

# Journal of Human Sciences and Extension

---

Volume 7  
Number 2 *Special Issue: Credible and  
Actionable Evidence in Extension Practice*

Article 13

---

6-28-2019

## Full Issue, Volume 7, Number2

Journal of Human Sciences and Extension

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



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

---

### Recommended Citation

Journal of Human Sciences and Extension. (2019). Full Issue, Volume 7, Number2. *Journal of Human Sciences and Extension*, 7(2), 13. <https://scholarsjunction.msstate.edu/jhse/vol7/iss2/13>

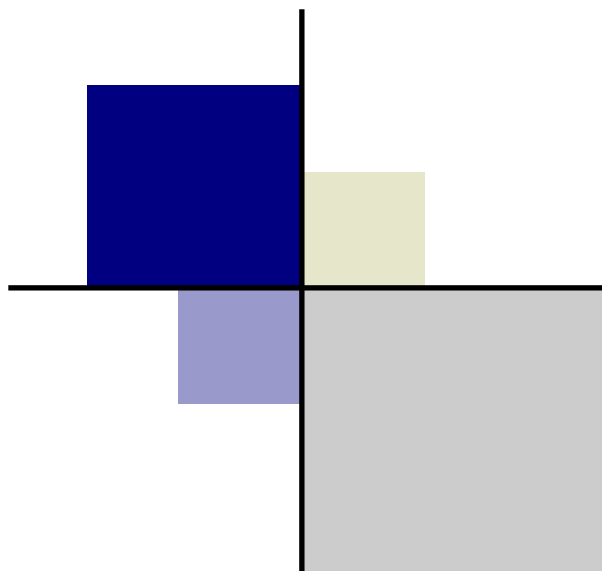
This Full Issue is brought to you for free and open access by Scholars Junction. It has been accepted for inclusion in Journal of Human Sciences and Extension by an authorized editor of Scholars Junction. For more information, please contact [scholcomm@msstate.libanswers.com](mailto:scholcomm@msstate.libanswers.com).

# Journal of Human Sciences and Extension

Volume 7, Number 2

June 2019

ISSN 2325-5226



Special Issue:  
Credible and Actionable Evidence in Extension Practice

Benjamin Silliman and Scott R. Cummings, Guest Editors

Donna J. Peterson and Rich Poling, Co-Editors



**Journal of Human Sciences and Extension**

**Volume 7, Number 2**

**June 2019**

**ISSN 2325-5226**

*Special Issue:*

*Credible and Actionable Evidence in Extension Practice*

**Table of Contents**

Editors' Introduction to the Special Issue

Rich Poling and Donna J. Peterson

Credible and Actionable Evidence in Extension Practice: Framing Issues, Contexts, and Principles

Benjamin Silliman and Scott R. Cummings

Whose Extension Counts? A Plurality of Extensions and Their Implications for Credible Evidence Debates

Thomas Archibald

Situational Complexity and the Perception of Credible Evidence

Scott Chazdon and Samantha Grant

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

Kenneth R. Jones, Eugenia P. Gwynn, and Allison M. Teeter

Measurement and Credible Evidence in Extension Evaluations

Marc Braverman

Credible and Actionable Evidence Across Extension Program Areas: A Case Example

Mary Marczak, Emily H. Becher, and Patricia Olson

Credible and Actionable Evidence Across Stakeholder Levels of the Cooperative Extension System

Nick T. Place, Richard M. Klemme, M. Ray McKinnie, Carrie Baker, Jean Parrella, and Scott R. Cummings

Communicating with Data: Telling the Extension Story in Credible and Actionable Ways

Diane D. Craig and Ruth H. Borger

How Evaluation Capacity Building Grows Credible and Actionable Evidence for Cooperative Extension Programs

Chelsea Hetherington, Cheryl Eschbach, and Courtney Cuthbertson

Closing Thoughts on Attaining More Credible Evidence for Extension Impact

Scott R. Cummings and Benjamin Silliman

Thank You to Reviewers for This Special Edition

## Editors' Introduction to This Special Issue

Over the last three decades, accountability has become increasingly important for agencies and organizations receiving funding from public or private sources. As the availability of public allocated funds has remained almost level or, in some cases, been reduced, accountability for those funds by agencies that receive the bulk of their funding from public funds has become increasingly scrutinized. Likewise, funds from private sources have become more and more dependent upon evidence that the allocated funds through grants and contracts are being used to produce the results desired by the funders. These situations have become even more precarious as more and more entities, both public and private, compete for the same funding resources.

Two important questions that need to be addressed in both of these situations are: What do funders and other stakeholders want in the way of accountability evidence that is both credible in the eyes of the funders and stakeholders as to the impacts and quality of the programs? and What evidence is actionable in ways to allow funders to act and make decisions about future funding by the funders, for the funding recipients to act and make decisions about modifying and improving programs, and to convince other stakeholders, including program participants, of the value of continuing the programs and the quality of the programs?

This issue of the *Journal of Human Sciences and Extension* focuses on both of these questions and provides readers with in-depth exposure to the meaning of credible and actionable evidence of program effectiveness and quality and how it can be addressed within an organization. The ten articles in this issue cover the basics of what credible and actionable evidence is; how such evidence can be identified, measured, and collected; how credible and actionable evidence can differ depending on different levels of an organization and the stakeholders wanting the evidence; how organizations can build capacity to collect credible and actionable evidence; and how this evidence can best be presented to program stakeholders.

The context in which these articles are presented is the Cooperative Extension System, at the national, state, and local levels of that system. However, the information in these articles can be just as valuable for any agency or organization that has to deal with accountability.

Dr. Ben Silliman and Dr. Scott Cummings served as Co-Editors for this special issue. We wish to thank them for their efforts in bringing together a distinguished group of authors for this special edition and their own contributions as authors. We also want to especially thank the individuals who volunteered their time and efforts to serve as peer reviewers of the articles found in this special edition. The list of these individuals can be found at the end of special edition.

Richard L. Poling and Donna J. Peterson, Co-Editors  
*Journal of Human Sciences and Extension*

## **Credible and Actionable Evidence in Extension Practice: Framing Issues, Contexts, and Principles**

**Benjamin Silliman**

*North Carolina State University*

**Scott R. Cummings**

*Texas A&M AgriLife Extension*

*Organizations that translate emerging science and provide community outreach, such as the Cooperative Extension Service and similar outreach engagement programs at universities, face ongoing challenges in establishing the credibility of program content and results as the pace of discovery of new knowledge, demand for effective applications, and diversity of audiences and other stakeholders continues to expand. This special edition of the Journal of Human Sciences and Extension (JHSE) explores the theme, “What is credible and actionable evidence in Extension programs?” Like a good evaluation, we begin this introductory article by framing the question, including academic, policy, and practical contexts; definitions of terms; discussion of the Extension context of credible evidence; and a sample of frameworks used to ground claims to credibility across disciplines and levels of reporting. A brief review of each article in the special edition concludes this overview of the JHSE special edition.*

**Keywords:** credible evidence, actionable evidence, Cooperative Extension, program evaluation, evidence-based practice, program evaluation standards, logic models

***“Unfortunately, seeking truth or agreement about what constitutes credible and actionable evidence does not seem to be an easy matter in most fields.”***

**—Stewart I. Donaldson (2015)**

### **Introduction**

Over many years of working with the Cooperative Extension Service (*Extension*), we have heard statements and questions similar to the following about Extension program effectiveness:

- “I know I am making a difference,” a confident young county Extension agent declares. “Our nutrition education program served 4,500 people last year.” “OK,” the county director replies, “So, how many of those participants and their families are eating healthy meals or saving money on food or medical bills?”

---

Direct correspondence to Benjamin Silliman at [bsillim@ncsu.edu](mailto:bsillim@ncsu.edu)

- “I know I am making a difference,” explains an experienced field crop Extension agent, “Producers are implementing conservation practices, trying drought-resistant varieties, and recognizing early-on when they have disease problems.” “Great,” replies a state Extension specialist, “But did producers “check off” those items on a list, or describe what they actually do? Have you been in the field with them to observe these changes?”
- “I know I am making a difference, an Extension program leader notes. “Three counties with long-standing financial management programs saw an increase of ten percent in families becoming self-sufficient. In three counties where there was never an interest in those programs, at least five participating families became self-sufficient and recommended the program to their friends.”
- “I know I am making a difference,” an Extension volunteer youth leader insists, “Our programs teach life skills, so they will be productive citizens in the future.” An interested county commissioner replies, “What exactly are those skills, and how do you know it is your program that turns youth into productive citizens?”

These statements and the follow-up questions raise an underlying issue. What counts as credible evidence for design and impact in community-based programs such as Extension and other engagement outreach programs of public and private universities? Our focus in addressing this question in this special edition of the *Journal of Human Sciences and Extension* (JHSE) will be on the programs of the Cooperative Extension System (CES), a partnership between the U.S. Department of Agriculture’s National Institute for Food and Agriculture (NIFA) and the nation’s land-grant universities, but the insights provided by the authors in this special edition will likely be applicable for a variety of organizations. As the scenarios above suggest, expectations for the amounts and types of evidence for program impact (i.e., making a difference) varies widely among Extension stakeholders.

Criteria for credible evidence include evidence-based practice, rigorous evaluation designs and measures, and usability of data for participants, practitioners, and decision-makers (i.e., stakeholders). Yet rigorous programming and evaluation entail different credibility criteria across diverse disciplines, settings, and stakeholder needs. Resources and conditions in non-formal, community-based educational programming settings rarely match those in research laboratories or model programs where evidence standards are established. Moreover, the process for translating science, delivering programs, and generating evidence for program effectiveness are not well-understood by many decisionmakers, participants or other citizen-stakeholders, or for that matter, many practitioners themselves. Thus, it is often challenging to know exactly how to support claims for “making a difference” or how to apply evidence of success to program improvement or policy decisions.

This special edition focuses on what types of evidence demonstrate the quality and impact of Extension programs, how evidence is generated, and how that evidence can be used by



stakeholders to make decisions on a wide range of concerns. Extension programs reflect diverse disciplines, strategies, objectives, contexts, and resources and serve stakeholders with diverse expectations. Stakeholders have differing capacities to understand and weigh program effectiveness evidence. We cannot answer all questions related to what evidence is *credible* (i.e., relevant and trustworthy) and *actionable* (i.e., useful for decisions about policy, practice, personal organizational change). We hope to show why credible evidence is often not implicit in Extension work, often not easy to produce, and in some cases, not universally acclaimed by stakeholders.

We also recognize that, for many stakeholders, credibility also connotes not simply program integrity or validity of impact data, but customer satisfaction with a program's processes, outcomes, and relationships. In a world of information overload, conflicting claims, and significant consequences for policies, programs, and personal decisions, skilled interpreters of credible and actionable evidence can add value to policy, program, and personal decisions. Cooperative Extension organizations that can translate knowledge, generate evidence of program impact, and facilitate understanding and use of credible evidence will sustain the mission of their land-grant universities.

In this introductory article, we describe the concepts and contexts of the broader credible evidence discussion and their relevance to Extension work. We will also highlight evaluation frameworks and resources that may be useful in thinking about what is credible and actionable evidence and useful in your own efforts to collect such evidence.

### **Framing the Discussion: What Counts as Credible and Actionable Evidence?**

#### **Defining Credible Evidence**

Credible evidence, in the broadest sense, is information that stakeholders perceive as trustworthy and relevant for answering their questions about a program (Centers for Disease Control and Prevention [CDC], 2012, *Program Evaluation Framework*, as cited in Donaldson et al., 2015, p. 7). The CDC Framework also notes that stakeholders may judge credibility based on how questions are posed or results interpreted, sources of information accessed, data collection methods and measures, and quality control procedures employed. Across different types of programs<sup>1</sup>, settings, and stakeholder priorities, the quantity and quality of evidence required for policy, practice, or funding decisions varies widely (Franz, 2013; Franz & Townson, 2008), and the definition of high-quality evidence is vigorously debated (Donaldson et al., 2015, p. 9). In practice, credibility of evidence is affected by stakeholders' engagement in identifying

---

<sup>1</sup> Program types include information campaigns, educational or training programs, implementation of policy recommendations, or other activities directed toward enrichment, prevention, or remediation (cf. Ripley et al., 2011; see also Franz & Archibald, 2018 on programming continuum).

evaluation questions and evidence criteria as well as practitioners' expertise, time, and resources for gathering evidence (Donaldson et al., 2015, p. 6; Franz & Archibald, 2018).

Mark (2015, p. 277) describes four attributes of evidence that influence decisions and actions: 1) credibility, or trustworthiness of processes and quality of the product; 2) relevance, or importance of the process or product; 3) inferential potency, or level of confidence in evidence based on critical criteria; and 4) comprehensiveness, or evidence across a broad range of questions. Health and safety criteria for agricultural processes and products reflect these attributes. Stakeholders, ranging from consumers to state health inspectors to dairy farmers, rely on a complex of procedural observations, lab tests, and anecdotal reports to establish a research evidence base and monitor practice from farm-to-fork. Different criteria may be relevant for different products. Food safety may be compromised or enhanced by a variety of expected and unexpected factors in the supply chain; thus, inferences about safety usually require more than one type of evidence. Moreover, evidence for food safety is interconnected with evidence about the environment, plant or animal as well as human health, economics, politics, ethics, and much more. Consequently, the credibility of a food system requires triangulated evidence from multiple sources, filtered through diverse perspectives and criteria. A similar process might be applied to Extension programming about food systems or other mission areas.

Relevance and trustworthiness of program outcome data are enhanced by collaboration between program evaluators, or the program deliverers who will conduct the evaluations, and program stakeholders who, together, work toward a common understanding of the problem or opportunity, prioritize evaluation questions, and agree on appropriate evidence, as time and resources permit (Donaldson, 2015, pp. 3–4). In this process, an Extension educator, supported by subject matter and evaluation specialists, can assess needs and assets and introduce diverse stakeholders to current evidence from basic and applied science. Together, guided by program theory (Donaldson, 2015, p. 5), collaborators can identify relevant evaluation questions and standards for credible evidence of program effectiveness. The planning process, which does not have to be long and contentious, also provides opportunities to build personal rapport (with the evaluator and among stakeholders) and organizational credibility, enhancing subsequent use of findings (Owen & Rogers, 1999, p. 117–120). Over time, an evaluator and/or planning facilitator can help stakeholders reflect on deeper value claims that influence their views of credible evidence, including personal preferences, perceptions of public value, legal or professional standards, contextual significance, and exemplary value (Scriven, 2007).

In this context, credible evidence, as interpreted by Donaldson and colleagues (2015), is not about what is true in an ultimate or absolute sense. Rather, credibility refers to what is relevant and trustworthy within a particular scientific paradigm or evaluation theory. Horgan (2012) notes that Kuhn (1962), argued that both falsification and verification each:

imply absolute standards of evidence, which transcend an individual paradigm. A new paradigm may solve puzzles better than the old one does, and it may yield more practical applications. But that does not make it absolutely more true than previous paradigms.

The new paradigm to which Kuhn refers may include new theories or evidence from science, alternative theories or approaches to evaluation, legal or policy criteria, cultural assumptions, and theories about reality. In this regard, professionals involved in evaluation must be consistently self-reflective and transparent about the personal and professional assumptions that guide their work and ready to question and/or better understand the credibility of established policy or practice (Miller, 2015, pp. 53–58).

In agriculture, for instance, paradigms and place are critical to credibility and actionability in agroecology (Reynolds, Smith, & Farmer, 2014; Valenzuela, 2016; Welsh & Rivers, 2011). In program delivery, research suggests that experiential learning facilitates motivation, learning, and creativity in many settings, but is not as efficient or effective as direct instruction for immediate memory and analytical skills (Baker & Robinson, 2018).

The traditional positivist approach to science produced many advances in discovery and application but often became too rigid in method or application of findings. Modern post-positivist evaluation science appreciates diverse voices, contexts, and processes, but may produce nonreplicable results (Christie & Fleischer, 2015, pp. 29–35). Consistent with discussions of the conduct of research (Pennock & O'Rourke, 2017) and *Guiding Principles for Evaluators* (American Evaluation Association, 2018), Extension professionals serve best as humble stewards who are honest about their assumptions, objectives, and limitations and patient interpreters, whether stakeholders are supportive or skeptical about the credibility of their evidence.

### **Defining Actionable Evidence**

Actionable evidence describes evaluation results that are adequate and useful for making policy or programming decisions (Julnes & Rog, 2015, p. 221). Credible evidence is not always actionable. For instance, we may recognize patterns of productivity or resiliency but not fully understand how to promote those outcomes programmatically or how to adjust for differences in organism or context. However, in many situations, educators may know enough to take preventive or proactive steps, even though they do not have complete knowledge of the change mechanisms, contexts, or other factors. As with credibility, what evidence is actionable depends on evaluation questions and contexts, and some methods may be more helpful than others in supplying that evidence (Julnes & Rog, 2015, p. 221). Julnes and Rog (2015, pp. 226–227) present Weiss' (1998) taxonomy of evaluation questions and Mark, Henry, and Julnes' (2000) related comment on evaluation tasks as a useful guide for program leaders or evaluators to decide what level of evidence is needed for action. For instance, at the Implementation level, similar to Rockwell and Bennett's (2004) output level, program reach or accessibility and fidelity

to evidence-based models are key questions for evaluation use. Blyth (2011) underlines this point regarding youth development: If programs are not accessible to all youth, how (or how much) does that compromise program claims for promoting positive youth development? Actionability questions at the Outcome level focus on the improvement of program participants relative to prior levels of performance (e.g., knowledge, attitudes, skills, and aspirations) and/or in comparison to a control group. At the level of Impact, or long-term sustained change, actionability focuses on aggregated impact (e.g., changes due to the program), disaggregated impact (e.g., relative benefits for participant groups), and causal mechanisms (e.g., program components critical to sustained changes).

Weiss (1998) also includes actionability questions related to cost-benefit analysis and critical program review (e.g., unanticipated effects, limitations and practical implications, recommendations for programs and policy changes, implications for new policies).

Actionability, like credibility, must be defined in terms of context and stakeholder questions, but the levels of analysis suggested by Weiss (1998) continue to provide a general rubric for thinking about practical uses of program evaluation data.

Julnes and Rog (2015) suggest that evaluations focused on program activities or outcomes require relatively less rigorous evidence than programs being piloted as models or programs being evaluated for cost-effectiveness or policy decisions. For instance, a drop-in program offering fitness activities and nutrition information in a community senior center might be monitored for participation (e.g., evidence of community interest, accessibility), program protocols (e.g., fitness screening) and quality (e.g., appropriate activities, supportive interaction). A more rigorous program would include these same participation, protocol, and quality checks, but also track indicators such as participants' weights, muscle tones, and blood pressures over time as well as conduct interviews of participants for details about the quality of their experiences and impacts of their participation. Targeted questions, representative sampling, and advanced analyses may help program leaders and evaluators weigh benefits based on participants' traits or the program's strategy to inform decisions about program improvements or expansion. Data on local health trends, comparisons across program sites, and with research on similar programs could further assist stakeholders in knowing how to invest resources in senior wellness programs. However, if decision-makers lack information on questions, such as where new services are needed or what additional organizations will contribute resources, they may not be able to act on program expansion.

A wide range of evaluation theories and methods have been developed to provide credible and actionable evidence to address particular evaluation questions and contexts. No one method can adequately address all questions and contexts, and mixed methods (e.g., quantitative and qualitative approaches) may be needed to provide compelling evidence on a single question.

Patton's Mountain of Accountability model (Patton & Blandin Foundation, 2014) describes an even broader range of indicators for organizational integrity and growth as well as program quality and outcomes. In this model, the first level, Basic Accountability for Management Processes, focuses on fiscal and program management, emphasizing intentional planning, effective management, fiscal and operational transparency. This level also involves due diligence in delivering activities and managing resources, consistent with sponsor expectations. Evidence for due diligence may be inferred as readiness to manage more complex or extended projects and achieve targeted impacts.

The next level, Accountability for Impact, entails the gathering of program quality and outcome data through internal and external evaluations utilizing diverse sources (e.g., staff, participants, boards, and broader stakeholders). The next higher level of Patton's mountain model, Accountability for Learning, Development, and Adaptation, focuses on reflective practice and process improvements that fuel learning and system change. Once programs consistently show desired results, a focus on continuous quality improvement is critical to sustaining or extending benefits, building capacity, and innovation. Significantly, the model emphasizes management and review functions of program evaluation and organizational learning that are ignored or presumed by other planning and evaluation models. Both the "bottom end" and "top end" evaluation questions are critical to credibility of programs and sponsoring organizations.

### **The Broader Practice and Policy Debate**

Organizations and policymakers in all fields face major challenges in determining credible evidence for a wide range of decisions, including issues with implications for life and death (e.g., medical treatments, technology innovations) and public or private investments (e.g., social, economic, environmental, social policies and program support). Stewart Donaldson and his colleagues addressed this ongoing debate in the book, *Credible and Actionable Evidence: The Foundation for Rigorous and Influential Evaluations* (Donaldson, Christie, & Mark, 2015). That book is a touchpoint for this special edition.

While conceding the idealism of Donald Campbell's *Methods of the experimenting society* (1991), where all policy decisions would be informed by rigorously-tested evidence, Donaldson et al. (2015), agreeing with Shadish, Cook, and Leviton (1991), recognized that "information or evidence judged to be poor by experimental scientific standards was often considered acceptable by key decision makers, including managers, politicians, and policy makers" (Donaldson, 2015, p. 8). Further discussion of the debate between these two paradigms is available at the American Evaluation Association's (2003) website (<https://www.eval.org/p/cm/ld/fid=95>). Articles in special issues of the publication *New Directions for Evaluation* also offer further discussion on credibility and validity (Chen, Donaldson, & Mark, 2011) and on mixed methods (Mertens & Hesse-Biber, 2013).

## Credibility in Extension Programming

### Diverse Criteria and Contexts for Credibility

The Cooperative Extension System (CES) provides one institutional context in which to reflect on the challenges of generating credible and actionable evidence. The goal of the CES, concisely stated on the NIFA website (National Institute for Food and Agriculture [NIFA], 2014), is “translating research into action: bringing cutting-edge discoveries from research laboratories to those who can put knowledge into practice.”

Given the breadth of Extension programming (Bull, Cote, Warner, & McKinnie, 2004; Kellogg Commission on the Future of State and Land-Grant Universities, 1999), Extension’s “research laboratories” are not limited to clean rooms populated by white coats and microscopes, although that may be the popular stereotype for scientific credibility. Field trials, non-formal educational programs, or 4-H camps illustrate settings with less controlled conditions than clinical laboratories and are typically more challenging contexts in which to establish credibility and actionability.

The NIFA website also states that the CES “empowers farmers, ranchers, and communities of all sizes to meet the challenges they face, adapt to changing technology, improve nutrition and food safety, prepare for and respond to emergencies, and protect our environment” (NIFA, 2018). In fact, the CES mission is much broader than described, including diverse disciplines, serving diverse stakeholders in diverse settings. Stakeholders in each discipline and decision-makers at each level of the CES (e.g., county governments, state land-grant universities, state legislatures, and the U.S. Department of Agriculture) hold differing standards of evidence and have preferences for different types of data, based on the different evaluation questions addressed by each field or type of program. Moreover, in each field, the complexity of the setting, strength of the research base, level of program maturity, and capacities of program staff and evaluators influence the quality of evidence that can be gathered about program merit and worth. In addition, new discoveries or innovative technologies often profoundly shift standards (e.g., hybrids/crop yields) or criteria (e.g., sustainability, environmental stewardship, animal welfare, farm labor health) for credible evidence in science and society.

Extension, like many organizations, is constantly challenged to collect and present evidence for program impacts in ways that are both scientifically credible and easy for stakeholders to understand. Making that challenge even harder in recent years has been social media and advocacy outlets swirling with either attacks on or support for the credibility of others that may or may not include the use of scientific-based data and often contains biased inferences. Credibility questions may extend beyond just program impacts to program and evaluation strategies, to organizational reputation, to questions of the common good and societal priorities. Extension professionals are challenged today, more than ever, to examine and interpret the substance and delivery of educational programming, the methods of evaluating program

strategies and impacts as well as their interactions with stakeholders who hold diverse views of what is credible evidence and how it is determined. The process of defining, generating, and interpreting credible evidence is continually a work in progress. At any given point, organizations and individuals can only determine what is “good enough” evidence in present circumstances.

