used to explain "blindsides" results from a quantitative method. A survey is followed by a focus group, interviews, or observation to explain or better understand what is happening in the quantitative results.

The exploratory sequential design explores potential patterns with a qualitative method and then verifies the patterns with a quantitative follow-up.

Deciding to use a mixed methods approach is a deliberate design decision. When choosing a mixed methods design approach, evaluators should consider several things, such as the stakeholder needs and wants; resources such as time, skill, and funding; and/or the complexity of the design (Creswell, 2003). Considering stakeholder needs and wants are essential to buy-in and success of the evaluation. The method(s) chosen should be realistic, given one's timeframe and budget for completion. In addition, more complex designs will require additional resources.

If both the know-how and the resources are available, using a mixed methods approach may be more effective since it allows for triangulation of findings. Triangulation allows one to tap into multiple ways to gather and ultimately analyze data, thus offering more credibility to actionable data (Greene, 2007). Furthermore, by using both quantitative and qualitative methods, the unique strengths of each approach can offset their respective limitations when used alone, thereby also increasing the level of credibility of the resulting data (Creswell & Creswell, 2018).

A mixed methods approach calls not only for the proper selection of various methods, but consideration must also be given to how data collection will occur. An evaluator must understand that whether collecting data sequentially (in phases) or concurrently, a personal choice of implementation could greatly affect the evaluation process (Creswell & Creswell, 2018).

Using mixed methods can offer a level of complexity that may be a challenge for novice evaluators and researchers, but on the other hand, may provide a level of objectivity that offers more quality evaluation results for stakeholders, assuming that steps have been taken to assess the validity and reliability of the data (Creswell, 2003). While it is beyond the scope of this article to provide a detailed description of each of the identified methods of data collection (both quantitative and qualitative), the reader is directed to additional resources such as Creswell and Creswell (2018) and Leavy (2017) for a more comprehensive review of both quantitative and qualitative methods.

### **Scientific Rigor**

Marquart (2017) defines scientific rigor as “the precision of a study in terms of planning data collection, analysis, and reporting” (p. 1). Simply put, rigor means to follow the required techniques and strategies for increasing both credibility (i.e., our trust and confidence in the research findings) and quality. For quantitative methods, validity and reliability are the golden standards of rigor (Coryn, 2007). The use of valid and reliable measures is crucial for both quality and credibility.

While quantitative methods call for the use of reliable and valid measures to create credible evidence, evaluators also aim to design and incorporate methodological strategies to ensure the credibility of the findings obtained from qualitative data. One strategy suggested by Creswell (1998) is to engage with at least one other individual to ensure that alternative interpretations of the data have been considered (Creswell, 1998). Lincoln and Guba (1985) provided several additional strategies for ensuring the credibility of qualitative data. These include:

- Maintaining a meticulous record of all decisions made throughout the process to ensure that data are consistent and transparent;
- Utilizing code-recode procedures as well as interrater reliability of the coding scheme with a Kappa statistic;
- Using more than one method to collect data on the same topic (i.e., triangulation) which allows for testing the consistency of findings obtained through different instruments; and
- Allowing participants the opportunity to respond to first drafts of reports to check for accuracy.

The importance of scientific rigor cannot be overstated and can be achieved with a variety of strategies not limited to the ones listed above (Santasier & Plack, 2007).

### **Methods and Misused Approaches**

Alkin and King (2017) provided interesting insight into the often misunderstood and misused approaches regarding evaluation methods. They argue that an inadequate evaluation can raise

major concerns of validity, which can cause credibility to wane. Alkin and King (2017) also describe a flaw in evaluation methods as similar to the perils associated with medical malpractice, potentially causing harm to those the evaluation intended to aid. Indeed, the methods selected have an influence on any evaluation, for they are the center of determining from whom to gather data and for whom the benefits of the evaluation will be rendered.

It is also important for an evaluator to adhere to certain ethical standards, such as the American Evaluation Association's (AEA) Guiding Principles for Evaluators (AEA, 2004), to stay clear of the tension that abides when working with stakeholders who are opposed to negative findings. In some cases, evaluators become very engaged in a program and can become biased toward a desire to present the results of the program in a more positive light than actually was the case, ultimately leading to an offering of biased evidence. Hence, the reason it is crucial to have a clear understanding of various methods that can be used to authenticate evaluative approaches for diverse clientele and communities. Having clear knowledge of the various evaluation methods can lead to the selection of methods that are appropriate for providing evidence to communities that the data were collected and analyzed in a way that minimizes bias.

Methodological decisions are not about selecting methods that may inherently buy into the decisions of the status quo (Alkin & King, 2017), but using those methods that will produce evaluation results that provide impact beyond individuals and the programs that serve them. A key step in the evaluation process is to provide additional opportunities for evaluators to not only engage users but also to educate them on the specific steps and procedures (Lamm, 2010; Lamm & Israel, 2013). This provides credible evidence for stakeholders and establishes credibility with those who will ultimately decide to use, misuse, or discard the results.

It is important to consider that data are what individuals and communities utilize to understand and ultimately make decisions to enhance the world around them. Moreover, they have a belief that the evidence will work to improve their own lives. In turn, data are important in helping to develop impact statements or success stories that Extension can use to communicate desired results to stakeholders. This lends credence to what Mark (2015) describes as the actionability of evidence. In other words, how relevant the outcomes are to real-life situations is a determining factor for most stakeholders.

Evaluation data are often presented to or read by a wide variety of audiences, many of whom are not trained researchers or evaluators. Different stakeholders will find different types of data more convincing than other types. For example, some individuals find quantitative data (e.g., means, frequencies, distributions) to be the most convincing way to express a project's impact. Others find qualitative data, such as participants' stories, more compelling. There are also different interest levels and amounts of time available for stakeholders to consume evaluation information. It usually takes less time and effort to review and interpret quantitative data than qualitative data.

Regardless of the methods used, findings must be communicated in understandable terms to ensure use of the evaluation results and lessons learned. Credible and actionable information is important, but this is often dependent on information being accessible and not overly technical for lay audiences. For instance, providing younger clientele who want to become more aware of resources for teens may be more user-friendly as a 1- or 2-page infographic than as a 20-page document. A busy elected official may also be more willing to read a brief summary that highlights a success story in her or his district than having to sit down for a long presentation to go through dozens of slides. Selecting applicable methods can aid in this endeavor.

No one evaluator's approach is the same nor will all questions posed to determine credible evidence be answered. Evaluation methods have roles to play and can make different contributions to the program evaluation process (ODAREACYF, 2016). The existing philosophical differences over which evaluation approaches are most suitable remains today. The answer still holds true, especially for Extension: which method is used depends on what we want to know and if the right questions are being asked to get the right information (National Research Council and Institute of Medicine, 2002; Patton, 2002; Secrest & Sidani, 1995). It is not necessary for every program evaluation to include highly technical qualitative, quantitative, or mixed methods. It should be within an evaluator's purview to be competent in determining which method is befitting of the questions at the inception of a program's development. Credibility is central in guiding the evaluation process from start to completion. What is most important is that despite the method of inquiry, it is implemented with rigor, consistency, and integrity.

### **The Use of Common Evaluation Measures and Methods**

Issues that affect the process of evaluation include the methods used in data collection as well as the skill level of the designated evaluator who is responsible for gathering such data. In most situations, this responsibility falls on the individual who is delivering the program directly to the designated clientele. In Extension, that individual is usually a county-level Extension educator, sometimes called a county Extension agent. For all practical purposes, this individual is at the grassroots level, providing educational programs that enable clientele to improve their lives and/or their communities. The county Extension educator is often awarded the pleasure of witnessing the action unfolding to improve the lives of agricultural producers, children, youth, and families, and the communities in which they live. However, none of these outcomes would be apparent if the methods an evaluator chooses to gather feedback, measure impact, and share the results are inappropriate or substandard.

When considering a state's entire Extension organization and viewing the roles of everyone from the campus-level administrators and state specialists to county-level educators, those individuals who have direct contact with clientele with their programs at the county or parish level are usually the most relevant team members to gather credible evidence that supports Extension's

impact on communities. However, county Extension educators often lack the capacity or confidence to carry out a thorough evaluative process that will render the desired results (Lamm & Israel, 2013). The vast majority of county Extension educators are hired for their passion, creativity, and capability for developing and implementing programs. A by-product of the skill sets of these educators is their ability to build relationships with clientele and stakeholders due to their living in the same communities where they work. After establishing Extension's credibility based on the positive connections/networks formed, many county/parish-level Extension educators are preoccupied with coming up with and delivering the next new program idea. This effort is usually the highest priority for Extension educators and often results in the development and implementation of an evaluation plan for that program, or other previously delivered programs, becoming a lower priority or not even considered at all (Rennekamp & Arnold, 2009).