### **Criteria for “Good Enough” Evidence in Public-Serving Organizations**

Given the diverse range of expectations and resources to support evidence-gathering, what is “good enough” evidence for a program’s impact and value to stakeholders? As with all credibility and actionability questions, the shortest answer is, “It depends.” We suggest three broad principles, including integrity, transparency, and adaptability, as a way to frame discussion and decisions on what qualifies as “good enough” evidence for Extension programs.

**Integrity to mission and standards.** As a public institution, Extension provides programs in the public interest and for the common good. Extension organizations have professionals who understand, apply, and help create standards and methods for generating and judging credible evidence both within their specific disciplines and within broader disciplines, such as communication, non-formal education, and leadership. Relying on program theory, evidence-based practices, and the use of high-quality measures and methods enables these professionals to produce outcomes at the higher levels of actionability (e.g., impact, cost-benefit, program review). While hierarchies of evidence are emerging in education (Institute of Education Sciences, 2019) and healthcare (Evans, 2002), rigorous and comprehensive evidence for practice is somewhat less advanced in areas such as agriculture (Virgona & Daniel, 2011) and social programs (Fraser, Richman, Galinsky, & Day, 2009).

Integrity to public value and the common good is equally critical to credibility (Franz, 2013; Greene, 2015; Kellogg Commission on the Future of State and Land-grant Universities, 1999). Across diverse disciplines, all practitioners are expected to conform to legal, professional, and ethical standards. However, outside the public sector, evidence is typically generated, interpreted, and accessible to only particular stakeholders. Proprietary information in the private sector is not generally accessible to those not designated to receive and use it. Sometimes, evidence may even be suppressed or overemphasized, or implications slanted to promote a particular product or organization. Historical examples of this include evidence on tobacco use and health (Brandt, 2012) and public vs. proprietary control of agricultural products (DeSchutter, 2011; Eisenberg & Nelson, 2002). Such practices can diminish the relevance and trustworthiness of that evidence in the general population. By contrast, access to high quality, understandable and unbiased evidence at the front end of programs (e.g., evidence for program content and delivery strategies) and the back end of programs (e.g., program results and actionable recommendations) is most likely to be judged relevant and trustworthy by a wide range of stakeholders.

**Transparency on practice and results.** High-quality evidence earns credibility not only because it is relevant and trustworthy to stakeholders but because it is presented and interpreted clearly and respectfully, at their level of understanding (Greene, 2015, pp. 208–109). Engaging stakeholders from initial needs assessment processes to actionability decisions typically enhances stakeholders' understanding of what, how, and why program strategies work and what evidence is needed to show program quality and impact.

Ideally, Extension professionals can develop and implement evidence-based programs and well-tested evaluation methods. Quite often, they encounter challenging situations and offer programming in settings where research evidence and the application of scientific standards are not well-established and often not practical. In these cases, the Extension professional must utilize and generate the best available evidence. Yet even programs based on well-tested models may yield weaker evidence than the original models, as we know with automotive fuel efficiency estimates. Realistically, many Extension educational programs may help participants gain the skills to make decisions or change behaviors but cannot eliminate the risks associated with those decisions being successful or guarantee that those changes actually occur. Transparency about program potential, limitations, and implications would help stakeholders judge a program's value and take appropriate actions based on evaluation evidence.

**Adaptability to conditions and criteria.** Program resources and conditions often limit the quantity and quality of evidence that can be gathered. Bamberger, Rugh, and Mabry (2012) identified effective strategies to generate evidence under budget, time, data, and political constraints. When inconsistent participation patterns and lack of program evaluation capacity limit the collection of outcome data, stakeholders may need to either focus more on program quality (Arnold & Cater, 2016) or scale up evaluation resources and capacity-building to generate a higher level of evidence (cf. Weiss, 1998, evaluation questions taxonomy). When evaluation resources and capacity are in short supply, one thing that program developers and implementors need to keep in mind is that not all programs require extensive evidence of merit or worth (Scriven, 2007). Typically, Extension professionals cannot conduct in-depth evaluations of all programs simultaneously, so decisions need to be made as to which evidence is most important and a priority for stakeholders and which evidence is not.

Even when evidence is relevant or trustworthy to one set of stakeholders at a given time and place, other evidence may be needed by stakeholders asking different questions. For instance, Federal officials may be interested in impact evidence of programs across states, whereas state or local officials may be satisfied with evidence just from their own jurisdiction. Thus, the process of generating “good enough” evidence is always a work in progress. Fortunately, a variety of frameworks and tool have been developed in the last two generations. A few of these are discussed below.



## **Evaluation Frameworks that Foster Credible Evidence**

### **Logic Models to Plan, Manage, and Interpret Credible Evidence**

Donaldson (2015, pp. 5–8) recommends that, regardless of the problem or opportunity being addressed, credible and actionable evidence is most likely to emerge from an intentional and systematic process that identifies, generates, and utilizes credible evidence. Logic models such as the Wisconsin Extension Program Evaluation Model (Taylor-Powell, Steele, & Doughlah, 1996) and the Centers for Disease Control and Prevention (CDC) Program Evaluation Framework (Centers for Disease Control and Prevention [CDC], 1999; Milstein, Wetterhall, & CDC Program Evaluation Working Group, 2000) provide such tools.

Each of these program logic models includes the input of program stakeholders in the program planning processes, including the needs assessment phase, the identification of desired program outcomes, and what evaluation measures will represent program successes. As diverse program stakeholders share in the program planning process, all those involved in the process will gain perspective on what evidence seems relevant and trustworthy to others. This exchange provides Extension professionals opportunities to review the research base and facilitate open discussion and reflection on program criteria, concerns, and consensus. Such dialogues also provide opportunities to explore the limits of credibility (e.g., probability vs. absolute certainty, assumptions underlying programming and evaluation decisions, variations in implementation strategies, possible alternative interpretations of data) and the significance of actionability (e.g., gathering data specific to making critical decisions and taking actions). Finally, stakeholder interaction during planning, implementation, and interpretation processes can result in credible indicators of the sustainability and effectiveness of the organization or partnership leading a program.

### **Developmental Evaluation**

Developmental evaluation, involving the continuous revision of evidence expectations and data-gathering strategies to fit changing conditions and goals, can provide a more flexible approach than a logic model, especially for new and complex initiatives (Franz, Garst, & Gagnon, 2015; Honadle, Zapata, Auffrey, vom Hofe, & Looye, 2014). Exploratory or start-up programs that want to establish parameters for credible evidence for processes (e.g., program delivery, evaluation, management, collaboration), impact (e.g., targeted outcomes and levels of change), or context (e.g., conditions and settings influencing change) may benefit from a developmental evaluation approach.

### **Evidence-based Practices**

The credibility of evidence for impact at the end of a program depends on the credibility of program design and implementation.

Ideally, program theory (Braverman & Engle, 2009; Sharpe, 2011) and implementation (Bauer, Damschroder, Hagedorn, Smith, & Kilbourne, 2015; Duerden & Witt, 2012) fit program content and delivery to audience needs in order to achieve desired outcomes. For programs such as pesticide management or youth shooting sports, protocols for program delivery must be followed closely to maintain safety and achieve positive outcomes. Other programming may allow more latitude for timing, instructional approaches, social and environmental conditions. However, not all Extension programs have a strong research and practice base. In such cases, the use of principles-focused evaluation (Patton, 2017) may aid decision-makers in tracking a program's processes (e.g., program delivery, program management, use of results) critical to achieving outcomes valued by an organization.

### **Program Evaluation Standards**

The Program Evaluation Standards (PES), developed for the Joint Committee on Standards for Educational Evaluation (JCSEE) (Yarbrough, Shulha, Hopson, & Caruthers, 2011), describe principles for effective and ethical practice of program evaluation that are integral to building credible and actionable evidence. In brief, the JCSEE PES include:

- 1) *Utility*, or usefulness of the process and results for stakeholders, facilitated by qualified evaluators who engage and communicate with all stakeholders in negotiating relevant purposes and promoting responsible and adaptive use of results.
- 2) *Feasibility*, or efforts to increase evaluation effectiveness and efficiency through good management, practical and responsive procedures, balancing political realities and stakeholder needs, and wise use of resources.
- 3) *Propriety*, or strategies to maintain what is “proper, fair, legal, right and just” in evaluations involving responsive and inclusiveness, protection of human rights, including formal agreements with stakeholders, evaluating and reporting in ways that are clear and fair, transparent and complete, disclosing conflicts of interest, and exercising fiscal responsibility.
- 4) *Accuracy*, or findings and interpretations that promote dependability and truthfulness, such as justifying conclusions in relation to context, valid and reliable information, explicit descriptions of program and context, with sound designs, interpretative judgments, and reporting accuracy.

The JCSEE PES also includes accountability standards to periodically explore and reflect on the purposes and processes of evaluations (e.g., meta-evaluation). Attention to the PES may seem like a time and resource investment that exceeds already-limited time and expertise for evaluation activities. Evidence from several fields indicates that the PES can be a valuable evaluation planning and capacity-building tool (Gill, Kuwuhara, & Wilce, 2016; Ruhe & Boudreau, 2013). The American Evaluation Association (2018) Guiding Principles for Evaluators also provide further professional guidance for the evaluation process.

### **Closing Thoughts: Building More Credible and Actionable Evidence**

Credible and actionable evidence is neither implicit in Extension work, easy to produce, or universally acclaimed. Some stakeholders, ranging from average citizens to policymakers, would count “anything reminiscent of Mom, the flag, and warm apple pie” as evidence-based practice or simply “good evaluation practice” (Shadish, Cook, & Leviton, 1991, as cited in Donaldson, 2015, p. 5). To paraphrase, “whatever evaluation produces results that satisfy clients is a good evaluation.”

Daniel Stufflebeam (2001) identified two evidence-gathering strategies that, while they may seem trustworthy and relevant, at least to some stakeholders, are likely to promote invalid or incomplete findings:

- 1) public relations-inspired studies designed to tout program value without solid scientific evidence, and
- 2) politically controlled studies, making claims that support an agenda or outcome favored by particular stakeholders (e.g., grantor, organization, interest group, or program leader) while withholding evidence that might conflict with their interests.

The former may take the shape of testimonials or marketing campaigns in lieu of rigorous evidence. The latter may include not only biased questions and methods but interpretations that overemphasize the positive and avoid the negative in order to impress funders or maintain a positive public image.

Because Extension’s mission is more than just keeping the customer satisfied, turning a profit, or doing science for its own sake, and because resources and data are almost always limited, the generation of trustworthy and usable evidence requires the use of programming and evaluation standards; professional judgments; and systemwide, long-term commitment to evaluation (Franz, Arnold, & Baughman, 2014; Lamm & Israel, 2013). Moreover, the way in which Extension engages, educates, and empowers stakeholders in the program development, implementation, and evaluation processes will likely influence not only the stakeholders’ perspectives of issue-related evidence but also their views of the credibility of the organization itself.

### **Reflecting on “Making a Difference” Statements**

We close by reflecting on some simple strategies and principles for building credible and actionable evidence of program effectiveness related to situations in which Extension professionals often find themselves, similar to the scenarios at the beginning of this article:

- Traditionally, program reporting focused on participant numbers, assuming that a broad range of citizens was served, and a significant portion would change behavior. More concrete evidence of behavior change is not only more credible but often provides actionable clues to next steps for programs and participants.

- Although much evidence can be gathered with a brief survey or checklist, more in-depth data on what, how, and why of program effects can be obtained with qualitative methods.
- Differences in context may necessitate different thresholds for credibility and strategies for data collection and interpretation. How can we compare programs at different stages and settings? It depends...
- Activities such as teaching do not necessarily produce outcomes, which is why programs are evaluated. More precise measures of specific skills with sustained outcomes provide more credible evidence than claims of broad skill change in a short period of time. In addition, simply citing resources such as the Targeting Life Skills Model (Hendricks, 1998) or research such as the national 4-H Study (Lerner, Lerner, & Colleagues, 2011) does not offer universal validation of all Extension youth programs.<sup>2</sup>

### Special Edition Topics

This special edition of the *Journal of Human Sciences and Extension* introduces Extension professionals at all levels as well as other professionals who are conducting similar types of educational programs, to key concepts related to using and generating credible evidence with diverse stakeholders in diverse situations. Authors in this special edition will address key issues and practices that should spark learning and debate on how we can plan, implement, and evaluate programs; tell our stories; and use program insights more effectively. We believe these steps are crucial in the pursuit of the land-grant mission and in the sustainability of evidence-based, public-serving programs across the length and breadth of the Cooperative Extension System.

Credible evidence begins with understanding the mission and meaning of Extension programming for diverse stakeholders and programs. In his article, “Whose Extension Counts? A Plurality of Extensions and Their Implications for Credible Evidence Debates,” Tom Archibald discusses how different understandings of Extension’s mission and program evidence has both enhanced and hampered Extension’s effectiveness. He goes on to show how engagement and empowerment of all stakeholders provides the best guide to setting objectives and achieving outcomes.

---

<sup>2</sup> Targeting Life Skills identifies potential life skills strategies and outcomes but does not provide a curriculum or evidence for specific amounts and types of training needed to produce specific changes in life skills. The National 4-H Study of Positive Youth Development surveyed child and adolescent 4-H participants whose civic engagement and career aspirations were higher than non 4-H peers. Other research suggests positive implications for adult development. A local program would need its own evidence for program quality (e.g., since no single model was noted in the National 4-H Study) and outcomes (e.g., to support its own claims to short- and long-term impact).

Credible evidence should be explained to stakeholders clearly and simply but is typically generated within complex contexts. In their article, “Situational Complexity and the Perception of Credible Evidence,” Scott Chazdon and Samantha Grant discuss how principles of developmental evaluation can help educators and evaluators navigate complex contexts with diverse stakeholders to produce a trustworthy and relevant process and a story of transformative change.

Much of the debate about credible evidence in the fields of evaluation and policy focuses on methods and measurement. Ken Jones, Eugenia Gwynn, and Allison Teeter, in their article, “Quantitative or Qualitative: Selecting the Right Methodological Approach for Credible Evidence,” describe how quantitative and qualitative methods—numbers and narratives—provide unique and complementary evidence for program accountability and improvement. In his article, “Measurement and Credible Evidence in Extension Evaluations,” Marc Braverman describes the qualities of good measures, advantages of matching measures to evaluation questions, possibilities and limitations of common measures for Extension program evaluation.

Credible evidence is never a “one-size-fits-all” proposition. In their article, “Credible and Actionable Evidence Across Extension Program Areas: A Case Example,” Mary Marczak, Emily Becher, and Patricia Olson illustrate how criteria for valuing and strategies for gathering evidence differ horizontally across Extension disciplines. Nick Place and colleagues explore differences vertically in their article, “Credible and Actionable Evidence Across Stakeholder Levels of the Cooperative Extension System,” as stakeholders at local, state, and federal levels value different kinds of evidence and communication about results.

Evidence often becomes more credible because of the way it is collected and interpreted to stakeholders. In the article, “Communicating with Data: Telling the Extension Story in Credible and Actionable Ways,” Diane Craig and Ruth Borger address the organizational and professional process of “telling the story,” including the use of traditional and emerging media.

Credible evidence can reach no higher than the evaluation capacity of Extension professionals and organizations. Chelsea Heatherington, Cheryl Eschbach, and Courtney Cuthbertson discuss key skills and strategic options for evaluation capacity building for generating and using evidence with a wide range of stakeholders in their article, “How Evaluation Capacity Building Grows Credible and Actionable Evidence for Cooperative Extension Programs.”

As co-editors of this special edition of *JHSE*, we close out the edition with reflections on these diverse themes, the challenges in using, generating, and interpreting credible evidence, and the implications of the credible and actionable evidence discussion for the future of Extension.

## References

- American Evaluation Association. (2003). *American Evaluation Association response to U.S. Department of Education Notice of proposed priority, Federal Register RIN 1890-ZA00, November 4, 2003: "Scientifically based evaluation methods."* Retrieved from <https://www.eval.org/p/cm/ld/fid=95>
- American Evaluation Association. (2018). *Guiding principles for evaluators*. Retrieved from <https://www.eval.org/p/cm/ld/fid=51>
- Arnold, M. E., & Cater, M. (2016). Program theory and quality matter: Changing the course of Extension program evaluation. *Journal of Extension*, 54(1), Article 1FEA1. Retrieved from <https://joe.org/joe/2016february/a1.php>
- Bamberger, M., Rugh, J., & Mabry, L. (2012). *Real world evaluation. Working under budget, time, data, and political constraints* (2<sup>nd</sup> ed.). Newbury Park, CA: Sage.
- Baker, M. A., & Robinson, J. S. (2018). The effect of two different pedagogical delivery methods on students' retention of knowledge over time. *Journal of Agricultural Education*, 59(1), 100–118. doi:10.5032/jae.2018.01100
- Bauer, M. S., Damschroder, L., Hagedorn, H., Smith, J., & Kilbourne, A. M. (2015). An introduction to implementation science for the non-specialist. *BMC Psychology*, 3(1), 32–43. doi:10.1186/s40359-015-0089-9
- Blyth, D. (2011). The future of youth development: Multiple wisdoms, alternate pathways and aligned accountability. *Journal of Youth Development*, 6(3), Article 110603FA011. Retrieved from <http://jyd.pitt.edu/ojs/jyd/article/view/182>
- Brandt, A. M. (2012). Inventing conflicts of interest: A history of tobacco industry tactics. *American Journal of Public Health*, 102(1), 63–71. doi:10.2105/AJPH.2011.300292
- Braverman, M. T., & Engle, M. (2009). Theory and rigor in Extension program evaluation planning. *Journal of Extension*, 47(3), Article 3FEA1. Retrieved from [https://www.joe.org/joe/2009june/pdf/JOE\\_v47\\_3a1.pdf](https://www.joe.org/joe/2009june/pdf/JOE_v47_3a1.pdf)
- Bull, N. H., Cote, L. S., Warner, P. D., & McKinnie, M. R. (2004). Is Extension relevant for the 21st century? *Journal of Extension*, 42(6), Article 6COM2. Retrieved from <https://www.joe.org/joe/2004december/comm2.php>
- Campbell, D. T. (1991). Methods for the experimenting society. *American Journal of Evaluation*, 12(3), 223–260. Retrieved from doi:10.1177/109821409101200304
- Centers for Disease Control and Prevention. (1999). Framework for program evaluation in public health. *Morbidity and Mortality Weekly Report: Recommendations and Reports*, 48(RR-11). Retrieved from <https://www.cdc.gov/mmwr/PDF/rr/rr4811.pdf>
- Chen, H. T., Donaldson, S. I., & Mark, M. M. (2011). Validity frameworks for outcome evaluation. In H. T. Chen, S. I. Donaldson, & M. M. Mark (Eds.), *Advancing validity in outcome evaluation: Theory and practice. New Directions for Evaluation*, 130, 5–16. doi:10.1002/ev.361

- Christie, C. A., & Fleischer, D. (2015). Social inquiry paradigms as a frame for the debate on credible evidence. 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. 27–38). Thousand Oaks, CA: Sage.
- DeSchutter, O. (2011). The right of everyone to enjoy the benefits of scientific progress and the right to food: From conflict to complementarity. *Human Rights Quarterly*, 33(2), 304–350. doi:10.1353/hrq.2011.0020
- Donaldson, S. I. (2015). Examining the backbone of contemporary evaluation practice: Credible and actionable evidence. 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. 3–26). 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.
- Duerden, M. D., & Witt, P. A. (2012). Assessing program implementation: What it is, why it's important, and how to do it. *Journal of Extension*, 50(1), Article 1FEA4. Retrieved from [https://www.joe.org/joe/2012february/pdf/JOE\\_v50\\_1a4.pdf](https://www.joe.org/joe/2012february/pdf/JOE_v50_1a4.pdf)
- Eisenberg, R. S. & Nelson, R. R. (2002). Public vs. proprietary science: A fruitful tension? *Academic Medicine*, 77(12), 1392–1399.
- Evans, D. (2002). Hierarchy of evidence: A framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing*, 12(1), 77–84. doi:10.1046/j.1365-2702.2003.00662.x
- Franz, N. (2013). Improving Extension programs: Putting public value stories and statements to work. *Journal of Extension*, 51(3), Article 3TOT1. Retrieved from <http://www.joe.org/joe/2013june/tt1.php>
- Franz, N., & Archibald, T. (2018). Four approaches to building Extension program evaluation capacity. *Journal of Extension*, 56(4), Article 4TOT5. Retrieved from <https://joe.org/joe/2018august/tt5.php>
- Franz, N. K., 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 <https://joe.org/joe/2014august/comm3.php>
- Franz, N. K., Garst, B. A., & Gagnon, R. (2015). The Cooperative Extension program development model: Adapting to a changing context. *Journal of Human Sciences and Extension*, 3(2), 3–12.
- Franz, N., & Townson, L. (2008). The nature of complex organizations: The case of Cooperative Extension. In M. T. Braverman, M. Engle, M. E. Arnold, & R. A. Rennekamp (Eds.), *Program evaluation in a complex organizational system: Lessons from Cooperative Extension*. *New Directions for Evaluation*, 120, 5–14. doi:10.1002/ev.272
- Fraser, M. W., Richman, J. M., Galinsky, M. J., & Day, S. H. (2009). *Intervention research*. New York, NY: Oxford University Press.