To increase the implementation of rigorous and influential program evaluations, involve more individuals than just the local Extension educators who are delivering the programs and who often do not have the training or time to adequately develop and conduct the types of evaluations that produce credible and actionable evidence of the programs' outcomes (Torock, 2009). Several states are doing this by engaging state Extension specialists to assist county Extension educators with their evaluation efforts. This includes taking an approach that employs common measures. Common measures are systematic evaluation tools aimed to assess the same or similar outcomes (see 4H.org, 2019; University of Minnesota, 2019; Weidner, 2017). The process of developing common measures usually begins with state-level Extension specialists working with county Extension educators to identify common issues across the state. Those common issues are then addressed by programs developed jointly by the county educators and the state specialists as well as state subject matter program leaders at times. As the county educators are charged with implementing the newly-developed programs at the local level, the state specialists are responsible for developing an evaluation instrument that is appropriate for measuring program impacts for each specific statewide program. These instruments can incorporate quantitative items commonly used in typical surveys or qualitative items (e.g., open-ended questions) to more thoroughly explore the experiences of clientele resulting from participation in the Extension program. These common measures evaluation instruments can be made easily accessible (e.g., via email, posted on a website) to any county Extension educator conducting a specific program. The evaluation instruments can then be downloaded by the county Extension educator whenever needed.

After program evaluation data have been collected using the common measures evaluation instruments, the county Extension educators can either send the evaluation results, summarized or raw, to the designated data collection person in the organization, or as some states have done, enter the data directly into an online portal created specifically to collect the evaluation data for each specific program. The data portal is often managed by Extension administrators in charge of accountability reports (e.g., state stakeholder accountability and reporting documents, the National Institute of Food & Agriculture (NIFA) Federal Plan of Work & Report of

Accomplishment reports). Once the county-level data have been entered, not only is there aggregated state-level data that can be used in multiple ways, but the information can also be sorted by county, giving the county educators a concise summary of the results based on the program that was implemented locally. In addition, one can compare results across county programs, if needed. This allows the county and state staff to both have an integral part in assessing the level of program impact. While state-level staff (e.g., subject matter specialists) have an opportunity to lend their expertise in designing and aiding in the implementation of collecting critical data, county agents/educators develop more self-assurance in knowing that they have adequately conducted a program that adds meaning to the lives of those they serve. In turn, evaluation is more readily adopted as a vital component of the program development process.

### **An Example from Extension Evaluation Practice**

All states have some means to gather Extension reports from county staff, state-level specialists, and/or faculty. However, in many cases, the data that are entered into these online portals are not reliable pieces of information, being “guesstimates” at best. In response to the need for consistency and validity, several states have taken significant steps to create reporting systems that provide access to credible data. Kansas State University Extension, for example, contracted with the University’s Office of Educational Innovation and Evaluation (OEIE) to create the Program Evaluation and Reporting System (PEARS) to help streamline data collection, evaluation, and reporting of evidence-based Extension and Supplemental Nutrition Assistance Program Education (SNAP-Ed) interventions. Evaluation data are entered in and pulled from PEARS in real time, fostering data-based decision-making related to program progress, implementation, and impact.

The PEARS Team at Kansas State University provides a product that strives for meaningful results. For example, *The SNAP-Ed Evaluation Framework* (U.S. Department of Agriculture Food and Nutrition Service [USDAFNS], 2016) contains 51 indicators and several hundred metrics with which the success of SNAP-Ed programming can be evaluated. PEARS modules have been developed in alignment with many of these indicators/metrics, allowing SNAP-Ed users to report programmatic impacts consistently and uniformly.

In addition, the USDAFNS requires each SNAP-Ed state to complete the Education and Administrative Reporting System (EARS) report annually (USDAFNS, 2017). While these reports have traditionally been compiled using numerous Excel spreadsheets within and across counties, users can now gain easy access to generate the report.

Many of the indicators/metrics from EARS have also proven to be excellent performance indicators for Extension programs, and the PEARS Team hopes to work with NIFA in the future to ensure the system continues gathering data that are aligned with Extension’s federal reporting requirements. In the meantime, PEARS provides a means for the Extension organization and

Extension nutrition educators to accurately evaluate and report on program impacts; make program improvements; and assess progress toward affecting policy, system, and environmental changes. These data currently help 28 states better understand what works, in what conditions, and why. This is ultimately helping them better allocate time and funding to maximize their impact, meet the needs of their communities, and build on their successes. As such, the PEARS Team has developed a system that places a particular emphasis on utility in improving Extension programs, policies, and accountability.

Taking such system-wide approaches benefits Extension in providing credible evidence. The use of these approaches allows the implementation of evaluation methods with a higher level of fidelity and credibility than would occur if a statewide program was administered with all counties determining their own way to gather data. In other words, methods processes, like common measures and the creation of more robust reporting systems, will help states identify a common core set of outcomes and indicators that will be useful in addressing critical issues facing Extension clientele and communities. Taking a systematic, consistent approach to developing and using common measures can also help state-level specialists work with county-level staff to more intently address needs that are unique to their communities and clientele. Not only does this approach provide a process for credible assessment, but it also offers the state Extension system a database for compiling reports that are action-oriented and ready for distribution at the request of stakeholders.

### Summary

Today's program evaluators must discern, from many angles, ways to be proactive in addressing individual and community needs and documenting the impacts and quality of the programs designed to address those needs. Even more, a paradigm shift must be implemented to ensure that one's lived experiences serve the same level of credibility as an equation that statistically predicts future occurrences. Although it is crucial to respect the fact that many stakeholders find credibility in numbers and percentages as a means to tell the true story, evaluators and evaluation stakeholders must embrace the fact that qualitative data should be given the same level of value and credibility as quantitative data.

Such a change in thinking when considering the credibility of evaluation data offers a reason to be versatile in analyzing and reporting data through multiple methods that speak to the expectations and needs of various audiences. Evaluation methods should never be underestimated in the quest to provide credible and actionable evidence. Method choice should certainly be contextual (Greene, 2007), but any framework viewed as an easy fix should be viewed with great caution. The merit of experimental or non-experimental approaches has and will continue to be used to assess the level of program effectiveness. To achieve this, the target audience (individuals, groups, communities) must be adequately considered.

This article provides a discussion on the purpose and need for the use of appropriate evaluation methods. Now is the time for Extension systems across the country to investigate what evaluation methods work for their needs and to determine which methods produce the most credible evidence for specific target audiences. Extension, throughout its history, has embraced new and innovative research-based concepts which have served both the land-grant universities and Extension's clientele and communities through outreach and engagement efforts. It is equally important to embrace the concept of identifying and applying appropriate evaluation methods that would benefit the organization's ability to collect and report credible and actionable evidence of Extension program impacts and quality that would increase Extension's public value among clientele and stakeholders.

### References

- 4H.org. (2019). *Common measures*. Retrieved from <https://4-h.org/professionals/common-measures/>
- Alkin, M. C., & King, J. A. (2017). Definitions of evaluation use and misuse, evaluation influence, and factors affecting use. *American Journal of Evaluation*, 38(3), 434–450. doi:10.1177/1098214017717015
- American Evaluation Association. (2018). *American Evaluation Association guiding principles for evaluators*. Retrieved from <https://www.eval.org/p/cm/ld/fid=51>
- Bennett, D. B. (1984). *Evaluating environmental education in schools: A practical guide for teachers*. Retrieved from <http://unesdoc.unesco.org/images/0006/000661/066120eo.pdf>
- BetterEvaluation. (2013). *Decide purpose*. Retrieved from [https://www.betterevaluation.org/en/rainbow\\_framework/frame/decide\\_purpose](https://www.betterevaluation.org/en/rainbow_framework/frame/decide_purpose)
- Bickman, L., & Reich, S. M. (2015). Randomized controlled trials: A gold standard or gold plated? In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *Credible and actionable evidence: The foundation for rigorous and influential evaluations* (2<sup>nd</sup> ed., pp. 83–113). Thousand Oaks, CA: Sage.
- Calba, C., Goutard, F. L., Hoinville, L., Hendriks, P., Lindberg, A., Saegerman, C., & Peyre, M. (2015). Surveillance systems evaluation: A systematic review of the existing approaches. *BMC Public Health*, 15, 448–460. doi:10.1186/s12889-015-1791-5
- Carlson, J. E., & von Davier, M. (2013). *Item response theory* [Educational Testing Service Research Report No. RR-13-28]. Retrieved from <https://www.ets.org/Media/Research/pdf/RR-13-28.pdf>
- Centers for Disease Control and Prevention. (2011). *Developing an effective evaluation plan*. Atlanta, GA: National Center for Chronic Disease Prevention and Health Promotion; Office on Smoking and Health; Division of Nutrition, Physical Activity, and Obesity.
- Charmaz, K., & Belgrave, L. (2012). Qualitative interviewing and grounded theory analysis. In J. F. Gubrium, J. A. Holstein, A. B. Marvasti, & K. D. McKinney (Eds.), *The SAGE handbook of interview research: The complexity of the craft* (pp. 347–366). Thousand Oaks, CA: Sage.