- Gill, S., Kuwahara, R., & Wilce, M. (2016). Through a culturally competent lens: Why the program evaluation standards matter. *Health Promotion Practice, 17*(1), 5–8, doi:10.1177/1524839915616364
- Greene, J. C. (2015). How evidence earns credibility in evaluation. 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. 205–220). Thousand Oaks, CA: Sage.
- Hendricks, P. A. (1998). *Developing youth curriculum using the targeting life skills model: Incorporating developmentally appropriate learning opportunities to assess impact of life skill development*. Ames, IA: Iowa State University.
- Honadle, B. W., Zapata, M. A., Auffrey, C., vom Hofe, R., & Looye, J. (2014). Developmental evaluation and the ‘Stronger Economies Together’ initiative in the United States. *Evaluation and Program Planning, 43*, 64–72. doi:10.1016/j.evalprogplan.2013.11.004
- Horgan, J. (2012). What Thomas Kuhn really thought about scientific “truth.” Retrieved from <https://blogs.scientificamerican.com/cross-check/what-thomas-kuhn-really-thought-about-scientific-truth/>
- Institute of Education Sciences. (2019). *What works clearinghouse*. Retrieved from <https://ies.ed.gov/ncee/wwc/>
- Julnes, G., & Rog, D. (2015). Actionable evidence in context: Contextual influences on adequacy and appropriateness of method choice in evaluation. 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. 221–258). Thousand Oaks, CA: Sage.
- Kellogg Commission on the Future of State and Land-Grant Universities. (1999). *Returning to our roots: The engaged institution* [Report 3]. Washington, DC: National Association of State Universities and Land-Grant Colleges. Retrieved from <http://www.aplu.org/library/returning-to-our-roots-the-engaged-institution/file>
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
- 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.
- Lerner, R. J., Lerner, J. V., & Colleagues. (2011). *The positive development of youth: Report of the findings for the first seven years of the 4-H study of positive youth development*. Chevy Chase, MD: National 4-H Council. Retrieved from <https://4-h.org/wp-content/uploads/2016/02/4-H-Study-of-Positive-Youth-Development-Full-Report.pdf>
- 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 public and nonprofit policies and programs*. San Francisco, CA: Jossey-Bass.
- Mertens, D. M. & Hesse-Biber, S. (2013). Mixed methods and credibility of evidence in evaluation. In D. M. Mertens & S. Hesse-Biber (Eds.), *Mixed methods and credibility of evidence in evaluation. New Directions for Evaluation, 138*, 5–13. doi:10.1002/ev.20053
- Miller, R. L. (2015). How people judge the credibility of information: Lessons for evaluation from cognitive and information sciences. 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. 39–61). Thousand Oaks, CA: Sage.
- Milstein, B., Wetterhall, S., & CDC Evaluation Working Group. (2000). A framework featuring steps and standards for program evaluation. *Health Promotion Practice, 1*(3), 221–228.
- National Institute of Food and Agriculture. (2014). *Agricultural Research, Extension, and Education Reform Act of 1998*. Retrieved from <https://nifa.usda.gov/resource/agricultural-research-extension-and-education-reform-act-1998>
- National Institute for Food and Agriculture. (2018). *Cooperative Extension System*. Retrieved from <https://nifa.usda.gov/cooperative-extension-system>
- Owen, J. M., & Rogers, P. J. (1999). *Program evaluation: Forms and approaches*. Thousand Oaks, CA: Sage.
- Patton, M. Q. (2017). *Principles-focused evaluation*. Newbury Park, CA: Sage.
- Patton, M. Q., & Blandin Foundation (2014). *Mountain of accountability: Pursuing mission through learning, exploration and development*. Grand Rapids, MN: Blandin Foundation. Retrieved from <https://blandinfoundation.org/articles/mountain-of-accountability/>
- Pennock, R. T., & O'Rourke, M. (2017). Developing a scientific virtue-based approach to science ethics training. *Science and Engineering Ethics, 23*(1), 243–262. doi:10.1007/s11948-016-9757-2
- Reynolds, H. L., Smith, A. A., & Farmer, J. R. (2014). Think globally, research locally: Paradigms and place in agroecological research. *American Journal of Botany, 101*(10), 1631–1639. doi:10.3732/ajb.1400146
- Ripley, J., Cummings, S., Lockett, L., Pope, P., Wright, M., Payne, M., Keith, L., & Murphrey, T. (2011). *Creating excellent programs* [Bulletin E-345]. College Station, TX: Texas A&M AgriLife Extension.
- Rockwell, K., & Bennett, C. (2004). *Targeting outcomes of programs (TOP): A hierarchy for targeting outcomes and evaluating their achievement*. Retrieved from <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1047&context=aglecfacpub>
- Ruhe, V., & Boudreau, J. D. (2013). The 2011 program evaluation standards: A framework for quality in medical education programme evaluations. *Journal of Evaluation in Clinical Practice, 19*(5), 925–932. doi:10.1111/j.1365-2753.2012.01879.x

- Scriven, M. (2007). The logic of evaluation. In H. V. Hanson et al. (Eds.). *Dissensus and the search for common ground* (pp. 1–16). Windsor, ON: OSSA. Retrieved from <https://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=1390&context=ossaarchive>
- Sharpe, G. (2011). A review of program theory and theory-based evaluations. *American International Journal of Contemporary Research*, 1(3), 72–75.
- Stufflebeam, D. (2001). Evaluation models. *New Directions for Evaluation*, 89, 7–98. doi:10.1002/ev.3
- Taylor-Powell, E., Steele, S., & Douglass, M. (1996). *Planning a program evaluation* [G3658-1]. Madison, WI: University of Wisconsin Extension, Program Development and Evaluation. Retrieved from <http://learningstore.uwex.edu/assets/pdfs/G3658-1.PDF>
- Valenzuela, H. (2016). Agroecology: A global paradigm to challenge mainstream industrial agriculture. *Horticulturae*, 2(1), 1–11. doi:10.3390/horticulturae2010002
- Virgona, J., & Daniel, G. (2011). Evidence-based agriculture: Can we get there? *Agricultural Science*, 23(1), 19–25.
- Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2<sup>nd</sup> ed.). Upper Saddle River, NJ: Prentice-Hall.
- Welsh, R., & Rivers, R. Y. (2011). Environmental management strategies in agriculture. *Agriculture and Human Values*, 28(3), 297–302. doi:10.1007/s10460-010-9285-7
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3<sup>rd</sup> ed.). Thousand Oaks, CA: Sage. Retrieved from <http://www.jcsee.org/program-evaluation-standards-statements>

*Benjamin Silliman*, Professor and Extension Specialist, Department of Agricultural and Human Sciences, North Carolina State University, specializes in program evaluation and youth development. He received his Ph.D. in Human Ecology from Kansas State University and Evaluation Practice Certificate from The Evaluators Institute.

*Scott R. Cummings*, Dr.P.H., is head of Organizational Development for the Texas A&M AgriLife Extension Service and Professor in the Department of Agricultural Leadership, Education, and Communications at Texas A&M University.

### Acknowledgments

The first author gratefully acknowledges helpful insights on program and delivery issues from Dr. Ben Chapman and Dr. Marshall Baker and guidance from Dr. Richard Poling on an earlier draft of the article. Larger insights on concepts and expression are *te Deum gloria*. The authors gratefully acknowledge support and guidance from the General Editors, Dr. Donna Peterson and Dr. Richard Poling throughout the development of the special edition.

## Whose Extension Counts? A Plurality of Extensions and Their Implications for Credible Evidence Debates

Thomas Archibald

Virginia Tech

*The debate over what counts as credible evidence often occurs on a methodological level (i.e., about what technical applications of systematic inquiry provide believable, justifiable claims about a program). Less often, it occurs on an epistemological level (i.e., about what ways of knowing are appropriate for making claims about a program). Even less often, the debate touches on ontological concerns (i.e., about what conceptualizations of reality, in general or in relation to a specific program, are in play when we wish to make claims about that program). For example, whether we understand Extension to be a vehicle for the dissemination of scientific knowledge or a site of grassroots democracy matters when we seek to evaluate Extension with credibility. The purpose of this paper is to examine the credible evidence debates through an ontological lens, showing why and how different narratives (or different realities) of Extension must be considered when we seek credible evidence about Extension.*

*Keywords:* credible evidence, randomized controlled trials, evaluation, ontological politics

***“[T]he reality we live with is one performed in a variety of practices. The radical consequence of this is that reality itself is multiple. An implication of this might be that there are options between the various versions of an object: which one to perform? But if this were the case then we would need to ask where such options might be situated and what was at stake when a decision between alternative performances was made.”***

—Annemarie Mol (1999, p. 74, emphasis in the original)

### Introduction

For the past two decades, the question of what counts as credible evidence in program evaluation and applied social science research has fomented a considerable amount of debate. In particular, divergent perspectives on whether randomized controlled trials (RCTs) should be considered the “gold standard” for producing credible evidence have occupied a central position in the debate (Donaldson, Christie, & Mark, 2009). In more recent years, quasi-experimental designs such as regression discontinuity design have also been lauded for their ability to generate evidence of

---

Direct correspondence to Thomas Archibald at [tgarch@vt.edu](mailto:tgarch@vt.edu)

impact just as well as RCTs (Pohl, Steiner, Eisermann, Soellner, & Cook, 2009). Long before these recent methodological skirmishes (hundreds of years before), questions about how to generate valid knowledge of the world around us—and specifically about the role of experimentation in that process—animated the scientific and aristocratic classes alike. For instance, Shapin and Schaffer (1985) examined the dispute between Robert Boyle and Thomas Hobbes over Boyle's air pump experiments in the 1660s, exploring acceptable methods of knowledge production and the societal factors related to different knowledge systems.

These seemingly esoteric methodological debates about credible evidence—be they historical or contemporary—are in fact fundamentally important questions about life in general and professional practice more specifically. This point is summed up by Trochim, who said:

The gold standard debate is one of the most important controversies in contemporary evaluation and applied social sciences. It's at the heart of how we go about trying to understand the world around us. It is integrally related to what we think science is and how it relates to practice. There is a lot at stake. (W. Trochim, unpublished speech transcript, September 10, 2007)

Along those same lines, in equally emphatic terms, Scriven (2008) wrote, “This issue is not a mere academic dispute, and should be treated as one involving the welfare of very many people, not just the egos of a few” (p. 24).

In this paper, I endeavor to show why so much is at stake in these contentious exchanges about credible evidence. In particular, I wade into the debates as they pertain to the context of Cooperative Extension. However, eschewing a frequently traced line of reasoning focused on the apparent merits and superiority of particular methodologies and designs, such as the RCT, I instead suggest that the debate cannot possibly be resolved unless we reconsider the very nature of the evaluand or object of inquiry—in this case, “Cooperative Extension.” In other words, I seek to recast the credible evidence debate as being just as much (if not more so) about *ontology*—the philosophical study of reality—as it is about epistemology and methodology. Going further, I suggest that the existence of multiple narratives about what Extension even *is*—the plurality of Extensions—precludes and prevents the possibility of any once-and-for-all summary statement about what counts as credible evidence in Extension evaluation.

In the remainder of this paper, I first briefly review the credible evidence debates in general, paying special attention to arguments that foreground ontological questions as an integral way of engaging with this topic. Then, I review a sampling of literature that opens up the possibility of seeing Extension as a pluralistic and shifting phenomenon or object rather than as a stable and agreed-upon one. Finally, I juxtapose these two bodies of literature to justify my central claim. The potential significance and practical applicability of this rather theoretical article are to help us better understand why it sometimes seems as though we are “talking past each other” when

debating what counts as credible evidence in Extension and related programmatic endeavors. As a result, I hope, we can reorient those debates to produce less heat and more light.

### A Brief Review of the Credible Evidence Debates

A large volume of work has been published in the fields of education research and evaluation about what counts as “evidence,” and especially about the privileged place of experimental and quasi-experimental designs in the production of credible evidence (Donaldson et al., 2009; Morrison, 2009; Mosteller & Boruch, 2002; Nelson & Campbell, 2017; Scriven, 2008; U.S. Department of Education, 2003; Walters, Lareau, & Ranis, 2009)—for Extension practitioners and other readers who are new to this topic, these works can provide background information on this issue, which is beyond the scope of this paper. In the field of program evaluation, an acute moment of conflict about what counts as credible evidence occurred in 2003 when the U.S. Department of Education’s Institute of Education Sciences (IES) “declared a rather wholesale commitment to privileging experimental and some types of quasi-experimental designs over other methods in applied research and evaluation funding competitions” (Donaldson, 2009, p. 7). As described by Donaldson (2009), the American Evaluation Association (AEA) submitted a strongly-worded public statement criticizing the enshrinement of RCTs as the best methodology or design to provide evidence of program effectiveness; in turn, a smaller group of prominent AEA members published a public rebuttal and refutation of AEA’s statement, signaling support for the RCT as gold standard. In the fifteen years since, the debate has waxed and waned, while the preeminence of RCTs has been ensconced in some notable and prominent places.

For example, the What Works Clearinghouse (WWC), managed by the IES, is a repository of evidence-based programs<sup>1</sup> in education shown to be effective through one or more high-quality RCTs. The WWC is explicit about its hierarchical view of research and evaluation designs to address the question of program effectiveness: “In order for a study to be rated as meeting evidence standards . . . , it must employ one of the following types of research designs: a randomized controlled trial or a quasi-experiment (including quasi-experiments with equating, regression discontinuity designs, and single-case designs)” (WWC, 2008, p. 5). The Campbell Collaboration—an international network that supports the preparation and dissemination of systematic reviews of evidence on the effectiveness of social programs, policies, and practices—has also established methodological standards that prescribe what constitutes the best available evidence about the effects of focal interventions, placing a clear emphasis on RCTs:

The critical feature of the research methods in this regard is the ability of the basic design to yield an unbiased estimate of the effects on the target outcomes relative to a defined counterfactual condition, that is, the *internal validity* of the research design (Shadish, Cook, & Campbell, 2002). With rare exceptions, the best evidence by this standard is

---

<sup>1</sup> For more on the debates specifically about contentious terms such as “evidence-based” or “evidence-informed” programs and practice, see Archibald (2015) and Nelson and Campbell (2017), among others.

provided by randomized controlled trials (RCTs). (Campbell Collaboration, 2017, p. 9, emphasis in the original)

In the field of international development (which includes interventions designed to strengthen rural advisory and extension systems), the privileged place of the RCT design has been championed by both the International Initiative for Impact Evaluation (3ie) and the Abdul Latif Jameel Poverty Action Lab (J-PAL) at the Massachusetts Institute of Technology (MIT). J-PAL sees “rigorous research” as essential to finding solutions to the world’s greatest challenges. For J-PAL, rigorous research, also called “high-quality impact evaluation,” is that (and only that) which employs randomization. Foreshadowing the ontological confusion that I examine in greater detail later in this paper, J-PAL’s director wrote, “Just as randomized trials for pharmaceuticals revolutionized medicine in the 20th Century, randomized evaluations have the potential to revolutionize social policy during the 21st” (Duflo & Kremer, 2003, p. 32). This quotation is an example of ontological confusion because the reality of pharmaceuticals and other insentient, physical materials—how they act and interact—is obviously categorically different from the reality of humans, a point many critics of the RCT as gold standard frequently evoke (e.g., Biesta, 2010; Scriven, 2008).

Randomized controlled trials did indeed revolutionize medicine (Baron, 2018). They had the same transformative effects in agricultural research, where much of the statistical analyses behind the RCT were originally developed (Box, 1978). The underlying principles of the RCT design are relatively straightforward; the design was created to increase the internal validity of study conclusions, to reduce the threat of bias in estimating the average effect of a specific treatment on a quantitative variable of interest. In its simplest form, the design is implemented by randomly allocating individual units of analysis (i.e., plants, people, schools, villages) to a treatment condition or to a control or comparison condition, absent the treatment being studied. As described by Scriven (2008), the RCT

is an experimental design involving at least two groups of subjects, the control group and the experimental group (a.k.a. study group, or treatment group), between which the subjects are distributed by a strictly random process (i.e., one with no exceptions), and which are not further identified or distinguished by any common factor besides the application of the experimental treatment to the experimental group. (p. 11)

The power of randomization is ascribed to its ability to methodologically address the “counterfactual question: how would individuals who participated in the program have fared in the absence of the program? How would those who were not exposed to the program have fared in the presence of the program?” (Duflo & Kremer, 2003, p. 3). Addressing the counterfactual question this way can be useful, in some cases, to answer some evaluation questions. Yet the RCT has been at the center of so much controversy over the past 20 years because there is a tendency—like in IES and J-PAL—to constitute it as “the best” and most credible type of evaluation, earning it the “gold standard” moniker.

### Why the RCT is Not the Gold Standard

It is not clear who first referred to the RCT as the gold standard, yet many prominent proponents of experimental designs (e.g., IES and J-PAL) reinforce this hegemonic superiority of the RCT atop the methodological hierarchy through rhetorical devices like, “often considered the gold standard” (e.g., Akobeng, 2005; Coalition for Evidence-Based Policy, 2003; Pillemer, 2011). What is clear is why many other prominent methodologists and academics—including those who promote RCTs—resist and reject the notion that the RCT is the gold standard. A brief overview of why the RCT ought not to be called the gold standard will help lay the foundation for my claim about the primacy of ontological questions when seeking credible evidence in Extension.

Howard White, founding director of 3ie and current Chief Executive Officer of the Campbell Collaboration (and thus a prominent proponent of RCTs), argues against the existence of a hierarchy of methods; evaluations should be led by the issues at hand, not by methods, and “having determined the evaluation questions, the best available method should then be used to answer them” (White, 2010, p. 162). In addition, White (2010) foreshadows a major claim in this article, that “there is no point in methodological debates unless they agree [on] a common starting point” (p. 153). The evaluation community is working from different assumptions about and definitions of “impact.” For some, impact refers to the final or most distal level of the program’s theory of change. In these cases, there is no way to say a priori which evaluation design or method is most appropriate. For others, impact

is defined as the difference in the indicator of interest ( $Y$ ) with the intervention ( $Y_1$ ) and without the intervention ( $Y_0$ ). . . . An impact evaluation is a study which tackles the issue of attribution by identifying the counterfactual value of  $Y$  ( $Y_0$ ) in a rigorous manner. (White, 2010, p. 154)

Using this definition of impact, the RCT is arguably the best or most appropriate method, though what is meant by “in a rigorous manner” is still open to debate. White himself suggests that rigorous RCTs should include a qualitative component to help elucidate not just *whether* a program or policy works, but also *how* it works. While White stops short of considering the plurality of evaluands and how that might matter for the question of what counts as credible evidence (i.e., of impact), he does foreground the importance of getting clear on the purposes of the inquiry and of letting that drive methodological decisions.

Angus Deaton (2010), winner of the 2015 Nobel Prize in Economics, argues that “experiments have no special ability to produce more credible knowledge than other methods, and that actual experiments are frequently subject to practical problems that undermine any claims to statistical or epistemic superiority” (p. 424). Scriven (2008) reiterates this second point, claiming the RCT has “essentially zero practical application to the field of human affairs” (p. 12) due to such implementation problems as being zero-blind rather than double-blind, among other limitations. Deaton (2010) claims evidence from randomized controlled trials can have no special priority:

Randomization is not a gold standard because “there is no gold standard” [citing contemporary philosopher Nancy Cartwright]. . . . Randomized controlled trials cannot automatically trump other evidence, they do not occupy any special place in some hierarchy of evidence, nor does it make sense to refer to them as “hard” while other methods are “soft.” These rhetorical devices are just that; metaphor is not argument, nor does endless repetition make it so. (p. 426)

Touching, at least tangentially, on the ontological faces of the debate about credible evidence, Deaton also points out the important distinction between macro- and microeconomic development interventions, and the difficult (if not impossible) task of parsing out the endogeneity or independence of the variables being studied. In other words, to make a reasonable claim that “RCTs are the best for generating credible evidence of impact,” one first must know if the evaluand consists of macro- or micro-processes, and if there is any way to know if randomization can really isolate operationalized variables. These conclusions apply, more generally, to the questions I present in the next section on the plurality of Extensions.

One additional noteworthy critic of the standard notions of RCTs as the best (or sole) fount of credible evidence is Gert Biesta, a policy-oriented philosopher of education based at Brunel University London. Biesta (2010) explicated the epistemological, ontological, and praxeological assumptions that inhere “evidence-based education.” The fundamental problem he identifies in the ontological domain is that “talk about ‘what works’ . . . operates on the assumption of a mechanistic ontology that is actually the exception, not the norm in the domain of human interaction” (Biesta, 2010, p. 497). A mechanistic ontology, on which the technological view of education (and Extension) is based,

relies on the idea that education can in some sense be conceived as a machinery where there are inputs, mediating variables and outcomes. The technological ambition, as mentioned, is to make the connection between inputs and outputs as secure as possible so that education can begin to operate as a deterministic machine. (Biesta, 2015, p. 16)

At the level of epistemology, this mechanistic ontology is associated with positivist technical-rationalistic assumptions about knowledge and about its role in guiding professional practice, whereby “professional activity consists in instrumental problem solving made rigorous by the application of scientific theory and technique” (Schön, 1983, p. 21).

Contrary to this mechanistic ontology, drawing from systems thinking and complexity theory, we see that educational systems (such as Cooperative Extension) are perhaps better characterized as open, semiotic, recursive systems. As described by Biesta (2015):

Education is an open system because it is in interaction with its environment rather than being completely disconnected from it. Education is a semiotic system because the interactions within the system are not interactions of physical push and pull, but of



communication, meaning making and interpretation. In addition, education is a recursive system because of the way in which the system evolves and feeds back into the further operation of the system—which, in more everyday terms, has to do with the fact that the “elements” in the system, teachers and students, are thinking beings with agency, that is, beings who can draw conclusions and can act upon those conclusions. (p. 16)

In the above quotation, replace “education” with “Extension,” replace “teacher” with “agent” or “educator,” and replace “students” with “program participants” and the underlying logic holds, suggesting that the ontology of Extension is more open, semiotic, and recursive rather than closed, deterministic, and mechanistic. Given these descriptions of what it means to be an open, semiotic, recursive system, the claim that the mechanistic ontology of “evidence-based education” is the exception rather than the rule seems warranted. Doing greenhouse trials, or maybe even field trials, on the best way to control mildew in potatoes is mechanistic; yet the social processes through which the Extension professional interacts with the potato producer to communicate about the knowledge derived from those trials are not. Rather, such processes have to do with relationships, meaning-making, and dynamic context-content interactions in a complex socio-political-economic system.

However, rhetorical efforts to equate research in medicine or agronomy to research in non-formal education and community development—efforts like those represented by the quotation from Duflo and Kremer (2003) shared above—apparently overlook the ways in which “the dynamics of education are fundamentally different from the dynamics of, say, potato growing or chemistry” (Biesta, 2010, p. 497). Biesta (2010) calls this the “efficacy deficit” of the evidence-based movement, “indicating that in the social domain interventions do not generate effects in a mechanistic or deterministic way, but through processes that . . . are open so that the connections between intervention and effect are non-linear” (p. 497).

### **A Plurality of Extensions**

If RCTs are not the gold standard, and if what counts as credible evidence depends not just on methodological norms and precepts, but also on the ontological characteristics of the object of inquiry, then what are the implications for our quest for credible evidence in Cooperative Extension? In this section, I review a small sample of literature that, from various perspectives, can help us rethink the seemingly settled fundamental notion of what Extension even *is*. Also, this literature helps us grapple with the question of whether there exists one solitary version of Extension, or whether we might be better served by recognizing and allowing for a plurality of Extensions.

For instance, although he was not directly addressing the uniquely American institution of Cooperative Extension, renowned educator Paulo Freire (1973) weighed in on the ontological foundations of extension by engaging in a rhetorical critique (via semantic analysis) of the very term “extension.” Through that analysis, Freire noted that often, “the role of extension agents is

to extend, not their hands, but their knowledge and their technical capacities” (p. 94). From a relational perspective, “the concept of extension which is characterized by the transference of techniques and knowledge is in direct contradiction to a truly humanist outlook” (Freire, 1973, p. 94), since it tends to be overly mechanistic (and messianic), reifying people as objects and negating the reflection and action that characterizes authentic educative encounters. The “field of association” (i.e., the discursive construction) of the term “extension” evokes a one-way directional transmission from one in a privileged position as the knower of that which is better to those that are inferior and passive. This is why Freire preferred the term “communication” over “extension.” The ontological nature of communication is more relational and dialogic, which is why Freire favored it.

The same dynamic tension is present in some of the earliest histories of Cooperative Extension in the United States. For instance, in *The People’s Colleges: A History of the New York State Extension Service in Cornell University and the State, 1876-1948*, Ruby Green Smith (1949/2013) foreshadowed this line of analysis that we find in Freire’s writing. As Peters (2017) points out, the dominant conceptualization of Cooperative Extension “is that extension was and still is a one-way conduit for transferring technology and information” (p. 73). This overly simplified storying of Extension is manifest in the literature spanning decades (e.g., Campbell, 1995; Eddy, 1957; Edmond, 1978; Mumford, 1940; National Research Council, 1996; Nevins, 1962; Rasmussen, 1989; Ross, 1942), and is implied by some of the slogans or taglines of Cooperative Extension systems (e.g., “Extending knowledge, changing lives;” “Putting knowledge to work;” “Taking the university to the people”).

Complicating this version of Extension, Ruby Green Smith emphasizes, like Freire, the more relational and dialogic elements of Cooperative Extension:

There is vigorous reciprocity in the Extension Service because it is with the people, as well as “of the people, by the people, and for the people.” It not only carries knowledge from the State Colleges to the people, but it also works in reverse: it carries from the people to their State Colleges practical knowledge whose workability has been tested on farms, in industry, in homes, and in communities. In ideal extension work, science and art meet life and practice. Mutual benefits result for the people and for the educational institutions they support. Thus the Extension Service develops not only better agriculture, industries, homes, and communities, but better colleges. (Smith, 1949/2013, p. ix)

Already, we see an ontological divide, in practice, as to whether Extension is about disseminating scientific knowledge or rather is about reciprocity (or both, or something in-between). What counts as credible evidence of Extension’s impacts cannot be established if we ignore this ontological uncertainty (or plurality).

While science-based technical facts and skills have been and will doubtlessly remain an important part of Cooperative Extension's reality, Smith impels us to think more broadly about purposes and thus about the core ontological foundations of Cooperative Extension, touching on the importance of practical wisdom and democratic living:

Extension workers need to have faith in spiritual values and to recognize the human relationships that contribute to what the ancient Greeks called “the good life.” They should believe that in the kind of homes, farms, and industries which are the goals of Extension service “man [sic] cannot live by bread alone;” that it is not enough for people to have food, shelter, and clothing—that they aspire also to find appreciation, respect for individuality and human dignity, affection, ideals, and opportunities. These are the satisfactions that belong to democratic living. (Smith, 1949/2013, p. 544)

Smith is getting at the virtues of Aristotelian *phronesis*, described by Flyvbjerg (2002) as an intellectual virtue of reasoned action that “concerns values, and goes beyond analytical, scientific knowledge (*episteme*) and technical knowledge or know-how (*techne*). It involves judgments and decisions made in the manner of a virtuoso social actor” (p. 26).

Smith's historical account, written in 1949, provides evidence that thinking of Extension as a relational, dialogic space for the exchange of knowledge in the pursuit of community development is not some new fad or some contemporary reimagining of what Extension is “really supposed to be.” Another such historical backing is provided by Shaffer (2017), who presents the important role of discussion groups and deliberative democracy in the earlier years of Cooperative Extension. Drawing on the *Report of the Commission on Country Life* (1911), texts from M. L. Wilson (assistant secretary of the USDA), and other extensive archival research, Shaffer (2017) shows how Extension has long “put into practice the role of supporting and catalyzing change in communities as facilitators of citizens' own agency in response to public issues” (para. 1) while also acknowledging that “There has always been a tension between a technocratic mindset and an approach that is more democratic, relational, and engaging” (para. 2). According to Shaffer, Wilson “championed efforts to approach Extension's work through a democratic lens, building on a belief that “free and full discussion [was] the archstone of democracy” (Wilson, 1939, p. 145) and that Extension agents could play a critical role in facilitating citizen discussion about a range of public issues” (2007, para. 4).

Contrast these historical, foundational descriptions of Extension with a contemporary perspective on an idealized notion of the role of research evidence in Extension—a perspective that, in light of these historical texts, seems rather shortsighted and misguided: Research Use by Cooperative Extension Educators in New York State (Hamilton, Chen, Pillemer, & Meador, 2013). The entire first paragraph of that article suggests that the authors have not read Smith or Wilson, or, if they have, that they have discounted those fundamental writings on the ontological reality of Extension as a relational and dialogic setting for deliberative democracy on public issues:

The Extension system exists to disseminate the findings of research beyond the academic community to practitioners, policy makers, and the general public. Extension educators thus serve as a bridge between scholars and the wider community. For example, scientists may find a way to apply pesticides more precisely or discover the benefits of serving low-fat milk to children. Extension staff then educate farmers or parents, respectively, about the new findings. These examples illustrate what Nutley, Walter, & Davies (2007) called the “knowledge-driven model” of research utilization in policy and practice. (Hamilton et al., 2013)

Empirical studies of the use of research knowledge in Extension education are indeed a welcome addition to the literature, but as the above paragraph suggests, Hamilton and his co-authors espouse a mechanistic ontology of Extension that harkens back to that which Freire, Biesta, Smith, and others have roundly rejected. Is Extension really knowledge-driven, or is it relationship-driven, or is it both? In any case, there are implications for what counts as credible evidence of successful Extension impacts, and for how we imagine research evidence to inform Extension practice.

Hamilton and his colleagues focus their article on the (general lack of) use of evidence-based programs (EBPs), that they describe as “becoming increasingly prominent to bridge the gap between research and practice . . . programs or curricula that have been *rigorously tested* to validate their effectiveness” (Hamilton et al., 2013, emphasis added). This sentiment evokes a notion I encountered as part of a qualitative study of efforts to make Cooperative Extension more “evidence-based” (see Archibald, 2015); in that study, one Extension administrator suggested to me during an interview that 4-H (the youth development component of Cooperative Extension) should perhaps no longer be allowed (i.e., funded) to implement programs (like livestock judging at the county fair), unless those programs could become more evidence-based, meaning that they had undergone at least one RCT that showed positive impact on the primary quantitative variable of interest. The understanding of Extension and the role of credible evidence in Extension manifest in that administrator’s suggestion throws the plurality of Extensions—and the stark ontological and epistemological politics and their consequences—into sharp relief. In similar ways, in other areas of Extension such as agriculture and natural resource programs, many practitioners focus on technical recommendations about a given content area; many times, such recommendations are predicated on experiments that do indeed operationalize a mechanistic ontology. However, the social side of Extension in such areas requires a different type of practice (see Peters, 2006), and thus a different type of evidence.

In other words, these debates raise the question: Should Extension be perceived and performed as an “infrastructure for the dissemination of scientific information or as a site of grassroots knowledge sharing” (Archibald, 2015, p. 145)? How we “see” Extension has stark consequences for the professional trajectories of Extension educators and also has real material implications for the lives and livelihoods of the community members Extension purports to serve. This is a

question about participation in the processes of inquiry and action that affect people's lives, echoing the title of Robert Chambers' (1997) book, *Whose Reality Counts?* This is fundamentally a question of ontological politics (Mol, 1999); it has to do with how "the real" is implicated in the "political" and vice versa, where "political" refers not to electoral or partisan politics but to the mundane, minute, active everyday processes of shaping and shifting contested options between varying versions of an object (such as Cooperative Extension).

### Conclusion

To sum up the arguments presented above, how we "see" Extension is a methodological and epistemological question about what counts as credible evidence to best know, to best render legible all that complexity that inheres Cooperative Extension programming. Yet "seeing" in one way (i.e., via an RCT) does political work of an ontological kind—it makes Extension to be more one way (i.e., technical-rationalistic, expert-driven, one-way dissemination of scientific knowledge) rather than some other way (i.e., relational, dialogic engagement in deliberative democracy to collaboratively address community issues). Instead of fixating on these apparent dichotomies, I suggest it may be more helpful to see the *plurality* of Extensions at play, rather than seeing Extension as one ontological way or another. We may better understand the reality of Extension as more of an ontological spectrum along which all activities and programs exist and move. It is clear that the RCT is not the gold standard, since there is no gold standard. The RCT is an inquiry tool which is well-suited for serving some purposes and achieving some ends, just as qualitative case studies are equally appropriate and rigorous inquiry tools if the purposes call for such a tool. It depends on the context and the purpose of the inquiry. In addition, methodological choices cannot be made (well) devoid of critical engagement with the philosophical assumptions about the evaluand or object of inquiry.

Where does this leave us, then, in the quest for credible evidence on Cooperative Extension programs? At the very least, I hope, this paper helps us open up new conversations about what Extension is and about the possibility of an ontological plurality of Extensions. If we build upon that premise, then we can turn—equipped with the other insights presented on credible evidence elsewhere in this volume—towards the generative effort of innovating or adapting methods and approaches that are well-aligned with the "real" core of Extension as is manifest in any particular context and with any specific evaluative end in mind. For instance, as stated by Peters:

Organizing opportunities for people to come together to address public problems and express and pursue their hopes and ideals has been a central part of what Extension has done throughout its first century. As it begins its second century, we should take time to work through different views about how this legacy can best be carried forward.  
(Imagining America, 2014)

In those contexts where Extension is serving the organizing role, an RCT is not likely the best evaluation design. We must explore other, better-suited designs to gather credible evidence of program impacts as we accompany Extension—with its ontology of open, recursive, dynamic, non-linear, values-laden practices, processes, and phenomena—into its next century.

### References

- Akobeng, A. K. (2005). Understanding randomised controlled trials. *Archives of Disease in Childhood*, 90(8), 840–844. doi:10.1136/adc.2004.058222
- Archibald, T. (2015). “They just know”: The epistemological politics of “evidence-based” non-formal education. *Evaluation and Program Planning*, 48, 137–148. doi:10.1016/j.evalprogplan.2014.08.001
- Baron, J. (2018). A brief history of evidence-based policy. *The Annals of the American Academy of Political and Social Science*, 678(1), 40–50. doi:10.1177/0002716218763128
- Biesta, G. J. J. (2010). Why “what works” still won’t work: From evidence-based education to value-based education. *Studies in Philosophy and Education*, 29(5), 491–503. doi:10.1007/s11217-010-9191-x
- Biesta, G. J. J. (2015). On the two cultures of educational research, and how we might move ahead: Reconsidering the ontology, axiology and praxeology of education. *European Educational Research Journal*, 14(1), 11–22. doi:10.1177/1474904114565162
- Box, J. F. (1978). *R. A. Fisher, the life of a scientist*. New York, NY: Wiley and Sons.
- Campbell Collaboration. (2017). *Campbell systematic reviews: Policies and guidelines* (Version 1.3, updated 24 November 2017). Retrieved from <https://campbellcollaboration.org/library/campbell-collaboration-systematic-reviews-policies-and-guidelines.html>
- Campbell, J. R. (1995). *Reclaiming a lost heritage: Land-grant and other higher education initiatives for the twenty-first century*. Ames, IA: Iowa State University Press.
- Chambers, R. (1997). *Whose reality counts?: Putting the first last*. London, UK: Intermediate Technology Publications.
- Coalition for Evidence-based Policy (CEBP). (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, 48(2), 424–455. doi:10.1257/jel.48.2.424
- Donaldson, S. I. (2009). In search of the blueprint for an evidence-based global society. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 2–18). Los Angeles, CA: Sage.
- Donaldson, S. I., Christie, C. A., & Mark, M. M. (Eds.). (2009). *What counts as credible evidence in applied research and evaluation practice?* Los Angeles, CA: Sage.

- Duflo, E., & Kremer, M. (2003). Use of randomization in the evaluation of development effectiveness. Paper prepared for the World Bank Operations Evaluation Department (OED) Conference on Evaluation and Development Effectiveness, Washington, DC. Retrieved from <http://economics.mit.edu/files/765>
- Eddy, E. D., Jr. (1957). *Colleges for our land and time: The land-grant idea in American education*. New York, NY: Harper & Brothers.
- Edmond, J. B. (1978). *The magnificent charter: The origin and role of the Morrill land-grant colleges and universities*. Hicksville, NY: Exposition Press.
- Flyvbjerg, B. (2002). Making social science matter. In G. Papanagnou (Ed.), *Social science and policy challenges: Democracy, values and capacities* (pp. 25–56). Paris, France: UNESCO Publishing.
- Freire, P. (1973). *Education for critical consciousness*. New York, NY: Continuum.
- Hamilton, S. F., Chen, E. K., Pillemer, K., & Meador, R. H. (2013). Research use by Cooperative Extension educators in New York state. *Journal of Extension*, 51(3), Article 3FEA2. Retrieved from <https://www.joe.org/joe/2013june/a2.php>
- Imagining America. (2014). *Extension Reconsidered initiative launches today*. Retrieved from <https://imaginingamerica.org/2014/02/11/extension-reconsidered-initiative-launches-today/>
- Mol, A. (1999). Ontological politics: A word and some questions. In J. Law & J. Hassard (Eds.), *Actor network theory and after* (pp. 74–89). Oxford, UK: Blackwell and the Sociological Review.
- Morrison, K. (2009). *Causation in educational research*. London, UK: Routledge.
- Mosteller, F., & Boruch, R. (Eds.). (2002). *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution.
- Mumford, F. B. (1940). *The land grant college movement*. Columbia, MO: University of Missouri Agricultural Experiment Station.
- National Research Council. (1996). *Colleges of agriculture at the land grant universities: Public service and public policy*. Washington, DC: The National Academies Press.
- Nelson, J., & Campbell, C. (2017). Evidence-informed practice in education: Meanings and applications. *Educational Research*, 59(2), 127–135. doi:10.1080/00131881.2017.1314115
- Nevins, A. (1962). *The state universities and democracy*. Urbana, IL: University of Illinois Press.
- Peters, S. J. (2017). Recovering a forgotten lineage of democratic engagement. In C. Dolgon, T. D. Mitchell, & T. K. Eatman (Eds.), *The Cambridge handbook of service learning and community engagement* (pp. 71–80). Cambridge, NY: Cambridge University Press. doi:10.1017/9781316650011.008
- Peters, S. J., O'Connell, D. J., Alter, T. R., & Jack, A. L. H. (2006). *Catalyzing change: Profiles of Cornell Cooperative Extension educators from Greene, Tompkins, and Erie counties, New York*. Ithaca, NY: Cornell University.

- Pillemer, K. (2011, April 23). *Randomized, controlled designs: The “gold standard” for knowing what works* [Evidence-Based Living blog]. Retrieved from <https://evidencebasedliving.human.cornell.edu/2011/04/23/randomized-controlled-designs-the-gold-standard-for-knowing-what-works-2/>
- Pohl, S., Steiner, P. M., Eisermann, J., Soellner, R., & Cook, T. D. (2009). Unbiased causal inference from an observational study: Results of a within-study comparison. *Educational Evaluation and Policy Analysis, 31*(4), 463–479. doi:10.3102/0162373709343964
- Rasmussen, W. D. (1989). *Taking the university to the people: Seventy-five years of Cooperative Extension*. Ames, IA: Iowa State University Press.
- Report of the Commission on Country Life*. (1911). New York, NY: Sturgis and Walton.
- Ross, E. D. (1942). *Democracy's college: The land-grant movement in the formative stage*. Ames, IA: Iowa State College Press.
- Schön, D. (1983). *The reflective practitioner. How professionals think in action*. New York, NY: Basic Books.
- Scriven, M. (2008). A summative evaluation of RCT methodology: An alternative approach to causal research. *Journal of MultiDisciplinary Evaluation, 5*(9), 11–24.
- Shaffer, T. J. (2017). Supporting the ‘Archstone of Democracy’: Cooperative Extension's experiment with deliberative group discussion. *Journal of Extension, 55*(5), Article 5FEA1. Retrieved from <https://www.joe.org/joe/2017october/a1.php>
- Shapin, S., & Schaffer, S. (1985). *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton, NJ: Princeton University Press.
- Smith, R. G. (1949). *The people's colleges: A history of the New York State Extension Service in Cornell University and the state, 1876–1948*. Ithaca, NY: Cornell University Press. [Republished in 2013].
- U.S. Department of Education. (2003). Notice of proposed priority: Scientifically based evaluation methods (RIN 1890-ZA00). *Federal Register, 68*(213), 62445–62447.
- Walters, P. B., Lareau, A., & Ranis, S. H. (2009). *Education research on trial: Policy reform and the call for scientific rigor*. Milton Park, Oxon: Routledge.
- What Works Clearinghouse [WWC]. (2008). *What works clearinghouse evidence standards for reviewing studies* (Version 1.0). Retrieved from <https://eric.ed.gov/?id=ED511668>
- White, H. (2010). A contribution to current debates in impact evaluation. *Evaluation, 16*(2), 153–164. doi:10.1177/1356389010361562
- Wilson, M. L. (1939). *Democracy has roots*. New York, NY: Carrick & Evans.

*Thomas Archibald* is an Associate Professor and Extension Specialist in the Department of Agricultural, Leadership, and Community Education at Virginia Polytechnic Institute and State University (Virginia Tech).



## Situational Complexity and the Perception of Credible Evidence

**Scott Chazdon**

**Samantha Grant**

*University of Minnesota*

*Situational complexity is the distinction between simple, technically complicated, socially complicated, and complex situations. Programs that operate in simple situations are usually able to follow a prescribed course of action, or recipe, while programs operating in more complicated or complex situations must be flexible and responsive. In this article, the authors present findings from an exploratory, multiple-case study of the credibility of evidence in four distinct program situations ranging from simple to complex. Key informant interviews were conducted with 16 key informants, both internal and external to Extension. The findings were generally that the more complex the situation, the more likely that flexible or mixed-methods approaches were employed to strengthen program credibility. Across all the cases, the relationships that Extension educators have built with stakeholders played a pivotal role in building credibility of evidence. We conclude that sometimes situational complexity matters, sometimes methods matter, sometimes reporting style matters, but what always matters is the trusting relationship between the delivery organization and the stakeholder.*

*Keywords:* complexity, credible evidence, evaluation, Extension education, program stakeholders, qualitative research, comparative case study

***"A situation can be described as more or less simple, complicated, or complex. Utility resides in examining the implications and insights generated by asking to what extent a situation is usefully approached as simple, complicated, or complex, or some combination of the three."***

**—Patton (2011, p. 85)**

### Introduction

Many of the articles in the *Credible and Actionable Evidence* text (Donaldson, Christie, & Mark, 2015) challenge us to think about the relationship between evaluation methodological choices and credibility. While this relationship is important, it is equally important to think about the relationship between the *situational complexity* of our programs and credibility. Researchers and evaluators have been attuned to the importance of situational and program context for decades

---

Direct correspondence to Scott Chazdon at [schazdon@umn.edu](mailto:schazdon@umn.edu)

(Rog, Fitzpatrick, & Conner, 2012; Stake, 1990). Our use of the term *situational complexity* refers specifically to the distinction between simple, technically complicated, socially complicated, and complex situations—a distinction attributed to the work of organizational theorists Ralph Stacey (1996) and David Snowden (2002) and applied more recently to program evaluation by Zimmerman (2001) and Patton (2011).

The situational complexity framework adapted from Stacey's work (Zimmerman, 2011), known as the Agreement & Certainty Matrix (see Figure 1), has two poles based on the level of certainty about cause and effect to solve a problem and the level of agreement among stakeholders about the desirability of the solution. To use a recipe metaphor (Patton, 2011), some Extension programs can actually provide a recipe—a relatively straightforward solution to known problems. These programs typically operate in simple situations where there is a high degree of certainty about cause and effect of a problem and a high level of agreement among stakeholders about the solution. Other Extension programs offer expertise-based frameworks or principles for action in response to more complicated or complex situations where there is more uncertainty about cause and effect of the problem, more social conflict or disagreement about solutions, or both. In these situations, following a recipe does not yield good results.

If program delivery needs to be attuned to situations of varying complexity, becoming less recipe-like as complexity increases, we wondered if evaluation evidence and perceptions of credibility of this evidence would also follow this pattern, becoming less recipe-like as complexity increases. If, as Greene (2015) noted, “method is always the servant of substance, never the master” (p. 206), surely the situational complexity of the program is part of the substance that we cannot ignore. Situational complexity is likely an important influence on credible evidence, so we sought to learn more about how situational complexity mattered.

In this article, we present an overview of situational complexity, and more broadly, program context, and present findings from four case studies examining the credibility of evidence in distinct program situations ranging from simple to complex. We used a comparative case study approach (Ragin & Amoroso, 2011) to learn more about patterns of similarity and difference in the perception of credible evidence. While this was an exploratory type of case study rather than an explanatory, hypothesis testing approach (Yin, 1993), we were curious to see if a credibility pattern would emerge that is similar to the pattern of program delivery, with more recipe-like approaches to evaluation being perceived as more credible for programs operating in simpler situations, while less recipe-like, perhaps mixed-method or participatory approaches would be perceived as more credible for more complicated or complex situations. Understanding the importance of situational complexity for evaluation has implications for how we design and develop future evaluative strategies.

### Situational Complexity and Evaluation Design

Before moving to a discussion of situational complexity, it is important to note that evaluation scholars have written about the importance of the related concept of program context more broadly. In fact, an entire issue of *New Directions for Evaluation* focuses on context and its influence on evaluation practice (Rog et al., 2012).

Noting that context matters deeply in evaluation, Rog (2012) identified five areas of context that affect evaluation practice: 1) the context of the problem or phenomenon, 2) the context of the intervention, 3) the broader environment of the intervention, 4) the context of the evaluation itself, and 5) the broader decision-making context (Julnes & Rog, 2015; Rog, 2012). Conner, Fitzpatrick, and Rog (2012) outlined a context assessment process with guiding questions about each of these areas of context that can be used for program planning and evaluation implementation. To clarify what the authors mean by areas of context, some examples of questions used for this context assessment process are shown in Table 1 below.

**Table 1. Examples of Context Assessment Questions**

Area of Context	Sample Questions for Context Assessment During Evaluation Planning
<b>General phenomenon/problem</b>	What is the problem the program is addressing? What groups prompted concern about it?
<b>Particular intervention</b>	Where is the program in its life cycle? Who does the program serve?
<b>Broad environment around the intervention</b>	Are there political and social views that affect perspectives on the program, its clients, or decision makers?
<b>Parameters of the evaluation</b>	What resources are available to support the evaluation?
<b>Broad decision-making arena</b>	What are the main decision makers/users of the evaluation? What are the political expectations for evaluation?

*Note:* From Conner et al. (2012), p. 96.

The idea of situational complexity seems closest to “broad environment around the evaluation,” and the idea that the views of program stakeholders surrounding an intervention are highly relevant to the credibility of the evaluation.

Our interest in situational complexity arose from our experience as internal and often developmental evaluators at University of Minnesota Extension. In our roles, we are confronted not only with variation in these five areas of context, but also with the fact that many of the situations in which Extension programs are either newly emerging or constantly changing, often unpredictable, and with a constant need to adapt programming as well as evaluation to these changing situations.

Based on the ideas of organizational theorist Ralph Stacey (Stacey, 1996), Brenda Zimmerman conceptualized a situational matrix to better understand the elements of complexity (Patton, 2011; Zimmerman, 2001). The Agreement & Certainty Matrix plots two features of a situation:

- 1) The level of certainty about cause and effect that can solve a problem, and
- 2) The level of agreement among stakeholders about the desirability of the solution.

In this matrix, there are four program situation types based on these two poles: simple, technically complicated, socially complicated, and complex. The two main distinctions that differentiate programs are the level of agreement and certainty about how to solve the problem. Certainty refers to the predictability about how to solve the problem, and agreement refers to the amount of conflict about how to solve the problem.

*Figure 1. Agreement & Certainty Matrix*

<b>Agreement</b>	Far from	Socially complicated	Complex
	Close to	Simple	Technically complicated
		Close to	Far from

**Certainty**

*Note:* Adapted from Zimmerman, 2001; found in Patton, 2011.

Simple situations are those that are close to certainty about cause and effect of the intervention to solve the problem and are close to agreement among stakeholders on the proposed solution to the problem. These are the recipe-like situations for which “best practices” can be found and agreed upon. Patton shared that in a simple program, the standard procedures that have worked to produce the desired outcomes in the past are highly likely to work again in the future (Patton, 2011). This is not to say that the program is simplistic; rather, it is more formulaic in design. In Extension, there are a variety of best practice programs in which research and evaluation have demonstrated that outcomes can be replicated with fidelity to the program design.

Technically complicated situations are far from certainty about how an intervention will produce a desired effect but are closer to agreement among stakeholders on the proposed solutions (Patton, 2011). Often in these situations, more than one area of expertise is needed, and therefore, must be coordinated and integrated. The solution to the problem is complicated but

knowable. Scaling programs across different program sites is one example of a technically complicated situation.

Socially complicated situations are far from agreement about solutions to the problem but are close to certainty about cause and effect (Patton, 2011). Situations with many different stakeholders offering differing perspectives, articulating competing values, and posing conflicting solutions are socially complicated. Relying on multiple perspectives means that each party has a different take on the situation. Multiple perspectives are important for both innovation and consensus but negotiating interpersonal or intra-group dialogue can be more challenging than actually solving the problem at hand.

Complex situations are both far from certainty and far from agreement. In these cases, high uncertainty about how to produce a desired result can fuel disagreement, and disagreements may intensify and expand the parameters of the uncertainty (Patton, 2011). These types of situations often call for innovative approaches, because there is not an easy—or known—solution to the problem.

For the sake of this article, we did not include the element of chaos that is often referenced in models of complexity, such as the Cynefin framework (Snowden, 2002). Chaos is when there is extreme uncertainty about how to solve the problem and strong conflict among stakeholders. Programs in a state of chaos are not in the right place to be evaluated because these programs are in a state of uncontrollability and unpredictability.

In this article, we have focused on situational complexity from the vantage point of two evaluators within the Extension program. In *Utilization-Focused Evaluation*, Patton (2008) discussed that the choice of an evaluation design is driven by the people involved and their situation. This involves negotiating the intended use of the evaluation for the intended users. Just as a program situation can be simple, complicated, or complex, an evaluation design can vary in its complexity. That is not to say that simple programs only have simple evaluations and complicated programs only have complicated evaluations.

Patton (2011) noted that situational complexity provides a framework to begin to explore evaluation questions. Patton suggested that simple situations lend themselves to logic models where evaluation is based on proving connections between inputs, activities, and outcomes. Evaluation designs in complicated situations should take into consideration the system, stakeholders, and the context of the program and thus lend themselves to more complex questions that investigate the linkages in the system. Developmental evaluation (Patton, 2011) was envisioned as an evaluation approach for complex situations. Evaluation designs for complex situations require the evaluator to stay attuned to changes in processes and outcomes to create feedback loops that continually inform the program and the stakeholders. This type of evaluation would need to be responsive to the program and the environment and takes system influences into consideration at all times.