- Chawla, S. (2015). *Analyzing stakeholders*. Slideshare retrieved from <https://www.slideshare.net/sahilchawla98229/analyzing-stakeholders-sahil-chawla>
- Corporation for National & Community Service. (2019). *How to develop the right research questions for program evaluation*. PowerPoint presentation retrieved from [https://www.nationalservice.gov/sites/default/files/resource/Asking\\_the\\_Right\\_Research\\_Questions.pdf](https://www.nationalservice.gov/sites/default/files/resource/Asking_the_Right_Research_Questions.pdf)
- Coryn, C. L. S. (2007). The holy trinity of methodological rigor: A skeptical view. *Journal of MultiDisciplinary Evaluation*, 4(7), 26–31.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed). Thousand Oaks, CA: Sage.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2<sup>nd</sup> ed.). Los Angeles, CA: Sage.
- Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage.
- Donaldson, S. I., Christie, C. A., & Mark, M. M. (Eds.) (2015). *Credible and actionable evidence: The foundation for rigorous and influential evaluations* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage.
- Driscoll, D. L. (2011). Introduction to primary research: Observations, surveys, and interviews. In C. Lowe & P. Zemliansky (Eds.), *Writing spaces: Readings on writing* (Vol. 2, pp. 153–174). Anderson, SC: Parlor Press.
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363–406.
- Fink, A. (2015). *Evaluation fundamentals: Insights into program effectiveness, quality and Value* (3<sup>rd</sup> ed.). Thousand Oaks, CA: Sage.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4<sup>th</sup> ed.). New York, NY: Pearson.
- Foley, G., & Timonen, V. (2014). Using grounded theory method to capture and analyze health care experiences. *Health Services Research*, 50(4), 1195–1210. doi:10.1111/1475-6773.12275
- Fowler, F. J., Jr. (2002). *Survey research methods* (3<sup>rd</sup> ed.). Thousand Oaks, CA: Sage.
- Franz, N. K. (2013). Improving Extension programs: Putting public value stories and statements to work. *Journal of Extension*, 51(3), Article 3TOT1. Retrieved from <https://joe.org/joe/2013june/tt1.php>
- Franz, N., Arnold, M., & Baughman, S. (2014). The role of evaluation in determining the public value of Extension. *Journal of Extension*, 52(4), Article 4COM3. Retrieved from <http://www.joe.org/joe/2014august/comm3.php>