From our own experiences as long-term evaluators in Extension, we have witnessed that the situational complexity of the program matters greatly for both program design and evaluation. As evaluators, we often start our projects by trying to better understand the program context before designing an evaluation. In each type of program, credibility of evidence is likely achieved through different approaches and methods. To better understand how approaches and methods changed in different contexts, four unique Extension programs were examined through case study interviews.

### **Research Design**

Our methodological choice for this project was a comparative case study. Ragin and Amoroso (2011) described the goal of comparative research is “to elucidate and explain the diversity within a particular set of cases” (pp. 135–136). In choosing cases for analysis, our analytic frame was situational complexity. Our approach was also interpretive, meaning that the goal was to gain a deeper understanding of the ways that situational context affects the perception of program credibility, not hypothesis-testing usually associated with post-positivist approaches to research. As noted by Greene (1994), interpretive approaches to evaluation and research focus on how a program is experienced by various stakeholders.

We identified four programs offered at the University of Minnesota Extension that exemplified the range of Stacey’s four situation types (Zimmerman, 2001), from simple to complex, and focused our analysis on patterns of similarities and differences across these four cases. We intentionally selected programs from all the different program areas within Extension. It is important to note that these programs were fit into the categories retrospectively based upon the evaluators’ knowledge of the program. Our goal was to find one program that best illustrated each situation type.

Although this was an exploratory type of multiple case study (Yin, 1993), we were curious if we would find a pattern that more recipe-like approaches to evaluation would be perceived as more credible for programs operating in more simple situations, while more flexible, and perhaps mixed-method or participatory approaches would be perceived as more credible for more complicated or complex situations.

To examine perceptions of the credibility of Extension program evaluation evidence, we identified and interviewed key program staff for four Extension programs:

- 1) Private Pesticide Applicator Workshops;
- 2) Children, Youth, and Families at Risk (CYFAR);
- 3) Parents Forever; and
- 4) McLeod for Tomorrow.

During our interviews with Extension staff, we asked for names and contact information for one or two key informant external stakeholders, people who Patton (2008) referred to as primary intended users, who would be in a position to reflect on the credibility of evaluation evidence. In some cases, these key informants were programmatic partners, and in some cases, they were funders. Key informants are people who have special insight or expertise about a topic (Patton, 2015, p. 284). The use of key informants in qualitative research arose in ethnographic studies and has continued to be useful in program evaluation when “there is a need to understand motivation, behavior, and perspectives of our customers and partners” (U.S. Agency for International Development, 1996, p. 1). In total, nine key informants internal to Extension and seven key informants external to Extension were interviewed.

Hour-long semi-structured interviews were conducted via phone or Google Hangout. Primary program staff and stakeholders were given interview questions before the interview. This helped the interviewees feel more comfortable and prepared for the discussion.

The selected case study programs are described in more detail in Table 2 and the narrative below. We also collected and analyzed secondary sources of data—program descriptions and program evaluation reports—from each program. Our interview questions for Extension staff and external stakeholders can be found in the Appendix.

**Table 2. Selected Case Study Programs and Situation Framing**

<b>Program</b>	<b>What Program Does</b>	<b>What Problem Is Program Addressing?</b>	<b>Brief Description of Situation (Focusing on Certainty/Agreement)</b>	<b>Situational Complexity</b>
<b>Private Pesticide Applicator Workshops</b>	Trains farmers to safely apply pesticides	Pesticides can cause damage to the person applying them and to the environment.	Close to agreement about the need to address the issue and close to certainty that success will come from educating a target audience.	Simple
<b>CYFAR</b>	STEM after-school education program for at-risk youth	Builds a sense of community and science skills in youth.	Close to agreement that youth programs matter but far from certainty about how to fit 4-H programs for the target audience.	Technically Complicated
<b>Parents Forever</b>	Parent education for divorcing or separating couples	Promotes resiliency for families transitioning through separation, divorce, and/or custody change.	Close to certainty that this type of intervention will produce desired effects but far from agreement about the need for this type of programming.	Socially Complicated

<b>Program</b>	<b>What Program Does</b>	<b>What Problem Is Program Addressing?</b>	<b>Brief Description of Situation (Focusing on Certainty/Agreement)</b>	<b>Situational Complexity</b>
<b>McLeod for Tomorrow</b>	Community leadership education	Promotes stronger connections among small towns in a county where competition among communities is the norm.	Far from agreement that this type of program is needed and far from certainty that this type of program will produce the desired effect.	Complex

### **Private Pesticide Applicator Workshops**

A simple program situation is one that has a high level of technical certainty about how to solve a problem as well as a high amount of social agreement about the program approach. Pesticide Applicator Workshops teach both private landowners (e.g., farmers) as well as professionals (e.g., commercial farmers, turf and landscape businesses) about proper pesticide application. For this article and interview, we focused on the training of private audiences. Minnesota has over 17,000 certified private pesticide applicators who require certification to be able to apply restricted use pesticides to their commodity cropland. Certification must happen every three years, and to keep certification, applicators must either attend a recertification training or complete and pass an exam. Approximately 41% of applicators needing recertification attend an Extension workshop annually. An exam needs to be passed if the applicator does not attend an Extension workshop.

The Private Pesticide Applicator Workshops exemplify a simple program as it serves well to manage a problem (incremental innovation of appropriate pesticide application techniques) through educating a specific audience—landowners. The problem addressed by this program is that pesticides can cause damage to the person applying them and to the environment if applied incorrectly. Problems exist with improper use of pesticides, including issues with health and safety, environmental protection, and agricultural pest management.

The Private Pesticide Applicator Workshop training is a long-running, mature Extension program. The benefit of such a mature program is that the program team has established a formulaic model for delivery and testing of the effectiveness. Extension staff deliver training on proper techniques to apply pesticides and work to ensure that research best practices are shared with the field to promote safer pesticide application. Extension staff members create day-long trainings and deliver them around the state. To create the trainings, Extension staff work with a variety of partners to identify the most pressing needs in pesticide application. They then develop content that is informative and engaging for their audiences and meets the requirements for certification.



Although we chose the Private Pesticide Applicator Workshops as a simple program, it could easily be classified as a socially complicated program as the proper mixing and application of pesticides is a technically (and sometimes socially) complicated challenge. The well-established design of creating adult-based pesticide training for a specific target audience was the reason for its selection as a simple program.

### **Children, Youth, and Families at Risk (CYFAR)**

A technically complicated program situation is one in which there is a high level of social agreement about how to solve the problem but a low amount of technical certainty as to the most effective methods. Technically complicated situations require more than one area of expertise that must be coordinated and integrated. Large-scale programs that have multiple local sites are often good examples, as what fits in one context might not fit in another.

The Children, Youth, and Families at Risk program (CYFAR) was our example of a technically complicated program. The CYFAR program is a national grant-funded program to support children, youth, and families at risk. University of Minnesota Extension has held CYFAR grant funding for this project for five years and developed a program model that had multiple years of testing and refinement. Program staff and funders believe that quality positive youth development programs are an important vehicle to get youth excited about science and more interested in science-related fields. CYFAR funding allowed program sites to design the best ways to deliver such programming, as investors were aware that the local context played a pivotal role in the way in which youth were engaged in the program, and they gave programs latitude to make program decisions that would promote positive youth development.

The CYFAR program was designed to ignite Somali middle school-aged youths' interests in learning about Science, Technology, Engineering, and Mathematics (STEM) while preparing them for higher education. Three clubs met in three different locations in Minneapolis, St. Paul, and Eden Prairie. Club participants were youth who experienced educational barriers such as lack of access to resources, high rates of truancy, and disengagement from school. During the school year, the clubs met weekly and were facilitated by two adults. Each club designed its own approach to building science skills based on the needs of the community, youth interest, and staff leader skills. Youth applied their learning to solve practical and scientific engineering problems. Youth also engaged in activities that built leadership skills needed to pursue higher education and careers. Each summer, youth participated in a University of Minnesota campus immersion where they learned about student life, explored academic interests, and identified steps toward college readiness.

### **Parents Forever**

A socially complicated situation involves many different stakeholders offering differing perspectives, articulating competing values, and posing conflicting solutions to a problem. We

chose a divorce education program as an example of a program operating in a socially complicated space. Although there may be technical agreement that providing parenting education to divorcing or separating couples helps them keep their conflicts away from their children, there is not widespread social agreement that this type of program is needed, or about the best approach for providing this type of intervention. Some critics of divorce education argue that the focal point of education efforts should be before marriage occurs in the first place. Other critics of divorce argue that the program should be required of all couples with children, while others think it should remain voluntary. Practicality has played a role in reducing the “dosage” of this program so that it is provided in an online format, and some critics argue that this does not provide as high quality of an intervention as the traditional face-to-face course. Some argue that programs of this type should require a scientific evidence base. There is also a lack of clarity as to whether this approach works for non-white, high conflict, or same-sex couples.

Minnesota Extension’s response to divorce education is the Parents Forever Program, which began in 1994. In 1998, the Minnesota legislature passed a requirement that in all contested custody or parenting time proceedings, parents of a minor child must attend a minimum of eight hours in an orientation and education program. Parents Forever is based on research that suggests factors such as conflict or financial stress can increase the risk of families going through transitions of separation, divorce, and/or changes in custody. The program is designed to encourage parents to pay attention to their own well-being, attend to their children’s development and the parent-child relationship, and improve the co-parenting relationship. Using these three primary mechanisms (well-being, parenting, co-parenting), the curriculum is aimed at increasing resiliency for families transitioning through separation, divorce, and/or custody change.

Parents Forever is highly regarded, often recommended, in counties across the state. Originally, Parents Forever was only a face-to-face program. More recently, Extension developed an online program offering as well. The online offering is popular because it removes transportation and child care barriers, making it easier for parents to attend. As noted by program staff, however, the online offering does not provide the social connections with other parents that have always been reported to be a valuable component of the program.

### **McLeod for Tomorrow**

Complex situations are those in which high uncertainty about how to produce a desired result fuels disagreement, and disagreements intensify and expand the parameters of the uncertainty. Communities in rural counties often find themselves competing with each other for scarce resources, often with a sense that some communities (often the county seat) get all the resources while other communities struggle. At the same time, rural residents often work or shop in communities other than their own, and problems such as workforce issues, economic

development challenges, natural resource concerns, or natural disasters do not begin or end at city limits. We examined McLeod for Tomorrow as an example of a program designed for this type of complex situation.

McLeod for Tomorrow is an example of “county bridging program” (Rasmussen, Armstrong, & Chazdon, 2011). These nine-month cohort programs are designed to strengthen county-wide community by creating “bridging” relationships of communication and understanding. This is often a challenge in counties where local heritage and pride play an important role, where deep-seated rivalries between communities exist, and where long-standing insider groups have controlled the decisions made. Therefore, these programs are explicitly designed to engage new or young residents of diverse backgrounds from the communities across the county.

In McLeod for Tomorrow, a variety of activities are embedded in the program’s design to build trust, mutual respect, commitment, and political awareness among the program participants and communities. To sustain the program, McLeod for Tomorrow became a non-profit 501(c)(3) organization several years ago, with approximately one-third of its budget coming from the county and the remainder coming from donations and grants. A paid, part-time coordinator manages most of these activities, and Extension’s role is to provide the leadership education component.

While it may sound like McLeod for Tomorrow has a recipe that is followed consistently from one year to the next, the mixture of personalities, ages, and cultural backgrounds of program participants varies each year, and the social and economic context is constantly changing. Furthermore, the program is funded in part by the county and must compete for resources with other programs that are less ambiguous in purpose.

### **Stakeholder Interviews on Credible Evidence**

Interviews with Extension staff and external stakeholders focused on the types of evaluation evidence collected, both formative efforts to improve program quality and summative efforts to measure results (Scriven, 1967), and the perceptions held by external stakeholders about the credibility of this evidence. In this section, we highlight key learning from each case about evaluation design and perception of credibility.

#### **Private Pesticide Applicator Workshops**

The evaluation approach at this point in the program’s maturity focused primarily on knowledge gains for participants. In this case, the program used a simple evaluation approach anchored around post-session evaluation as well as checks for understanding using Turning Point technology polling during the sessions. Feedback from these evaluations are viewed by the trainers and helped to inform future training efforts. Educators intentionally chose an evaluation focus on educational gains versus public impact (such as environmental impact on water quality),

as they believe that the program has control over the education but not adoption/compliance with proper pesticide use. Moving into an evaluation design that was focused on environmental impact would require control in the design of the study to understand the connection between training from Extension and changes in practice that would ultimately result in environmental shifts.

In addition to the teaching evaluation, a summative impact evaluation study was conducted annually in 2011, 2012, and 2013 of program participants from that year to learn if the trainings were meeting program goals and if any changes should be made to better meet the needs of the audience. Evaluations were mailed to all 1,000 randomly selected participants who had completed the training, and 44% responded.

Results demonstrated that:

- Nearly 73% of workshop participants made at least one pest management decision based on what they learned at the workshop, and 45% made two or more such decisions;
- Farmers found the workshop modules on Personal Protective Equipment to be the most useful in making pest management decisions; and
- 91% of participants planned on attending a workshop again in the future.

Credibility from the side of the program staff was evidenced by positive evaluation data that were fed back into the program to continue to make changes. The program team meets to review all data, and some of the teaching data were also shared with program partners to justify the continued need for the program. The three-year evaluation gave critical feedback to help change content offered. Enrollment for the courses remains high, with an increase of people opting to take the course rather than the exam. In addition, investors continue to give money to the program and are happy with the partnership with Extension.

**Perceptions of credibility.** Primary stakeholders for this program were investors of the program, including the Minnesota Department of Agriculture, U.S. Environmental Protection Agency, and Extension leaders. Extension staff noted that external stakeholders continue to come back to Extension for this programming, which is strong evidence of the program's credibility as well as trust in the program. They noted that Extension collects summative participant feedback about the program and continues to use formative feedback to improve teaching. One staff member noted that external stakeholders are mostly attuned to "credibility more in how the training is offered or how certain groups are represented."

The external stakeholder interviewed represented a state agency partner who worked closely with University of Minnesota Extension. She expected the Private Pesticide Applicator Workshop training to show results in the form of the summative program outcomes—properly training individuals to adequately apply pesticides without harming themselves or the environment

according to the Minnesota Agriculture Department guidelines. Evaluation approaches, according to her, should also measure the applicability of material and should allow program staff to get feedback about ways to improve the workshop in the future.

This stakeholder was satisfied with the evidence from the program. She trusted the process of collecting the data as well as the quality of data collected. When asked what would cause her to lose confidence in the program, she stated that she would lose confidence if one of three things occurred:

- 1) If there was no evidence that the participants are getting something out of it (no learning occurring),
- 2) If Extension used outdated data to provide recommendations or to adjust the program settings, and
- 3) If there was a mismatch between what the data provides and what was observed in the field.

Interestingly, when asked to think beyond results, the stakeholder noted the important role that Extension staff play in being experts in their field. She stated that she wanted “to see that the University of Minnesota Extension program staff have ample opportunities for professional development and access to the latest research and technologies to stay current with industry trends and maintain credibility from their audience.” This was an example of a stakeholder valuing the people in the organization as much as they value the evidence collected for evaluation purposes.

## **CYFAR**

During the final year of the CYFAR grant, the program team wanted to learn more about the impact of the CYFAR model. Much work had been put into developing a model that could be adapted across program sites, and over the tenure of the grant, the model had been practiced in urban as well as rural environments.

Summative evaluation methods included youth pre- and post-surveys utilizing the common measures provided by the CYFERnet Evaluation Team at The University of Arizona and Virginia Tech. The common measures used in this project measured the impact of 4-H CYFAR program in science and positive youth development. The staff also utilized a formative assessment strategy to document the youth development practices that each site was using each time they engaged with youth. In addition, a mid-year focus group and a final showcase of learning allowed youth to articulate their goals and challenges.

Evidence collected supported the intended program outcomes. Youth liked the program, retention was high, and there was a positive impact on youth’s STEM capabilities. In addition, youth showed increased interest in science careers and STEM abilities. A major thread in the

evaluation was tracking attendance data for youth. In past years of the grant, attendance data were collected by program leaders but not reported back to the principal investigators or evaluators. The team hypothesized that youth with greater participation in the program would see greater science outcome gains. The evaluation did not support this hypothesis, as there were no significant outcome differences for high- and low-participation youth. The majority of youth had high outcome attainment.

The project team remarked that their adherence to mainly quantitative data collection in the form of survey or collection of attendance data did not capture all of the rich stories they heard from program staff. The team decided to conduct both a mid-term focus group and final showcase of learning to better illustrate the benefits of youth programs.

Results have been utilized by program staff in program reports. The team also created videos that have been shared across the 4-H system to document the impact of this program model. Evaluation successes for this group were attributed to the importance of building relationships with program partners as well as with youth. All partners were invested in collecting high-quality data across multiple program sites.

**Perceptions of credibility.** We interviewed two external stakeholders for the CYFAR program. One was a staff member from a partner non-profit organization that works side-by-side with University of Minnesota Extension to plan, deliver, and evaluate the CYFAR program. In addition, we interviewed a local Minnesota funder of the CYFAR program.

Program staff and partners expressed interest in gathering summative data showing the value of the CYFAR program to scale the program up to other sites. There also was a strong value placed on understanding formative data to inform improvements in the program design that would ultimately increase the impact for youth. This partner stated the evaluation data had been used in meaningful ways for program improvement by stating:

I use the information provided by Extension to improve the program in ways that are in accord with the wants and needs of participants and their parents. We try to shape the program with the vision of participants, and Extension's evaluations are helping us achieve that goal.

The funding stakeholder had great interest in the CYFAR program because it serves a hard to reach an audience of Somali youth. He stated, "We want to help the University and Extension reach communities that they would not have reached otherwise."

When asked about the credibility of the evidence, both stakeholders vouched for its credibility. One stakeholder knew this program had a steady stream of supporting data that had been shared about the program but could not remember specific evaluation findings. This is another example of a time where the trust in both the organization and the reputation of strong work with the right

audience influenced the credibility of the evaluation. The other stakeholder shared that the participatory approach to evaluation added to the credibility of the findings. He shared:

The source of the data, program participants and their parents, adds to the credibility of the evaluation evidence provided by Extension.

When asked under what circumstances they would lose confidence in the program, one respondent stated:

If it becomes a one-way street. It means that the university adopts a top-down approach and imposes [on] us a way of implementing the program. Also, if the structure of the organization changes, shifting from being flexible to being a “one size fits all” type of organization.

Another respondent stated that he would lose confidence if there were an unexplained decrease in the effectiveness of the program or a decrease in the quality of the services.

### **Parents Forever**

The Parents Forever program collects evaluation evidence with participants after each course, whether in-person or online. The evaluation survey includes a series of 14 Likert-scale questions about participant outcomes in the three main focal areas of the curriculum: parent well-being, parent-child relationships, and co-parenting relationships. These items are intended to collect data for both formative and summative purposes. In addition, participants are asked about their likelihood of following through on five specific action steps:

- 1) Adjusting my parenting to better meet the needs of my child(ren),
- 2) Taking steps to improve my support network,
- 3) Identifying my goals,
- 4) Using one of the co-parenting strategies I learned about, and
- 5) Using one of the financial tools I learned about.

Short reports using data visualizations are then prepared annually for each of the eleven sites that provide Parents Forever training. In addition, quarterly evaluation surveys are conducted with facilitators of the program.

In 2005, program leaders published an article in the *Journal of Extension* based on an impact study of the program (Dworkin & Karahan, 2005). The article cited the program’s success in meeting its objectives. Since then, Extension research staff have conducted numerous studies with the goal of testing and documenting the effectiveness of the Parents Forever curriculum in supporting parents through separation, divorce, and/or custody change (Becher et al., 2015; Becher et al., 2018; Cronin, Becher, McCann, McGuire, & Powell, 2017). Program staff are

working on submitting evaluation evidence to various registries to enable the program to received certification as “evidence-based.”

**Perceptions of credibility.** We interviewed two Extension staff members and two program stakeholders about their views on credible evidence for the Parents Forever program. One stakeholder leads the Community Education department in a rural Minnesota community and provides referrals to divorcing or separating parents who are required by a court order to complete divorce education. The second stakeholder was a family law attorney who helped provided background on legal issues for the development of the curriculum.

The staff members we interviewed noted that facilitators of the program are social workers and are fiercely devoted to this program. They are on the front lines with the parents. One staff member stated, “When you are in a class with parents for eight hours, it is amazing to see parents come in with their arms crossed, disgruntled because they have to be there, but then they leave telling us thank you and ‘I wish I would have taken this class before we decided to divorce.’” The facilitators also are not familiar with evaluation or questions of methodology. Instead, they bring evaluation into their program by continually making program changes based on conversations with participants, intuition, and feedback from other facilitators.

When asked about the type of evaluation evidence she would like to see, one external stakeholder mentioned the rate of people taking a parent education class (versus not taking such a course) in going back to court in the future because they are unable to come to an agreement. This evidence could focus both on dollar savings for the court system as well as reduced conflict. The stakeholder acknowledged, however, that this type of study would be costly and difficult to conduct.

The other external interviewee commented on the evaluation reports that she receives annually. She noted that the reports are difficult to interpret, sometimes using a numbering system for the 14 program outcomes that is hard to follow. She requested that the reports be made more user-friendly.

However, when asked if they believed the evaluation evidence was credible, both interviewees agreed it was. One respondent noted, “I don’t know what else you could ask! I love land grant universities, and I totally trust everything that comes from there.” The other stated, “I have no reason to believe it isn’t credible.” This same respondent went on to state she would only lose confidence in the program

if the program started ignoring experts in the field, whether they be child development people, other legal experts. If the information presented wasn’t based in scholarly or professional knowledge, or if they kept getting evaluations back from parents that the program isn’t helpful and they ignored that. But I don’t think they would do that.



Due to the relationships that have been formed between stakeholders and the Parents Forever program, a high level of negative evidence would be needed to discredit the program.

### **McLeod for Tomorrow**

Evaluation for the McLeod for Tomorrow leadership program involves pre- and post-surveys conducted during the first and last session of the nine-month cohort program. The pre-survey collects data on the organizational involvement of program participants as well as data on their level of civic involvement. These same questions are repeated in the post-program survey. Analysis of the pre-post data shows behavior change as measured by the percentage of participants who actually increase their level of involvement in organizations as well as the percentage who increase their engagement in civic activity. The post-program survey also includes a retrospective pre-post set of questions about the achievement of leadership competencies in the areas of civic engagement, self-efficacy, self-awareness, cross-community knowledge, and shared vision for the future. These competency data are used for formative as well as summative purposes.

In recent years, evaluation staff have begun to collect community impact data using Ripple Effects Mapping and follow-up surveys with program alumni. Ripple Effects Mapping is a participatory group process that engages program participants as well as other community stakeholders, in paired interviews and large group dialogue about the chain of effects produced by a program (see Chazdon, Emery, Hansen, Higgins, & Sero, 2017; Hansen Kollock, Flage, Chazdon, Paine, & Higgins, 2012). The Ripple Mapping session identified several county-wide efforts that had been created by alumni of the McLeod for Tomorrow program. This awareness then led program staff to design a further evaluative study to quantify, in dollar terms, the economic contribution of the program for the county. This study found, overall, that the program returned a value of \$6.40 for every dollar invested (Tuck, Chazdon, Rasmussen, & Bohn, 2019).

**Perceptions of credibility.** We interviewed two Extension staff members and two external stakeholders. One of the external stakeholders is the coordinator of the non-profit organization that runs the McLeod for Tomorrow program. The other external stakeholder is a county government administrator.

The Extension Leadership educator noted that when she has presented information about the program to the county board, she does not have to work hard to justify the program. They say, “We love this program, and we know it works.” In fact, one of McLeod’s county commissioners went on to participate in the Minnesota Agricultural and Rural Leaders program, another well-known Extension offering. County Commissioners attend McLeod for Tomorrow workshops when they are invited as well as continually provide about one-third of McLeod for Tomorrow’s funding.

The external stakeholder emphasized that Extension's evaluation efforts have helped her communicate with donors and sponsors about the importance of the program. She did note that materials from Extension need to be "easy to consume," meaning that sometimes the tables and information are too complicated, and thus, it is important to keep reports brief and only present the most important information to stakeholders. She also expressed some concern that program participants may rush through the completion of evaluation surveys, which tends to reduce the credibility of the evidence collected.

When asked under what circumstances she might lose confidence in the leadership program, the interviewee spoke specifically about the Extension educator who has always delivered the leadership content, noting that "if she retired, I'd have to build a new confidence in her replacement."

The county administrator noted, "I also tend to trust information that is provided or compiled by reliable resources such as the University of Minnesota Extension office; organizations that have ethics and good practice in place typically care about the type of information presented by those representing their organization." In reaction to the recent economic contribution study report, she noted that "people like to read about people, most of the figures given are numbers—although great, it won't stick in people's minds. Faces and stories stick in people's minds and cause them to take action."

Table 3 summarizes our findings on the credibility of evidence in these four program contexts.

**Table 3. Situation Framing and Insights About Credibility**

<b>Program</b>	<b>Situation Framing</b>	<b>Evaluation Design</b>	<b>Insights About Credibility of Evidence</b>
<b>Pesticide Applicator</b>	Simple	End of training participant survey; Follow up survey	Evidence of learning gains is necessary and credible; Properly trained Extension staff are the most important source of credibility.
<b>CYFAR</b>	Technically Complicated	Youth pre- and post-survey; Program attendance data; Program lesson plans	Source of data (both parents and youth) adds to credibility; credibility flows from responsiveness of program, not a "pre-canned" approach
<b>Parents Forever</b>	Socially Complicated	End of training participant survey; Facilitator survey; Quasi-experimental study	Keep it simple; Maintain trust; if possible, show monetary value (reduced court expenditures)
<b>McLeod for Tomorrow</b>	Complex	Pre and post surveys, alumni follow-up; Ripple Effects Mapping; Return on Investment study	Keep it simple; Protect integrity of data collection; Maintain trust; If possible, show monetary value

### **Discussion and Implications for Extension Evaluation**

This multiple case study project was exploratory. We were interested in patterns of similarity and difference based on situational complexity. We were curious to see if more recipe-like approaches to evaluation were perceived as more credible for programs operating in simpler situations. Similarly, we wanted to see if less recipe-like, perhaps mixed-method or participatory approaches, would be perceived as more credible for more complicated or complex situations.

Our findings were generally that the more complex the situation, the more likely that flexible or mixed-methods approaches were employed to strengthen program credibility. The Private Pesticide Applicator Workshops and Parents Forever programs were able to rely upon simple survey findings to measure learning gains, and this evaluation design was perceived as credible. The other programs also used survey methods but combined these methods with participatory and qualitative methods to better document the richness of the program experience. In the case of the CYFAR program, staff used focus groups, video stories, and a final showcase of learning to strengthen the credibility of evidence. Evaluation of McLeod for Tomorrow has involved both Ripple Effects Mapping and a combined qualitative-quantitative analysis of the economic contributions of the program to the county to strengthen the program's credibility.

While less recipe-like evaluation approaches were used in the more complicated and complex programs in our case study, it was particularly interesting to note that external stakeholders for these more complex programs emphasized that evaluation reporting should "keep it simple."

One of the external stakeholders interviewed for the McLeod for Tomorrow program felt that compelling stories were just as valuable as the quantitative evidence she was presented. This finding from our exploratory study is worth further research. Is there a relationship between "simplicity" of the situational context and credibility of quantitative forms of evidence? Or stated conversely, do stakeholders in complex situational contexts trust narrative more than numbers because they know the numbers cannot tell the whole story?

A limitation of our case study design was that while we set out to select programs that exemplified the four quadrants of the Agreement & Certainty Matrix, we learned through the interviews that programs did not fit these boxes very well. Situations can trend toward "simple," but they are never simple all the time. For example, stakeholders in the Private Pesticide Applicator Workshops could, over time, become more skeptical about pesticide use in general. Patton (2011) noted that simple, complicated, and complex situations are not always easy to distinguish: "There is no complexity thermometer that gives degrees of uncertainty and disagreement on a standardized, all-purpose scale" (p. 95). Programming situations that are close to agreement or close to certainty are increasingly rare. For this reason, it is best to assume that a range of methods and participatory evaluation strategies are worth pursuing to increase credibility.

In addition, we interviewed key informant stakeholders who already had a relationship with Extension programs. This likely biased some of their responses about the credibility of evidence, because these people were largely friends of Extension. However, we believe this limitation did not compromise our findings of this study because the reality of delivering educational programs is that they are driven by stakeholder investment and the stakeholders chosen for the interviews were the primary intended users of the evaluation findings (Bryson, Patton, & Bowman, 2011; Patton, 2008).

While our study showed a pattern of differences based on situational complexity, we also saw an important commonality among the four cases. The cases consistently highlighted the importance of stakeholder trust and Extension's credibility as a delivery organization. Years ago, Weiss and Bucuvalas (1980) found that "trustworthiness" of a research study was a crucial component influencing decisionmakers' likely use of the study findings. But the type of stakeholder trust we heard about was different than the "trustworthiness" of the evaluation research itself. It was more about the trustworthiness of the delivery organization. As noted by Greene (2015), "the credibility of evaluative evidence is not automatically granted via the use of particular empirical methodologies but rather is earned through inclusive, relational, and dialogic processes of interpretation and action that happen on the ground, in context, and in interaction with stakeholders" (p. 206).

In earlier work on the relationship of public value to evaluation, Chazdon and Paine (2014) found that the credibility of the delivery organization, defined as "stakeholder perceptions of the quality of the public program, as well as the reputation of the delivery organization," was a key component in the public value of a program (p. 108). As noted by Chazdon and Paine (2014):

The 4-H program has a long history and is often revered in rural communities. It is likely that some of the public value resulting from this program is derived from this reputation, but this reputation must be carefully safeguarded and cannot always be taken for granted. Moreover, public universities often benefit from a perception that they offer unbiased analyses and reports that hold up to public scrutiny better than analyses or reports produced by for-profit companies. Yet, even prestigious research universities may lose this reputation as a result of a breach of integrity. (p. 109).

Our case study interviews supported this notion. CYFAR had higher sights in their evaluation to demonstrate the value of their program model. The team thought about the different contexts for youth and different stakeholders as well as showing impacts. Staying attuned to stakeholders' needs is a big part of what keeps Extension credible with stakeholders. For both Parents Forever and McLeod for Tomorrow, staying attuned resulted in using data visualization and brief report formats that were easy for external stakeholders to understand. An implication for evaluators is

that maintenance of organizational credibility is a shared responsibility between evaluators and program staff.

The relationships that Extension educators have built with stakeholders play a pivotal role in building credibility of evidence. Mathison (2015) noted that “information becomes evidence through lived experiences, including professional practice. . . . The more context provided for evidence, the better able we are to judge its credibility” (p. 158). Miller (2015) discussed the way that people judge credible information and finds that trust in an individual is a crucial component. In Extension, these trusting relationships proved to be an important part of how credible evidence was assessed by stakeholders.

For that reason, evaluations in Extension should strive, if possible, to use participatory approaches in which both evaluators and program staff are engaged in the design of the evaluation. Our Extension staff have rich relationships built through ongoing work with stakeholders. They understand the needs and wants of their stakeholders, and in our case studies, they were attuned to the evaluation that would help support their program.

Program support and growth, from our interviews with Extension staff, came both from formative evaluation to improve the delivery of the program as well as summative evaluation that positioned the Extension program favorably with stakeholders. Evaluators should use the expertise of program staff to help craft rigorous evaluations that are attuned to both the improvement of the program and stakeholder needs for impact data.

Looking across our four case studies, a common theme was that credibility is as much, or more, about programs and personnel than it is about evaluation rigor. Evaluation is vital, but it is vital because it protects the integrity of the program delivery organization. Regardless of the program, the external stakeholders we interviewed for this study wanted to generally know that Extension programs were evaluating, and they wanted to know that we know our audiences well, but they did not necessarily want to see the evaluations! When they did want to see them, they wanted them to be short and easy to interpret.

We began this study with a concern that too much emphasis has been focused on methodology in discussions of the credibility of evaluation evidence. We thought situational context was also an important influence on the credibility of evidence. Through our case studies, we have learned that sometimes situational complexity matters, sometimes methods matter, sometimes reporting style matters, but what always matters is the trusting relationship between the delivery organization and the stakeholder. As concluded by Greene (2015), “well beyond good method, making meaningful and consequential judgments about the quality and effectiveness of social and educational programs requires engagement, interaction, listening, and caring” (p. 219).

So it seems that the more relevant question for Extension is “what makes the program credible?” rather than “what makes the program *evaluation* credible?” Evaluation is part of what makes a program credible, but it does not stand on its own.

### References

- Becher, E. H., Cronin, S., McCann, E., Olson, K. A., Powell, S., & Marczak, M. S. (2015). Parents Forever: Evaluation of an online divorce education program. *Journal of Divorce & Remarriage*, 56(4), 261–276. doi:10.1080/10502556.2015.1025900
- Becher, E. H., McGuire, J. K., McCann, E. M., Powell, S., Cronin, S. E., & Deenanath, V. (2018). Extension-based divorce education: A quasi-experimental design study of the Parents Forever program. *Journal of Divorce & Remarriage*, 59(8), 633–652. doi:10.1080/10502556.2018.1466256
- Bryson, J. M., Patton, M. Q., & Bowman, R. A. (2011). Working with evaluation stakeholders: A rationale, step-wise approach and toolkit. *Evaluation and Program Planning*, 34(1), 1–12. doi:10.1016/j.evalprogplan.2010.07.001
- Chazdon, S., Emery, M., Hansen, D., Higgins, L., & Sero, R. (2017). *A field guide to ripple effects mapping*. Minneapolis, MN: University of Minnesota Libraries Publishing. Retrieved from <https://www.lib.umn.edu/publishing/monographs/program-evaluation-series#Book%202>
- Chazdon, S., & Paine, N. (2014). Evaluating for public value: Clarifying the relationship between public value and program evaluation. *Journal of Human Sciences and Extension*, 2(2), 100–119.
- Conner, R. F., Fitzpatrick, J. L., & Rog, D. J. (2012). A first step forward: Context assessment. *New Directions for Evaluation*, 135, 89–105. doi:10.1002/ev.20029
- Cronin, S., Becher, E. H., McCann, E., McGuire, J., & Powell, S. (2017). Relational conflict and outcomes from an online divorce education program. *Evaluation and Program Planning*, 62, 49–55. doi:10.1016/j.evalprogplan.2017.02.008
- 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.
- Dworkin, J., & Karahan, A. R. (2005). Parents Forever: Evaluation of a divorce education curriculum. *Journal of Extension*, 43(1), Article 1RIB6. Retrieved from <https://www.joe.org/joe/2005february/rb6.php>
- Greene, J. C. (1994). Qualitative program evaluation: Practice and promise. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 530–544). Thousand Oaks, CA: Sage.
- Greene, J. C. (2015). How evidence earns credibility in evaluation. 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. 205–220). Thousand Oaks, CA: Sage.

- Hansen Kollock, D., Flage, L., Chazdon, S., Paine, N., & Higgins, L. (2012). Ripple effect mapping: A "radiant" way to capture program impacts. *Journal of Extension*, 50(5), Article 5TOT6. Retrieved from <https://www.joe.org/joe/2012october/tt6.php>
- Julnes, G., & Rog, D. (2015). Actionable evidence in context: Contextual influences on adequacy and appropriateness of method choice in evaluation. 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. 221–258). Thousand Oaks, CA: Sage.
- Mathison, S. (2015). Seeing is believing: Using images as evidence in evaluation. 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. 157–176). Thousand Oaks, CA: Sage.
- Miller, R. L. (2015). How people judge the credibility of information. 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. 39–61). Thousand Oaks, CA: Sage.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4<sup>th</sup> ed.). Thousand Oaks, CA: Sage.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York, NY: Guilford Press.
- Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4<sup>th</sup> ed.). Thousand Oaks, CA: Sage.
- Ragin, C. C., & Amoroso, L. M. (2011). *Constructing social research: The unity and diversity of method* (2<sup>nd</sup> ed.). Los Angeles, CA: Sage.
- Rasmussen, C. M., Armstrong, J., & Chazdon, S. (2011). Bridging Brown County: Captivating social capital as a means to community change. *Journal of Leadership Education*, 10(1). Retrieved from [https://journalofleadershiped.org/wp-content/uploads/2019/02/10\\_1\\_Rasmussen\\_Armstrong\\_Chazdon.pdf](https://journalofleadershiped.org/wp-content/uploads/2019/02/10_1_Rasmussen_Armstrong_Chazdon.pdf)
- Rog, D. J. (2012). When background becomes foreground: Toward context-sensitive evaluation practice. *New Directions for Evaluation*, 135, 25–40. doi:10.1002/ev.20025
- Rog, D. J., Fitzpatrick, J. L., & Conner, R. F. (Eds.). (2012). *Context: A framework for its influence on evaluation practice*. *New Directions for Evaluation*, 135.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39–83). Chicago, IL: Rand McNally.
- Stacey, R. D. (1996). *Complexity and creativity in organizations*. San Francisco, CA: Berrett-Koehler Publishers.
- Stake, R. E. (1990). Situational context as influence on evaluation design and use. *Studies in Educational Evaluation*, 16(2), 231–246. doi:10.1016/S0191-491X(05)80027-6
- Snowden, D. (2002). Complex acts of knowing: Paradox and descriptive self-awareness. *Journal of Knowledge Management*, 6(2), 100–111.

- Tuck, B., Chazdon, S., Rasmussen, C., & Bohn, H. (2019). *Economic value of the McLeod for Tomorrow leadership program, 2017*. St. Paul, MN: University of Minnesota Extension. Retrieved from <https://extension.umn.edu/community-research/mcleod-tomorrow-leadership-program>
- U.S. Agency for International Development. (1996). *Performance monitoring and evaluation tips: Conducting key informant interviews*. Retrieved from [https://pdf.usaid.gov/pdf\\_docs/PNABS541.pdf](https://pdf.usaid.gov/pdf_docs/PNABS541.pdf)
- Weiss, C. H., & Bucuvalas, M. J. (1980). Truth tests and utility tests: Decision-makers' frames of reference for social science research. *American Sociological Review*, 45(2), 302–313.
- Yin, R. K. (1993). *Applications of case study research*. Thousand Oaks, CA: Sage.
- Zimmerman, B. (2001). *Ralph Stacey's agreement & certainty matrix*. Toronto, Canada: Schulich School of Business, York University. Retrieved from [https://www.betterevaluation.org/en/resources/guide/ralph\\_staceys\\_agreement\\_and\\_certainty\\_matrix](https://www.betterevaluation.org/en/resources/guide/ralph_staceys_agreement_and_certainty_matrix)

*Scott Chazdon*, Ph.D., is an Evaluation and Research Specialist with the Extension Center for Community Vitality, University of Minnesota. Dr. Chazdon oversees program evaluation for the Center and conducts applied research activities on rural community development topics.

*Samantha Grant* is Evaluation Director for the University of Minnesota Extension Center for Youth Development. She evaluates the Minnesota 4-H program and focuses on teaching others about how to present data effectively through visual reports and presentations.



## **Appendix: Interview Protocol**

### **Interviews with Program Staff**

1. Describe the program/intervention.
2. Who are the stakeholders in the program? Who is not a stakeholder? Who are trying to influence with your evidence?
3. To what extent is there certainty among various stakeholders about how to solve the problem, to what extent is there agreement/conflict among stakeholders about how to achieve desired outcomes?
4. Describe the evaluation design. What type of evaluation evidence did you collect? Please tell us more about why you chose the evaluation methods you chose?
5. How was the evaluation implemented? What were the results? Did the stakeholders think the evidence was credible? Why did they think it was credible? Or did they want something different? How have you responded?
6. How have the results been used?

### **Interviews with External Stakeholders**

1. Please tell us about your relationship with the Extension program and/or program staff?
2. What results do you expect from the program?
3. What results do you expect Extension to be able to measure?
4. Are you satisfied that Extension's evaluation efforts provide you the evidence you need that the program is achieving intended results?
5. Beyond the results, are there other aspects of the program that you care about? Please explain.
6. Do you believe the evaluation evidence presented by program staff is credible? Why or why not?
7. Under what circumstances would you lose confidence in the program?
8. Is there other evidence you would want to see about the program?

## 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](mailto: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).

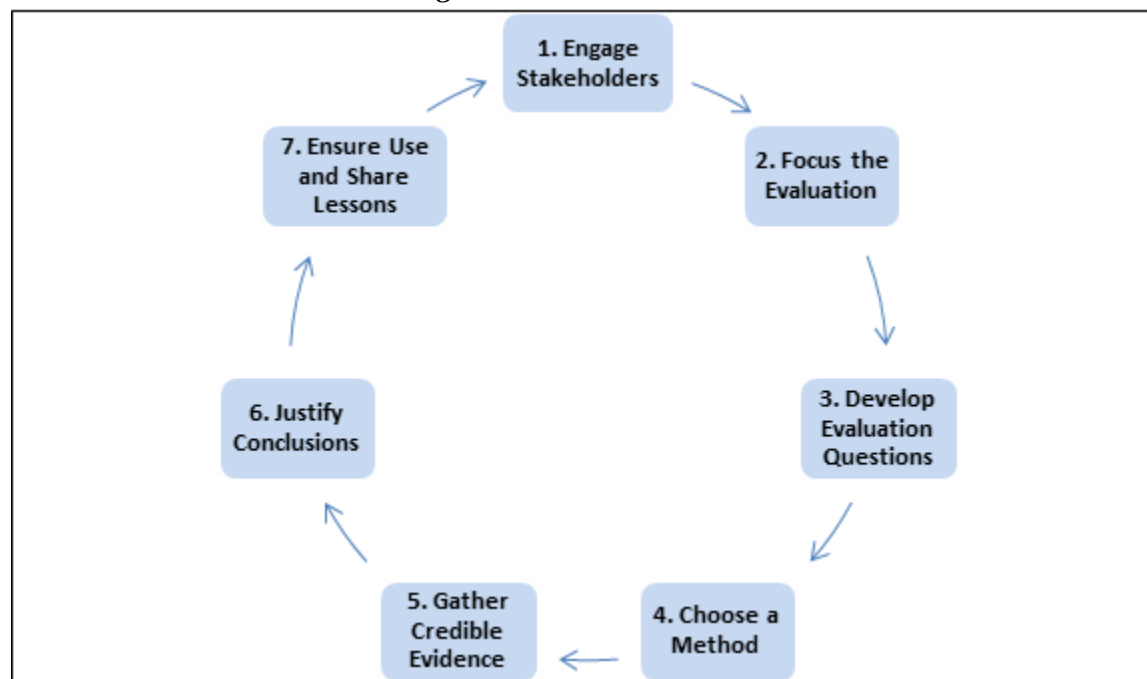
Utilizing the wrong methods could cause significant limitations on what can be interpreted from the results. Having a keen understanding of research methods allows evaluators and program participants alike, to become more astute consumers of evidence (Gooden & Berry-James, 2018).

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.

## Measurement and Credible Evidence in Extension Evaluations

**Marc T. Braverman**

*Oregon State University*

*This article examines the concept of credible evidence in Extension evaluations with specific attention to the measures and measurement strategies used to collect and create data. Credibility depends on multiple factors, including data quality and methodological rigor, characteristics of the stakeholder audience, stakeholder beliefs about the information source, and the evaluation context. Measurement planning involves a process of making thoughtful decisions about choosing study variables, measurement strategies, and specific measures that adequately reflect the content and goals of the program being evaluated. The use of specific measures may also entail implicit assumptions, e.g., that the respondent is being truthful and accurate, which must be accepted if resulting data are to be viewed as credible. The article discusses aspects of measurement quality, including reliability and validity, for both quantitative and qualitative forms of data. Program stakeholders should be encouraged to be attentive, reflective, and critical in their analysis of evaluation evidence, and their views on what makes data credible must be understood and considered. The use of common measures in evaluating multi-site programs can be valuable, but only if the measures are fully appropriate for local sites. The article concludes with a summary of implications and recommendations for Extension evaluation practice.*

*Keywords:* common measures, credible evidence, evaluation planning, Extension evaluation, measurement strategy, validity

***How both program processes and outcomes are measured will largely determine the degree to which a program is determined to be effective.***

—Schwandt (2015, p. 82)

### Introduction

The credibility of an evaluation refers to the likelihood that stakeholders will accept the evaluation results as convincing and will accept the conclusions and recommendations as reasonable and justified. Judgments about the soundness, credibility, and persuasiveness of evidence set the stage for follow-up action and utilization, and thus, one can speak of the *actionability* of evaluation findings (Mark, 2015). Credibility for a particular audience depends on numerous factors including its timeliness, the relevance of the primary questions, the use of a rigorous design to answer those questions, and the quality of both the evidence and the

---

Direct correspondence to Marc T. Braverman at [Marc.Braverman@oregonstate.edu](mailto:Marc.Braverman@oregonstate.edu)

conclusions (Donaldson, 2015). Measurement is at the heart of the process and is the focus of this article. The choice of measurement strategies and instruments and the effectiveness of data collection produce the raw material on which the analyses and interpretations rest.

In this article, I discuss several dimensions of measurement in Extension evaluations, and the implications for how convincing—that is, how credible and actionable—the findings and recommendations will be for program stakeholders. I begin with some clarification about the concept of credibility and why it is broader than simply an assessment of data quality, and I present a model of the components of credibility with regard to evaluation evidence. I then describe some relevant principles of the measurement process in program evaluation and their relevance to credibility. Based on these analyses, I conclude with recommendations for increasing the credibility and actionability of Extension evaluations through thoughtful measurement decisions.

### **Isn't Credibility Just a Reflection of Data Quality?**

The concept of credibility, widely discussed in the current evaluation literature (e.g., Donaldson, Christie, & Mark, 2015), is universally acknowledged to depend, at least in part, on rigorous methods of investigation. So a good place to start is to ask why we speak of the *credibility* of our results and conclusions rather than simply the *quality* and *rigor* of the data. Some writers who use the term “credible evidence” do take this approach, concentrating only on the rigor of the methods used to produce the evidence. These methodological factors include the selected measurement instruments, the sampling procedures, and the evaluation design (often favoring randomized controlled trials). However, my own view is that credibility, though it undoubtedly depends on data quality and methodological rigor, is a more complex and multi-faceted concept. Since it refers to the likelihood that evidence is to be believed and judged as accurate, there must be a human angle involved. Observers will often disagree on what constitutes the strongest methodologies for collecting data, or even on what a particular response means. Some stakeholders are more skeptical than others about program results, methods, or assumptions about evaluation data.

### **A Model of the Influences on the Credibility of Evaluation Findings**

Stakeholder judgments about credible evidence in Extension evaluations may be influenced by four kinds of factors:

- ***Data quality.*** This term encompasses a wide-ranging set of considerations, including the reliability of measures, the formats of the data, the timing of data collection, and the validity of the conclusions stemming from the evidence.
- ***Characteristics of the stakeholder audience(s).*** Most Extension evaluations will have multiple primary audiences, which may include administrators, internal program staff, clients (including adult participants as well as parents of participants in the case

of youth programs), elected officials, and other community members. Some of these stakeholders will be more informed, more skeptical, more interested, more invested in the success of the program, better able to understand evaluation methodology, and/or more actively engaged in the evaluation process than others. Their tendencies to accept results as credible will vary significantly as a result of these predispositions.

- ***Stakeholder beliefs about the information source.*** Communications will be more readily accepted by stakeholders if they accept the information source as objective, trustworthy, and knowledgeable. The information source might be an organization such as Extension, a public agency, a private business enterprise, or an individual contractor.
- ***The context for the evaluation.*** All program evaluations take place within a larger context, which includes the nature of the organization delivering the program, the time and resource constraints on the evaluation, the decisions that might be riding on the evaluation results, and so on. That context will influence both how the evaluation is conducted and how it is received and accepted by its stakeholders.

Only the first of these factors, data quality, is directly related to what is thought of as methodological rigor. The other factors, to varying degrees, are dependent on perceptions, potential biases, predispositions, and political priorities, thus adding to the complexity of the concept of *credibility* in the assessment of evidence.

### **Perspectives on the Measurement Process in Program Evaluation**

Several important issues about the measurement process provide background context for making assessments about data quality and credibility.