- Garbarino, S., & Holland, J. (2009). *Quantitative and qualitative methods in impact evaluation and measuring results*. Birmingham, UK: University of Birmingham.
- Gavazzi, S. M., & Gee, E. G. (2018). *Land-grant universities for the future: Higher education for the public good*. Baltimore, MD: Johns Hopkins University Press.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine Publishing.
- Gooden, S. T., & Berry-James, R. M. (2018). *Why research methods matter: Essential skills for decision making*. Irvine, CA: Melvin & Leigh.
- Greene, J. C. (2007). Methods choice are contextual, contingent and political. In G. Julnes & D. J. Rog (Eds.), *Informing federal policies on evaluation methodology: Building the evidence base for method choice in government sponsored evaluation. New Directions for Evaluation, 113*, 111–113.
- Hessler, R. M. (1992). *Social research methods*. St. Paul, MN: West Publishing Company.
- Holland, J., & Campbell, J. R. (Eds.). (2005). *Methods in development research: Combining qualitative and quantitative approaches*. London, UK: Practical Action Publishing.
- Kelley, K., Clark, B., Brown, V., & Sitzia, J. (2003). Good practice in the conduct and reporting of survey research. *International Journal for Quality in Healthcare, 15*(3), 261–266. doi:10.1093/intqhc/mzg031
- Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied Research* (5<sup>th</sup> ed.). Thousand Oaks, CA: Sage.
- Lamm, A. J. (2010). The latest take on utilization-focused evaluation. *Journal of Extension, 48*(3), Article 3TOT4. Retrieved from <https://joe.org/joe/2010june/tt4.php>
- Lamm, A. J., & Israel, G. D. (2013). A national examination of Extension professionals' use of evaluation: Does intended use improve effort? *Journal of Human Sciences and Extension, 1*(1), 49–62.
- Lamm, A. J., Israel, G. D., & Diehl, D. (2013). A national perspective on the current evaluation activities in Extension. *Journal of Extension, 51*(1), Article 1FEA1. Retrieved from <http://www.joe.org/joe/2013february/a1.php>
- Leavy, P. (2017). *Research design: Quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches*. New York, NY: Guilford Press.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Mark, M. M. (2015). Credible and actionable evidence: A framework, overview, and suggestions for future practice and research. In S. I. Donaldson, C. A. Christie, M. M. Mark (Eds.), *Credible and actionable evidence: The foundation for rigorous and influential evaluations* (2<sup>nd</sup> ed., pp. 275–302). Thousand Oaks, CA: Sage.
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding and improving policies and programs*. San Francisco, CA: Jossey-Bass.

- Marquart, F. (2017). Methodological rigor in quantitative research methods. In J. Matthes (Ed.), *The international encyclopedia of communication research methods* (Vol. 2, pp. 1–9). Hoboken, NJ: Wiley-Blackwell. doi:10.1002/9781118901731.iecrm0221
- Marshall, C., & Rossman, G. B. (1989). *Designing qualitative research*. Newbury Park, CA: Sage.
- McClure, M. M., Fuhrman, N. E., & Morgan, A. C. (2012). Program evaluation competencies of Extension professionals: Implications for continuing professional development. *Journal of Agricultural Education*, 53(4), 85–97. doi:10.5032/jae.2012.04085. Retrieved from <http://www.jae-online.org/attachments/article/1700/53.4.85%20McClure.pdf>
- McDavid, J. C., & Hawthorn, L. R. L. (2006). *Program evaluation & performance measurement: An introduction to practice*. Thousand Oaks, CA: Sage.
- Mertens, D. M., & Wilson, A. T. (2012). *Program evaluation theory and practice: A comprehensive guide*. New York, NY: Guildford Press.
- Morgan, D. L. (1997). *Focus groups as qualitative research* (2<sup>nd</sup> ed.). Newbury Park, CA: Sage.
- Mullins, J., Chapman-Novakofski, K., Franck, K., Olson, B., Serrano, E., Townsend, M. S., & Wong, S. S. (2015). Food and nutrition Extension programs: Next generation impact evaluation. *Journal of Extension*, 53(4), Article 4FEA1. Retrieved from <https://www.joe.org/joe/2015august/a1.php>
- National Minority Aids Council. (n.d.) *Program evaluation*. Retrieved from <http://www.nmac.org/pdf/program%20evaluation.pdf>
- National Research Council and Institute of Medicine. (2002). *Community programs to promote youth development*. Washington, DC: National Academies Press. doi:10.17226/10022
- Newman, I., & Benz, C. R. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum*. Carbondale, IL: Southern Illinois University Press.
- Nsubuga P., White M. E., & Thacker, S. B., Anderson, M. A., Blount, S. B., Broome, C. V., . . . Trostle, M. (2006). Public health surveillance: A tool for targeting and monitoring interventions. In D. T. Jamison, J. G. Breman, A. R. Measham, G. Alleyne, M. Claeson, D. B. Evans, P. Jha, A. Mills, & P. Musgrove (Eds.), *Disease control priorities in developing countries* (2<sup>nd</sup> ed., pp. 997–1015). Washington, DC: The World Bank and New York, NY: Oxford University Press.
- Office of Data, Analysis, Research and Evaluation Administration on Children Youth & Families [ODAREACYF]. (2016). *Qualitative research methods in program evaluation: Considerations for federal staff*. Retrieved from [https://www.acf.hhs.gov/sites/default/files/acyf/qualitative\\_research\\_methods\\_in\\_program\\_evaluation.pdf](https://www.acf.hhs.gov/sites/default/files/acyf/qualitative_research_methods_in_program_evaluation.pdf)
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Pell Institute. (2019). *Evaluation toolkit: Evaluation guide*. Retrieved from <http://toolkit.pellinstitute.org/evaluation-guide/>

- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly*, 68(1), 109–130. doi:10.1093/poq/nfh008
- Rallis, S. F. (2015). When and how qualitative methods provide credible and actionable evidence: Reasoning with rigor, probity, and transparency. In S. I. Donaldson, C. A. Christie, C.A., & M. M. Mark (Eds.). (2015). *Credible and actionable evidence: The foundation for rigorous and influential evaluations* (2<sup>nd</sup> ed., pp. 137–156). Thousand Oaks, CA: Sage.
- Rennekamp, R. A., & Arnold, M. E. (2009). What progress, program evaluation? Reflections on a quarter-century of Extension evaluation practice. *Journal of Extension*, 47(3), Article 3COM1. Retrieved from <http://www.joe.org/joe/2009june/comm1.php>
- Rolka, H., Walker, D. W., English, R., Katzoff, M. J., Scogin, G., & Neuhaus, E. (2012). Analytical challenges for emerging public health surveillance: A CDC perspective. *Morbidity and Mortality Weekly Report*, 61(03), 35–39.
- Santasier, A. M., & Plack, M. M. (2007). Assessing professional behaviors using qualitative data analysis. *Journal of Physical Therapy Education*, 21(3), 29–39. doi:10.1097/00001416-200710000-00005
- Secrest, L., & Sidani, S. (1995). Quantitative and qualitative methods: Is there an alternative? *Evaluation and Program Planning*, 18(1), 77–87. doi:10.1016/0149-7189(94)00051-X
- Siebold, W. (2011). *Examples of data collection methods*. Retrieved from [www.ncdsv.org/images/Siebold\\_ExamplesOfDataCollectionMethods\\_updated\\_11-7-2011.pdf](http://www.ncdsv.org/images/Siebold_ExamplesOfDataCollectionMethods_updated_11-7-2011.pdf)
- Teutsch, S. M., & Churchill, R. E. (Eds.). (2000). *Principles and practice of public health surveillance* (2<sup>nd</sup> ed.). New York, NY: Oxford University Press.
- Torock, J. L. (2009). Experiential learning and Cooperative Extension: Partners in non-formal education for a century and beyond. *Journal of Extension*, 47(6), Article 6TOT2. Retrieved from <https://www.joe.org/joe/2009december/tt2.php>
- U.S. Department of Agriculture Food and Nutrition Services. (2017). *Education and administrative reporting system form*. Retrieved from <https://snaped.fns.usda.gov/program-administration/ears-form-training>
- U.S. Department of Agriculture Food and Nutrition Services. (2016). *SNAP-Ed TOOLKIT - SNAP-Ed evaluation framework*. Retrieved from <https://snapedtoolkit.org/framework/index/sectors-of-influence/>
- U.S. General Accounting Office (U.S. GAO). (1987). *Case study evaluations*. Retrieved from [https://www.legistorm.com/reports/view/gao/15651/Case\\_Study\\_Evaluations.html](https://www.legistorm.com/reports/view/gao/15651/Case_Study_Evaluations.html)
- University of Minnesota. (2019). *CYFAR-approved common measures*. Retrieved from [https://cyfar.org/ilm\\_common\\_measures](https://cyfar.org/ilm_common_measures)
- Weidner, A. (2017). *Common measures 2.0*. Retrieved from <https://extension.purdue.edu/hub/wp-content/uploads/2018/07/Common-Measures-2.0-Presentation.2017.09.27-1-1.pdf>

*Dr. Ken Jones* is a Professor and Director of Program & Staff Development for University of Kentucky (UK) Extension.

*Dr. Eugenia Gwynn* is an evaluation specialist for the North Carolina A&T Extension Program.

*Dr. Allison Teeter* is an evaluator for the Office of Educational Innovation and Evaluation at Kansas State University.