#### **The Links Between Constructs, Variables, and Measures**

In an earlier paper (Braverman, 2013), I described a model for developing evaluation measures that involves a four-step process. “The measurement specification process generally begins with the broadly conceived target construct, which reflects, often in everyday language, the issue that the program is designed to address” (p. 102). Examples of these constructs could be “healthy eating,” “parenting skills,” “interest in science,” “leadership skills,” “knowledge about common garden pests,” and so on. Once the target construct is decided on, ideally with participation from stakeholders, the evaluator must decide how the construct will be translated into a variable to be measured, which measurement approach will be used, and finally, what specific instrument will be included in the evaluation. The instrument will consist of the specific questions (or sometimes a single question) that will be used to measure the target construct. Some instruments do a far better job of representing the target construct than others. If an evaluation has sufficient resources and the construct is of central importance, the evaluation planners might decide to measure the construct using multiple variables and instruments. The sequence for some of the decisions to be made in this process is illustrated for several sample constructs in Table 1.

## Measurement Rigor

Braverman and Arnold (2008) defined methodological rigor as “a characteristic of evaluation studies that refers to the strength of the design’s underlying logic and the confidence with which conclusions can be drawn. An evaluation that incorporates attention to methodological rigor will be in a better position to afford evidence and conclusions that can stand up to critical analysis” (p. 72). With specific regard to aspects of measurement in evaluation, Braverman and Arnold described several components of rigor that relate to measurement strategies. These include the conceptualization of program outcomes, decisions about how those outcomes will be represented by the evaluation measures, and the data collection strategies to be used.

**Table 1. Measurement Planning: Identifying Potential Options in the Progression from Construct to Evaluation Measures**

<b>General Construct</b>	<b>Specific Variables That Might Be Used to Represent the Construct (Selected)</b>	<b>Related Variables That Could Potentially Also Be Used as Relevant Outcomes (Selected)</b>	<b>Potential Measurement Strategies</b>
<b>Parenting skills</b>	<ul style="list-style-type: none"> <li>• Identification of parenting style</li> <li>• Parenting self-efficacy</li> <li>• Parent-child communication</li> <li>• Parent-child interactions:               <ul style="list-style-type: none"> <li>• Expression of warmth</li> <li>• Empathy</li> <li>• Responsiveness</li> </ul> </li> <li>• Discipline practices</li> <li>• Monitoring</li> </ul>	<ul style="list-style-type: none"> <li>• Parental stress</li> <li>• Parenting satisfaction</li> <li>• Parent-child relationship quality</li> <li>• Positive child behaviors</li> </ul>	<ul style="list-style-type: none"> <li>• Survey self-report questionnaire (scales or specific items):               <ul style="list-style-type: none"> <li>• Self-ratings of knowledge gain</li> <li>• Behavioral self-report</li> </ul> </li> <li>• Observation of parent-child interaction:               <ul style="list-style-type: none"> <li>• Live observations</li> <li>• Videotaped interactions</li> </ul> </li> <li>• Interview</li> </ul>
<b>Physical activity</b>	<ul style="list-style-type: none"> <li>• Daily, weekly, or monthly total minutes of <i>Moderate to Vigorous Physical Activity</i> (MVPA)</li> <li>• Number of days per week with at least 1 hour MVPA</li> <li>• Average or total number of steps per day</li> <li>• Physiological tracking (heart rate)</li> </ul>	<ul style="list-style-type: none"> <li>• Body mass index</li> <li>• Sedentary behavior (e.g., sitting time) per day</li> <li>• Overall physical fitness</li> </ul>	<ul style="list-style-type: none"> <li>• Survey self-report questionnaire (scales or specific items)</li> <li>• Activity logs or diaries</li> <li>• Interview</li> <li>• Activity monitors:               <ul style="list-style-type: none"> <li>• Pedometers</li> <li>• Accelerometers</li> </ul> </li> <li>• Direct observation</li> </ul>
<b>Healthy eating</b>	<ul style="list-style-type: none"> <li>• Overall eating patterns</li> <li>• Food consumed in past week (or day or month)</li> <li>• Meal observation</li> <li>• Food available in home</li> <li>• Eating intentions</li> </ul>	<ul style="list-style-type: none"> <li>• Family eating practices</li> <li>• Knowledge of:               <ul style="list-style-type: none"> <li>• Nutrition</li> <li>• USDA’s MyPlate</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Survey self-report questionnaire (scales or specific items):               <ul style="list-style-type: none"> <li>• Food frequencies</li> <li>• Dietary recall (e.g., over 24 hours)</li> </ul> </li> <li>• Tracking of food purchases (e.g., from debit card)</li> <li>• Pantry inventory inspection</li> </ul>



### **Kinds of Data: Quantitative and Qualitative**

The measurement process will differ in significant ways depending on whether the data to be collected and analyzed are quantitative (in numerical form) or qualitative (in text form). The form of the data may change between data collection and analysis. For example, short-answer responses may be coded into categories or numerical quantities for certain kinds of data analysis. Qualitative data will be the product of free-form, open-ended responses on surveys, extended answers to interview questions, daily log entries, text-based descriptions from observers, and so on.

There are multiple and varied approaches to analyzing both categories of data, but in general, the analyses of quantitative and qualitative data tend to be distinctly different processes. Indeed, they often reflect different goals for what is to be described and learned. Quantitative data analyses generally take the form of summarizing the dataset in terms of descriptive or inferential statistics, e.g., through calculating mean scores to present a picture of the sample “on the average,” or of exploring quantitative relationships between variables. By contrast, qualitative data analyses examine interviews, raw video and audio evidence, narratives, and observational notes to generate “rich” and “thick” descriptions of programs. These approaches can often provide unexpected insights that cannot be captured by checklists, surveys, or tests.

Mixed-methods evaluation designs, which make use of both quantitative and qualitative forms of data (e.g., Mertens, 2018), benefit from the distinct strengths and advantages of each format. However, because the analysis methods for the two forms of data differ greatly (e.g., Bazely, 2017), the criteria and approaches for making judgments about data quality are very different as well. Nevertheless, crucial data quality considerations, such as the relevance of the measurement approaches for addressing the major evaluation questions, the recognition of implicit assumptions, the awareness of ambiguities in the data, the recognition of potential biases, the appropriateness of interpretations and conclusions, etc., are of primary importance for both types of data (Creswell & Creswell, 2018).

### **Implicit Assumptions in the Use of Measures**

The use of any measurement strategy or instrument entails some assumptions if we are to accept the resulting data as accurate and valid. Consider, for example, survey self-report, which is frequently used in Extension evaluations to measure attitudes, opinions, values, behavioral histories, behavioral intentions, assessments of programmatic success, and other types of outcomes. Several implicit assumptions are involved, and observers who disagree about the reasonableness of these assumptions will also disagree about the credibility of the responses. These assumptions include the following:

**1. *The respondent is trying to be truthful in reporting.*** In most cases, evaluators and evaluation audiences assume that the survey respondent is being truthful and honest. However, there are times when this assumption might be questionable, e.g., because respondents may be motivated

to provide a socially desirable response (SDR; Dillman, Smyth, & Christian, 2014; Tourangeau, Rips, & Rasinski, 2000). Questions that ask about sensitive topics, such as household income, financial habits, drug use, sexual activity, or the respondent's compliance with regulatory requirements, are especially vulnerable to this form of bias. In these instances, the honesty of the survey respondents—or at least a subset of those respondents—might be reasonably judged to be open to question.

Social desirability scales are sometimes used to estimate the degree of bias due to SDR (Perinelli & Gremigni, 2016), but these add length to the survey and have limited effectiveness. In addition, a technique known as *randomized response* (Höglinger, Jann, & Diekmann, 2016) has been suggested as a formal methodological strategy to deal with sensitive survey questions, but it is complex and cumbersome. Most often, especially in Extension settings, common-sense strategies are employed to minimize social desirability bias, such as making the questionnaire anonymous, with the underlying assumption being that respondents will perceive that there is no reason to be dishonest if they cannot be identified. In practice, making questionnaires anonymous for sensitive question content, while undoubtedly helpful, is not fully satisfactory, because respondents' inclination to be honest and forthcoming is not entirely guided by logic (Tourangeau et al., 2000).

In addition to motives of self-protection, respondents may be motivated to answer in a way that they perceive as desired by the evaluator or program staff. For example, participants in Extension health education classes may be aware that it will benefit the program if they report positive personal impacts, such as increases in healthy eating and regular exercise. Thus, sometimes respondents may be trying to protect themselves by providing what they perceive as socially desirable responses, and sometimes they may be trying to protect the program in which they have participated.

**2. *The respondent is able to be reasonably accurate in answering the questions and is willing to make the effort to do so.*** Some questions that appear on surveys require cognitive effort to respond accurately, e.g., involving thoughtful judgment or memory recall. Examples include autobiographical behavior questions that ask about behavioral history or behavioral frequency, such as: “Which of the following foods did you eat yesterday?”, “How many minutes of moderate-to-vigorous physical exercise do you usually engage in each week, on average?”, and “When was the last time you had a physical examination from your doctor?”. The last of those questions, which requires the respondent to accurately recall the amount of time that has passed since a previous event, is especially subject to errors of either overestimating or underestimating time periods, a phenomenon known as telescoping (Braverman, 1996).

Assuming that respondents are able to provide the necessary information, one must also trust that they are willing to engage in the concentrated effort needed to answer accurately. In many cases, respondents have been found to expend just enough effort to provide what they consider a “good

enough” answer, which might not reflect the degree of accuracy that the evaluator desires or expects; this phenomenon has been called *satisficing* (Krosnick, Narayan, & Smith, 1996; Tourangeau et al., 2000). Thus, the evaluator must be prepared to consider the question: Even if respondents are able to answer these questions, can we justifiably assume that they are motivated to do so?

**3. *The respondent is an appropriate source for the information desired.*** In addition to cognitive inaccuracies, respondents might also simply not have access to the requested information. For example, parents may be asked to estimate the number of minutes of screen time in which their children engage each week, their children’s average amount of exercise, or the hours they spend doing homework, even though parents may not know these aspects of their children’s daily lives. The fact that parents’ level of information about these topics might be inadequate does not always stop evaluators from asking about them, and, quite often, that inadequacy does not even stop parents from answering the questions. Survey researchers have found that respondents will often answer a survey question with a complete guess, rather than selecting “I don’t know” or leaving the response blank (Groves, 1989).

In summary, assumptions about respondents’ honesty, accuracy, motivation, and knowledge are sometimes justified, sometimes not. However, these and other assumptions are frequently taken for granted without critical appraisal. If even a relatively small percentage of the respondents are unreliable in these respects, the overall quality of the dataset will be compromised. Furthermore, concerns about these assumptions should be accentuated when the respondents are children. Depending on their ages, many children’s power of recall and their motivation may be insufficient to answer certain questions.

### **Characteristics of Measurement that Affect the Credibility of Evidence**

Given this background context, let us examine several specific factors that, in many cases, might influence credibility. Most of these directly involve data quality and rigor, but others involve the perspectives of different evaluation audiences and the larger context in which the evaluation is conducted.

#### **The Reliability and Validity of Measures**

*Reliability* refers to the consistency of a measurement. If a survey item or scale is found to produce widely varying responses across conditions in which consistency is expected, the accuracy of scores from that item or scale will be suspect. That consistency, or lack of it, will influence the confidence we can place in the scores.

Our expectation of consistency is tied to our understanding about the variable we are trying to measure. For example, we would expect weight, height, and body mass index to be very stable if measured twice within 30 minutes (assuming no eating in the interim), but we would not have

that same expectation for blood pressure, which varies to some degree every time it is taken (Bandalos, 2018). Thus, proper interpretation of reliability requires that we understand our variables, particularly with regard to prior expectations about the consistency and replicability of scores.

The “consistency” of measures can refer to replications across different time periods, different versions of a test, individual questions within a scale, or different individuals who are making judgments to produce the scores. Several of the major categories of reliability, as applied to quantitative forms of data, are the following:

- **Test-retest reliability.** This refers to consistency across short periods of time for variables that we expect to be relatively stable. Test-retest reliability is not appropriate for variables that experience change, e.g., indicators of mood or fatigue.
- **Interrater reliability.** Some outcome scores are based on judgments by raters. Examples might include essay tests to assess writing ability, athletic efforts to assess mastery of physical skills, or observations of parent-child interactions to assign scores on parenting style. The accuracy of these scores depends on the skill of the rater, and there should be minimal variation in scores based on who is doing the rating. Interrater reliability, often measured with the statistic Cohen’s kappa, is a measure of consistency across judges.
- **Internal consistency.** This refers to the consistency of items that make up a scale. For example, if we have a brief, 8-item scale to measure leadership style, each item should correlate positively with the other items and contribute toward the overall score. Internal consistency is usually measured with the statistic Cronbach’s alpha and is relevant for knowledge tests, attitude scales, and measures of psychological constructs. The statistic can also be used to produce the strongest scale from a set of candidate items.

Several qualitative research theorists have discussed how the concept of reliability can be applied to the analysis of qualitative data, although there is no consensus on this topic. Creswell and Poth (2018) place emphasis on the coding process: “In qualitative research, *reliability* often refers to the stability of responses to multiple coders of data sets” (p. 264). Miles, Huberman, and Saldaña (2014) note, with regard to what they call the “reliability/dependability/auditability” of qualitative data: “The underlying issue here is whether the process of the study is consistent, reasonably stable over time and across researchers and methods. We are addressing issues of quality and integrity: Have things been done with reasonable care?” (p. 312).

*Validity* refers to the appropriateness of interpretations, judgments, and conclusions that are made on the basis of scores. Validity theory has evolved significantly in the last several decades. The modern conception, developed by Samuel Messick (1989) and others, rejects the previously dominant view that validity is an inherent, identifiable quality of measures and tests. It is not

accurate to talk about “a valid test,” because for any given test or measure, some uses will be valid while others will not. Instead, validity is a property of the ways that measurement scores are used. Thus, one can talk about valid uses, inferences, or conclusions that are based on the information from one or more measures. (See Bandalos, 2018, or other recent texts on measurement and psychometrics for further discussion.)

To cite an example relevant to Extension, after a series of trainings for volunteers in a food preservation program, participants may be asked to rate the amount they learned with regard to food safety precautions. Self-ratings of this type cannot be considered a rigorous measurement strategy for the assessment of knowledge (as will be discussed further below). Thus, it might be considered a valid use of those ratings to make relatively low-consequence decisions about how the training sessions can be revised to be more interesting and comprehensive. However, it would not be a valid use of the self-rating scores to determine and certify which of the volunteers are sufficiently prepared to provide advice to the public regarding the safety of specific food preservation practices.

For qualitative data, as with the concept of reliability, there are competing perspectives on how the concept of validity can or should be applied. Miles et al. (2014) summarize the debate: “*Validity* is a contested term among selected qualitative researchers. Some feel that this traditional quantitative construct. . . has no place in qualitative inquiry. Alternative terms such as *verisimilitude* and *a persuasively written account* are preferred. But other qualitative methodologists continue to use the term purposefully because it suggests a more rigorous stance toward our work” (p. 313). Goodrick and Rogers (2015) prefer the term *inference quality* in place of *validity* for qualitative data, and they describe multiple strategies for strengthening the quality of inferences from qualitative analyses, depending on the specific analytic approach.

### **Scaling and Interpretive Clarity**

Several principles of effective measurement relate to the way that potential responses are scaled. Consider a behavioral frequency question that may be asked in a nutrition education program, about the consumption of sugar-sweetened beverages (SSBs): “I drink sugar-sweetened beverages,” with response options consisting of *Yes* and *No*. Most evaluators would consider that wording to be inadequate for several reasons. First, it is unclear what the dividing line should be between *Yes* and *No* if there is no guidance provided within the question. The respondent might interpret *Yes* to mean either “ever” or “regularly”—two very discrepant meanings. Due to the question’s ambiguity, different respondents will probably make different judgments, and it will not be possible to understand precisely what information they have given us. Furthermore, even if the dividing line between *Yes* and *No* is clear, this behavior should be best expressed as a *range* of frequencies, since the two options are insufficient to capture the variability that exists in people’s lives. Respondents might drink SSBs every day, once a week, once a month, or never. To be useful, the collected data should be able to reflect real-life variation, to whatever degree is needed for our intended uses.

In addition to considering the number of response options, the wording of those options is important as well. For behavioral frequency, many Extension evaluations use some variation of *Rarely / Sometimes / Often / Always*. Although the inclusion of four options may provide sufficient spread, the labels are vague with regard to what the options actually mean, leading to ambiguity in the resulting responses. Survey research texts (e.g., Dillman et al., 2014) recommend using a set of options that are worded as clearly as reasonably possible. An example might be: *Rarely or never / About once a month / About once a week / More than once a week*.

To summarize, if we ask a question in a way that fails to represent the range of variation that exists among our respondents, our data will lack important information. Similarly, if we do not really understand what our respondents have told us with their answers, our data will be ambiguous. In assessing evidence, less informed stakeholders may not recognize these measurement problems, but more knowledgeable stakeholders will find such results confusing or untrustworthy.

### **Acknowledging and Compensating for the Limitations of Individual Measures**

Every measure used in evaluation has limitations. The use of multiple approaches to measure a single critical construct, known as *triangulation*, can increase our confidence in the findings by exploiting strengths and compensating for limitations in individual measures.

Consider the construct of *healthy eating*, illustrated in Table 1. This can be measured in numerous ways, including direct self-report on survey questionnaires, food diaries, food pantry surveys, and plate waste studies (Braverman, 2013). Self-report is probably the most commonly used due to its convenience, time efficiency, low expense, and capacity to address past time periods. Yet as noted above, survey self-report also entails drawbacks, such as the risk of deliberate misrepresentation and potential problems with memory recall, question comprehension, and/or motivation to respond accurately. Therefore, an evaluator may choose to include additional methods of assessing eating behavior to supplement the information gained from self-report. If the information from multiple measures provides a consistent picture, the strength and credibility of the findings will be enhanced.

The major disadvantage of using multiple measures is the extra time and effort it requires for data collection. Respondents might lose patience or be confused by what they perceive as redundancy. Since the time allotted for data collection is almost always limited, other variables may be left out if a great deal of attention is devoted to accurately assessing one particular construct. The evaluator must weigh the advantages and disadvantages of using multiple measurements to strengthen the inferences regarding a single component of the evaluation, compared to addressing a broader set of questions.

## Examining the Implicit Assumptions Associated with Measures and Measurement Strategies

As described earlier, all measures involve assumptions, and credibility is related to the reasonableness of those assumptions, relative to scientific and evaluation standards, or personal criteria of program stakeholders. An example is the use of self-ratings to determine levels of learners' skills and knowledge, a widely-used measurement strategy in Extension program evaluations. Rather than give participants a subject matter test on the content of an educational program (e.g., on gardening, diet and nutrition, parenting, personal finance, etc.), participants are simply asked to report the degree of their knowledge and/or the amount they have learned from the program. The level of self-rated knowledge can be compared pre- and post-program, and if it has increased, the conclusion will usually be drawn that the program has been effective.

Self-ratings are considered to be a form of *indirect* rather than direct measurement (Banta, 2004; Braverman, 2013). With regard to rigor, this approach is a weak strategy for assessing subject matter knowledge because many people will either under- or overestimate their own levels of mastery, and there is usually little or no evidence to support the accuracy of those judgments. The most obvious alternative strategy is to assess knowledge directly with a subject matter test. However, this will typically be more logistically difficult in several ways, which explains why it is not more commonly used. First, it would require more time for measurement, probably involving multiple questions about the subject matter, whereas a self-rating might involve only a single survey item with a 4- or 5-point rating scale. Second, an appropriate knowledge test that closely matches the Extension program curriculum would probably be available only rarely, and thus would often need to be created by the program staff or evaluator. Therefore, one can see why the use of self-ratings might be preferred based on evaluation logistics. However, with regard to the data quality itself, the credibility of the direct test is far superior, and the reliance on self-ratings in Extension program evaluations can result in reduced credibility of the findings.

Another illustration of potentially questionable assumptions is the measurement of behavioral intentions in place of measuring the actual behaviors of interest. For example, following an educational program, an evaluation may measure program participants' *intentions* to engage in regular physical activity or personal financial planning. These intentions can be a valid type of outcome on which to focus in an evaluation, but it is a mistake to assume that positive intentions can be equated with actual behavior changes, which are usually the outcomes of greatest interest in the assessment of program impact. Intentions can be easier to measure than behaviors because they can be assessed immediately following the end of the program, whereas the assessment of actual behaviors often requires the passage of time before those behaviors kick in. This would necessitate additional contacts with the program clients, many of whom might not respond. However, the danger of equating intentions with behaviors was demonstrated by Lohse, Wall, and Gromis (2011), who reported that participants' intentions to increase fruit and vegetable consumption following an Extension nutrition education class correlated poorly with their actual

consumption three weeks after the program's end. Yet sometimes these two kinds of variables—behavioral intentions and actual behaviors—are treated interchangeably in making claims for a program's success.

### **Evaluation Stakeholders: Variations in Sophistication, Judgments, and Priorities**

As discussed earlier, the credibility of data is, in part, subjectively and individually determined. Credibility, by definition, refers to believability, and whether someone believes a set of conclusions based on a body of evidence is not entirely under the control of the evaluator, even if the evaluation has been designed and implemented in an exemplary fashion. Credibility depends to a certain extent on the perspectives of stakeholders. And as Miller (2015) notes, "it is not always the case that people will evaluate the credibility of evidence or information through rigorous analytical means. Indeed, the default appears to be *not* to analyze information rigorously and to rely instead on an initial intuitive judgment that is based largely on peripheral informational cues" (p. 49).

Ideally, stakeholder judgments about the credibility of evidence will be based on considerations of evaluation rigor. But this implies the presence of a critical, attentive, and knowledgeable audience. Without this orientation, the notion of credible evidence becomes irrelevant: even the least rigorous evaluation evidence will do.

Why would an evaluation audience not be engaged in this process? One reason is that they may feel they lack the necessary expertise, in which case they may relinquish responsibility to the evaluator for understanding and interpreting the results. Indeed, specialized knowledge is often needed, such as in decisions about how best to measure constructs when multiple options may be available. An evaluator must make measurement choices based on time, resources, and other considerations. Nevertheless, interested audiences may want to know the reasoning behind a particular choice and the evaluator's justification for why it was considered best.

A second reason for a lack of stakeholders' engagement would be if they are heavily invested in a particular result, e.g., finding evidence of outstanding program success, which could influence their willingness to be objective. If the evaluation comes back to their liking, they may embrace those results, disregarding considerations about rigor and the relative strength or weakness of the evidence. In such cases, it is up to the evaluator to strive for objective interpretation, recognizing and acknowledging whatever limitations in the measurements may exist. Even if some stakeholder audiences, such as program participants, are willing to take the evaluator's word about the strength of evidence, it would be a mistake to assume that this is true for all key stakeholders. At some point, it is likely that the evaluator will face tough questions about the methodological choices that were made.

Active and engaged stakeholders can help to ensure that evaluation results are used appropriately. To promote this orientation among stakeholders, they should be encouraged to



engage in *evaluative thinking*, which is being increasingly recognized as a key component of evaluation use. Evaluative thinking (ET) has been described as “in essence, critical thinking applied to contexts of evaluation” (Buckley, Archibald, Hargraves, & Trochim, 2015, p. 376). It is especially relevant to the goal of building evaluation capacity within organizations (Patton, 2018; Vo & Archibald, 2018).

Evaluative thinking is important because credibility judgments are enhanced by the growing sophistication of stakeholder audiences. Being able to understand the basis for strong evidence leads to more appropriate and effective evaluation use. Evaluation users who demand or value rigorous evaluation methods will be better able to use evidence to build more effective programs. Buckley et al. (2015) make this point in noting that evaluative thinking “is the substrate that allows evaluation to grow and thrive. . . . ET is a protective factor to prevent against the risk of senseless, mindless evaluation” (p. 378). In other words, promoting evaluative thinking among our Extension program stakeholders with regard to measurement and other evaluation elements will eventually result in stronger programs.

### **The Use of Common Measures to Evaluate Multi-site Programs**

Many Extension programs are implemented across multiple community sites. Some of these, such as 4-H Youth Development, are typically delivered in every county within a state. In addition, innovative projects funded by external federal, state, or foundation grants frequently involve wide implementation, sometimes covering sites in multiple states. In many of these programs, the goals, objectives, and target outcomes across sites are highly similar, and it can be logical to seek to measure common outcomes using standardized measurement instruments and strategies. For example, based on this reasoning, the National 4-H Council has developed common measures to be used in evaluating youth programs in the areas of science, healthy living, civic engagement, college/career readiness, and positive youth development (National 4-H Council, 2019), with the expectation that these measures will be used in programs that the Council funds within these topic areas.

Common measures are also used by the National Institute of Food and Agriculture’s (NIFA) Children, Youth, and Families at Risk (CYFAR) Initiative (CYFAR Professional Development and Technical Assistance Center, 2018). CYFAR provides grants to every state to implement innovative projects that serve at-risk families. The projects are planned at the state level, and thus there is no expectation of continuity or coordination of specific program activities from one state to another. Further, within a state, there is often diversity in how the program is shaped and delivered across counties, and even between community sites within counties.

This broad tapestry of projects, many of which share common aims, presents a daunting challenge for the task of measuring their impact on target outcomes. In response, a national CYFAR evaluation team developed a series of scales to be used as common measures across sites (Payne & McDonald, 2012, 2015). Targeted constructs that are intended for measurement

in every CYFAR project include program quality, youth program participation, and “core competencies” such as caring, decision-making, and social conscience. In addition, common measures are made available for specific program outcomes, such as leadership, parenting, nutrition, physical activity, and workforce preparation, to be used by local projects that target those outcomes (CYFAR Professional Development and Technical Assistance Center, 2019).

The use of common measures across program sites can be enormously valuable, by providing continuity and standardization in the evaluation process (Table 2). Results at different sites can be aggregated to allow for evaluative conclusions at the level of the broad program initiative, while different program delivery options can be compared for relative effectiveness. Thus, the credibility and actionability of evaluation evidence will often be considerably enhanced, especially for program funders and other stakeholders at levels of administrative and policy decision-making. For example, the use of common measures to assess critical program outcomes across states allows NIFA to report about the broad impacts of the CYFAR initiative to its parent agency, the U.S. Department of Agriculture. Common measures can enhance the credibility of evaluation data at the community site level as well, if it is communicated that the data stem from highly regarded, widely used instruments.

But Table 2 presents some cautions as well. The focus and delivery of programming at different sites within a project will usually not be uniform, due to either deliberate design or natural variations based on program personnel, location, and scheduling. Therefore, a commonly used measure may have differing degrees of relevance and importance at different sites. In cases where local site staff are given relatively broad latitude to make decisions about program focus and design, the use of externally imposed measures may be an uncomfortable force-fit. The program staff may be most interested in selecting measures that reflect their own priorities, and a requirement to use common measures may introduce extra time, redundancy, and/or irrelevance into the data collection process.

For example, Lewis, Horrillo, Widaman, Worker, and Trzesniewski (2015) examined the psychometric properties of four 4-H common measures, using exploratory factor analysis on responses from 721 California youth. They found that several scales had significant levels of missing responses due to the limited applicability of some items for their respondents (e.g., “I wear a helmet when riding an all-terrain vehicle,” from the Healthy Living scale). Payne and McDonald (2015), in reviewing the CYFAR Initiative’s common measures for parenting and youth citizenship, asked program staff from seven states to rate the relevance of each of the scale items to their own state CYFAR programs. Out of 22 items making up the two scales, only 66% were rated “completely relevant”; that is, about a third of the scale items did not closely align with the content of the local programs. In sum, the appropriateness of using common measures depends on the uniformity of goals, objectives, and program content across the program sites.

**Table 2. Using Common Measures Across Program Sites: Benefits and Cautions**

<b>Potential Benefits</b>
<ul style="list-style-type: none"> <li>• <b>Consistent interpretation of constructs.</b> The consistency of measurement across program sites allows for consistent definition and interpretation of the program's core constructs. This is a strength—if that consistency across sites is an expectation of the project. (See <i>Cautions</i> below.)</li> <li>• <b>Reduced burden on local site staff.</b> Providing site staff with ready-to-go instruments makes data collection easier and can encourage evaluation practice at the sites.</li> <li>• <b>Quality control in measurement.</b> A common measure that has been carefully developed or selected can help to ensure that best practices in measurement are being followed. However, evaluators must still pay attention to issues of validity and reliability at the local sites. (See <i>Cautions</i> below.)</li> <li>• <b>Conclusions about collective impact and overall program accountability.</b> Using common measures for core constructs allows for evaluation conclusions to be drawn at a higher-order program level, combining results from individual sites. These conclusions can be highly credible and actionable for decision-making by funders, administrators, and legislators.</li> <li>• <b>Comparisons across sites.</b> Community sites may differ, in large and small ways, in how program delivery takes place. Using common measures can allow for the effectiveness of program variations to be compared with a standardized measuring stick.</li> </ul>
<b>Cautions</b>
<ul style="list-style-type: none"> <li>• <b>Appropriateness for local site circumstances.</b> Multi-item scales used as common measures might include individual items that are not relevant for a local site, e.g., if they refer to content outside the scope of the site's curriculum. This could result in program sites being improperly evaluated based on irrelevant content.</li> <li>• <b>Potential for redundancy.</b> For a variety of reasons, a local site might need to utilize its own instrument to measure the construct of interest, e.g., to track progress over time with an instrument used in prior years. In that case, the need to use the common measure in addition to the local measure may create redundancy in the measurement process.</li> <li>• <b>Psychometric properties.</b> It cannot be automatically assumed that validity and reliability estimates for the common instrument will be adequate and equivalent at each site. If site participants differ in significant ways from the populations with which the instrument was validated, e.g., in terms of age, aptitudes, cultural background, etc., validity and reliability must be established for the local context.</li> <li>• <b>Requirements for administering the measure.</b> Data collection using a common measure may require uniformity of process, e.g., with regard to time allocation, instructions, observation processes, etc. If those uniform procedures are not followed by site-level evaluators, the appropriateness of the measure may be compromised.</li> </ul>

### **Implications and Recommendations for Extension Evaluation Practice**

Based on the preceding discussion, a number of recommendations can be offered for Extension practice with regard to credible measurement in evaluation.

#### **1. Prioritize measurement quality in evaluation planning, but know why you are doing it.**

No evaluation or research study has unlimited resources. Data quality comprises many components, and even the most carefully conducted studies require decisions about the best that can be done under the circumstances. All study designs have weaknesses and limitations, and

decisions about data quality involve trade-offs between reasonable options. Design weaknesses can be minimized, but they cannot be eliminated completely.

Measurement choices should be made with the goal of making the strongest possible case for a credible study. In many cases, there will not be a single correct choice. The evaluator needs to make decisions about how each of the available options affects the study's overall strength of evidence. In many cases, each option in a decision context will have its own strengths and weaknesses.

## **2. Monitor and communicate indicators of quality and rigor in the data.**

Researchers have developed procedures and standard practices to help us understand the quality of our data. Assessing reliability, in its various forms, is one important strategy to accomplish that. Evaluators should also track and report the response rate for surveys and, if relevant, the degree of program attrition between pretest and posttest (that is, the number or proportion of participants who were included in the pretest sample but later dropped out of the program and were not part of the posttest sample). Those indicators provide background on how well our intended sample has been captured. Information should also be provided about missing data on individual measures, including the extent of missing scores and how that phenomenon has been handled in the analysis of data.

## **3. Engage with the Extension program's stakeholder audiences early in the evaluation planning process, to determine the factors influencing their judgments about the credibility of evaluation evidence.**

To the extent possible, we should bring our Extension audiences into the evaluation planning process, where methods-related issues can be discussed, digested, and decided upon. There are several advantages to this approach. First, involvement in evaluation planning will increase their buy-in to the evaluation and their interest in the results. Second, they will provide perspectives and details that the evaluator can use in weighing options about evaluation design. Third, stakeholders involved in planning and measurement decisions will be more likely to understand and accept the evaluation results, even if those results are disappointing.

## **4. Educate your Extension evaluation audiences and promote evaluative thinking.**

Extension audiences need to be reasonably educated and attentive about evaluation. If our audiences uncritically accept our evaluation conclusions, without attempting to understand the basis for our claims regarding the data, then our Extension clientele are not being well served, and neither are our programs or our organization. The stakeholders of Extension programs need to be partners in the decision process following a program evaluation. That is the basis for their need to think *evaluatively* about what has been learned and what actions are suggested by the evidence.

**5. Recognize and communicate that decisions about best practices with regard to measurement are often not clear-cut. Make thoughtful trade-offs between the need for high data quality and the availability of resources.**

In many contexts, trying to decide how best to ensure quality can be an uncertain, ambiguous, and unpredictable process. Recommendations from the research literature or lessons from practitioner experience may be inconclusive. We often need to make decisions based on partial information. In these cases, decisions must be based on knowledge of the relevant research literature, our understanding of the target population, the lessons learned from our past practices, and the kinds of data quality limitations that we are most willing to tolerate. Well-informed evaluation stakeholders will be more willing to accept and be supportive of limitations in measurement rigor if the evaluator can demonstrate that the choices made were reasonable and logical in light of the reality of available resources.

**6. Consider using common measures for the evaluation of programs with multiple sites.**

The availability and use of common measures can offer significant benefits for the data quality and credibility of multi-site program evaluations, *if* there are basic commonalities and emphases across the sites that justify the standardization of measurement processes. Common measures can allow for broad statements about overall program impact using data that are aggregated across sites, and can also allow for direct comparisons between those sites. However, if there is significant variation in the goals and/or the programming content of those local sites, the use of common measures may not be appropriate and may present difficulties for understanding program outcomes.

### **Conclusion**

In this paper, I have examined factors that contribute to the credibility of measurement in evaluation. The measurements, design, and implementation of an evaluation are largely under the control of the evaluator, but credibility also depends on the mindset, understanding, and sophistication of stakeholder audiences. Credibility has sometimes been discussed as an objective phenomenon, a stable attribute of a body of evidence, without consideration of the variability in people's perspectives. However, since credibility—or “believability”—implies that a judgment is being made by one or more audiences, it is a characteristic that exists, at least in part, in the eye of the beholder. Program stakeholders, including staff, administrators, funders, planners, professional associates, and participants, will play a critical role in determining how an evaluation gets interpreted and how it influences the fate of the program. In sum, the credibility of evaluation evidence depends on the intersection of how evidence has been created and the responses of the humans who determine what that evidence means.

## References

- Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. New York, NY: Guilford.
- Banta, T. W. (Ed.). (2004). *Hallmarks of effective outcomes assessment*. San Francisco, CA: Jossey-Bass.
- Bazeley, P. (2017). *Integrating analyses in mixed methods research*. Thousand Oaks, CA: Sage.
- Braverman, M. T. (1996). Sources of survey error: Implications for evaluation studies. In M. T. Braverman & J. K. Slater (Eds.), *Advances in survey research. New Directions for Evaluation*, 70, 17–28. doi:10.1002/ev.1032
- Braverman, M. T. (2013). Negotiating measurement: Methodological and interpersonal considerations in the choice and interpretation of instruments. *American Journal of Evaluation*, 34(1), 99–114. doi:10.1177/1098214012460565
- Braverman, M. T., & Arnold, M. E. (2008). An evaluator's balancing act: Making decisions about methodological rigor. In M. T. Braverman, M. Engle, M. E. Arnold, & R. Rennekamp (Eds.), *Program evaluation in a complex organizational system: Lessons from Cooperative Extension. New Directions for Evaluation*, 120, 71–86. doi:10.1002/ev.277
- Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation*, 36(3), 375–388. doi:10.1177/1098214015581706
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5<sup>th</sup> ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4<sup>th</sup> ed.). Thousand Oaks, CA: Sage.
- CYFAR Professional Development and Technical Assistance Center. (2018). *CYFAR annual report 2017: Promoting the well-being of Children, Youth and Families At-Risk*. St. Paul, MN: University of Minnesota. Retrieved from <https://cyfar.org/resource/2017-cyfar-annual-report>
- CYFAR Professional Development and Technical Assistance Center. (2019). *CYFAR approved common measures*. St Paul, MN: University of Minnesota. Retrieved from [https://cyfar.org/ilm\\_common\\_measures](https://cyfar.org/ilm_common_measures)
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4<sup>th</sup> ed.). Hoboken, NJ: Wiley.
- Donaldson, S. I. (2015). Examining the backbone of contemporary evaluation practice: Credible and actionable evidence. 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. 3–26). Thousand Oaks, CA: Sage.
- Donaldson, S. I., Christie, C. A., & Mark, M. M. (Eds.), *Credible and actionable evidence: The foundation for rigorous and influential evaluations* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage.

- Goodrick, D., & Rogers, P. J. (2015). Qualitative data analysis. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (4<sup>th</sup> ed., pp. 561–595). Hoboken, NJ: Wiley.
- Groves, R. M. (1989). *Survey errors and survey costs*. Hoboken, NJ: John Wiley & Sons.
- Höglinger, M., Jann, B., & Diekmann, A. (2016). Sensitive questions in online surveys: An experimental evaluation of different implementations of the randomized response technique and the crosswise model. *Survey Research Methods*, *10*(3), 171–187. doi:10.18148/srm/2016.v10i3.6703
- Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. In M. T. Braverman & J. K. Slater (Eds.), *Advances in survey research. New Directions for Evaluation*, *70*, 29–44. doi:10.1002/ev.1033
- Lewis, K. M., Horrillo, S. J., Widaman, K., Worker, S. M., & Trzesniewski, K. (2015). National 4-H common measures: Initial evaluation from California 4-H. *Journal of Extension*, *53*(2), Article 2RIB3. Retrieved from <https://joe.org/joe/2015april/rb3.php>
- Lohse, B., Wall, D., & Gromis, J. (2011). Intention to consume fruits and vegetables is not a proxy for intake in low-income women from Pennsylvania. *Journal of Extension*, *49*(5), Article 5FEA5. Retrieved from <https://joe.org/joe/2011october/a5.php>
- 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.
- Mertens, D. M. (2018). *Mixed methods design in evaluation*. Los Angeles, CA: Sage.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3<sup>rd</sup> ed., pp. 13–104). Washington, DC: American Council on Education.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3<sup>rd</sup> ed.). Thousand Oaks, CA: Sage.
- Miller, R. L. (2015). How people judge the credibility of information: Lessons for evaluation from cognitive and information sciences. 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. 39–61). Thousand Oaks, CA: Sage. doi:10.4135/9781483385839.n4
- National 4-H Council. (2019). *Common measures*. Retrieved from <https://4-h.org/professionals/common-measures/>
- Patton, M. Q. (2018). A historical perspective on the evolution of evaluative thinking. In A. T. Vo & T. Archibald (Eds.), *Evaluative thinking. New Directions for Evaluation*, *158*, 11–28. doi:10.1002/ev.20325
- Payne, P. B., & McDonald, D. A. (2012). Using common evaluation instruments across multi-state community programs: A pilot study. *Journal of Extension*, *50*(4), Article 4RIB2. Retrieved from <https://joe.org/joe/2012august/rb2.php>

- Payne, P. B., & McDonald, D. A. (2015). Common evaluation tools across multi-state programs: A study of parenting education and youth engagement programs in Children, Youth, and Families At-Risk. *Journal of Extension*, 53(3), Article 3FEA5. Retrieved from <https://joe.org/joe/2015june/a5.php>
- Perinelli, E., & Gremigni, P. (2016). Use of social desirability scales in clinical psychology: A systematic review. *Journal of Clinical Psychology*, 72(6), 534–551. doi:10.1002/jclp.22284
- Schwandt, T. A. (2015). *Evaluation foundations revisited: Cultivating a life of the mind for practice*. Stanford, CA: Stanford University Press.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.
- Vo, A. T., & Archibald, T. (Eds.). (2018). *Evaluative thinking. New Directions for Evaluation*, 158. doi:10.1002/ev.20317

*Marc T. Braverman* is a Professor in the School of Social and Behavioral Health Sciences and an Extension Specialist in the Family and Community Health program at Oregon State University. His research interests include evaluation theory, measurement, applied research methods, community program development, tobacco control policy, and adolescent health.

### Acknowledgments

I am grateful to Ben Silliman for his suggestion of an earlier version of the model that appears on page 2. I also thank my Oregon State University colleagues Kathy Gunter and Shauna Tominey for their very helpful suggestions about sections of Table 1.



## Credible and Actionable Evidence Across Extension Program Areas: A Case Example

**Mary Marczak**

**Emily H. Becher**

**Patricia Olson**

*Extension Center for Family Development*

*University of Minnesota, Twin Cities*

*What is credible and actionable evidence can vary by program discipline in the Cooperative Extension Service. Through a series of interviews with program leadership from Agriculture, Natural Resources, Youth Development, Family Development, and Community Vitality, a case study of one state's Extension system is described. While programs certainly vary in epistemological, ontological, and methodological underpinnings, each collects and delivers credible and actionable evidence in ways unique to their stakeholder audience. Diverse Extension programs do share a common mission and a common delivery mechanism (i.e., each works to promote knowledge gain, skill development, and behavior change in people). Therefore, the future of common credible and actionable evidence in Extension may rest on investing in systems that promote a shared science of delivery, engagement, and continuous improvement.*

*Keywords:* Extension history, credible and actionable evidence, interdisciplinary, key-informant interviews

***“Even though there are no ways of knowing for sure, there are ways of knowing for pretty sure.”***

**—Lemony Snicket (1999)**

### Situation

In academia, what is considered credible and actionable evidence is largely discipline-based (Scriven, 2015). However, the Cooperative Extension Service (*Extension*), with its cross-disciplinary content and program areas, requires a broader way of understanding credible and actionable evidence. Add in the fact that Extension is a blending of content (research, science, academic knowledge) and process (how we teach or disseminate the knowledge for application in everyday life) and the notion of credible and actionable evidence grows even more unclear and more complex.

---

Direct correspondence to Mary Marczak at marcz001@umn.edu

This article intends to explore the concepts of credible and actionable evidence using University of Minnesota Extension as a case example to illustrate the complex and multifaceted nature of what makes evidence relevant and consequential in Extension work. University of Minnesota Extension's research and outreach is organized broadly into four content *centers*: Agriculture, Food and Natural Resources (AFNR); Community Vitality (CV); Family Development (FD); and 4-H Youth Development (YD). Beyond content, these centers also differ in terms of primary audiences for outreach, educational processes, and funding sources. Yet as the centers share the same mission and promise to discover science-based solutions, deliver practical education and engage citizens to build a better future, they also have commonalities and experience similar challenges in building credible and actionable evidence.

We have a relatively unique situation in University of Minnesota Extension. In 2006, the Dean of Extension requested and received ongoing funding from the University to embed Evaluation and Program Development Specialists in each center/program area. These evaluation specialists then leveraged grants and other resources to organize teams of campus faculty, state, regional and local Extension educators on the promotion track, graduate research assistants and research associates to build a research base for each center to better understand our practice and to study the impacts of Extension programs. These evaluation specialists across the centers meet with the Senior Associate Dean of Extension monthly to discuss evidence and the public value of Extension. We believe this history and context can offer a unique cross-disciplinary perspective when discussing credible and actionable evidence in Extension.

To broaden our perspectives beyond Evaluation specialists, the authors interviewed individual program leaders across centers to learn more about the programmatic aspects related to credible and actionable evidence. In cases where there are two core program areas (e.g., Agriculture and Natural Resources), we interviewed program leaders from each content area. Additionally, if a program leader felt they wanted a perspective from another key person, we accommodated. In total, seven key informants were interviewed. The interviews were semi-structured, informal, and conversational, and are used to illustrate and describe versus to generate empirical and generalizable data. The key informants included program leaders and faculty specialists across program areas. Though the one- to two-hour conversations were informal, the questions we asked were organized around the following foci:

**Ontological:** What is the nature of what you do and how did your program come to be in Extension? What is the primary contribution of your program area in terms of societal betterment?

**Epistemological:** How do you know what you do works? What knowledge is deemed consequential and relevant? What theories or frameworks are foundational?

**Methodological:** What are your thoughts about what may count as credible and actionable evidence for your program? What evidence is determined to be trustworthy,

relevant, and consequential? How does your program gather data to show evidence? Who do you have to convince?

All interviews were conducted by one or two of the authors (depending on schedule) and were recorded and reviewed to add depth to the article. Our ongoing conversations as well as perspectives of the key informants were organized into core discussion areas to consider when exploring credible, actionable evidence in Extension. The core discussion areas will follow a brief historical overview of Extension programs to add further context. We hope the unique vantage point of this case example can contribute to the overall conversation offered through this special issue of the *Journal of Human Sciences and Extension*.

### **Historical Antecedents of Programs**

A brief historical overview of Extension programs is pertinent because “what makes us Extension and how we came to be” has not only endured the test of time but continues to add critical context for program development, delivery, and evidence building. In describing this history, it becomes clear how credibility and evidence differ, in part, because of each Extension discipline’s historical roots and differences among primary stakeholders.

With the Morrill Act of 1862, the initial land-grant mission was to make university-discovered technological and science-based knowledge accessible to all people, and not just the aristocratic class. During the late 1800s and early 1900s, most land-grant universities developed a series of farmers’ institutes, where farmers could avail themselves of university knowledge without attending a university. The scope of the farmers’ institutes quickly grew in response to demands to include content in which women and youth were also interested, often because the farmers were, at times, unavailable to attend themselves. 4-H clubs started to emerge in New York and Ohio, dedicated to the practical instruction of youth in skills needed for their work and home life (Seevers & Graham, 2012). Participation by women, and inclusion of topics on issues important to women, also started to grow in Extension during this time, in conjunction with the growing science of home economics and food safety. Extension home demonstration agents taught cooking skills, sewing, nutrition, and various useful skills like home sanitation, financial resource management, and even carpentry, including the construction of ice refrigerators and kitchen cabinets (Seevers & Graham, 2012).

Meanwhile, the science of agriculture grew along with the demand from farmers for access to the latest technology and scientific developments. These demands outpaced the supply of and availability of university professors and scientists so, in response, the land-grant university system created a formal arm dedicated to conducting outreach education with the passage of the 1914 Smith-Lever Act (McDowell, 2003), which officially created the Cooperative Extension Service. What is important to note is that both home economics and agriculture were grounded in a parallel scientific discipline. The goal was to increase variables with highly measurable metrics, including reduced illness from foodborne pathogens and increased crop yields. Youth

development efforts were slightly different and were focused on providing experiential opportunities for youth skill-building and leadership. The metric, in that case, was the experience offered versus a measured output. The credibility of 4-H and Youth Development is firmly grounded in the positive experiences that youth have and the adults who observe and appreciate the value-added by those experiences. This is an important distinction with ontological, epistemological, and methodological implications even today.

Throughout the 1900s, Extension also played a role in community development and organizing, particularly in rural areas. The great depression in the 1930s deeply impacted how Extension worked in rural communities because people were less interested in production and more interested in financial survival. Due to lack of funding for professional positions, Extension turned to local rural volunteers to help deliver education and outreach. As a result, a tradition of Extension promotion of rural leadership development was born. During the Great Depression, Extension worked in rural communities to organize farm cooperatives and generate forums on economics, encouraging local leadership and civic engagement. “Cooperative Extension became the single federal agency having a direct educational link with rural America” (Seevers & Graham, 2012, p. 37). This special relationship between Extension and community leadership development, particularly in rural areas continues today in the area of Community Vitality and Community Resource Development. This role was formalized in 1993 when the USDA “declared community development to be a priority for programming emphasis” (Seevers & Graham, 2012, p. 92). In Minnesota, while Community Vitality does have an Extension educator in the Minneapolis-St. Paul seven-county metro area, historically as an Extension center it has focused on encouraging the leadership, vibrancy, and economics of small towns in rural Minnesota. Community Vitality’s credibility has historically been rooted in the value that small towns find in their partnerships with and quality of education offered by the Community Vitality Center.

While natural resources, such as soil, water, air, have always been connected to agricultural work, the 1924 Federal Clarke-McNary Act provided funding for the first forestry Extension program in Minnesota. The funding was managed by the U.S. Forest Service, and the focus of the initial work was establishing farm windbreaks, which involved planting trees around farms to protect farmland from erosion. The 1970s saw a large increase in environmental programming in Extension, expanding much of the work previously done around integrated pest management and soil conservation (Abraham, 1986). The area of natural resources as a separate content area to agriculture grew with a renewed focus on the environment and the role of individual decision-making to promote and protect soil and water quality. While natural resources’ ontology is different from traditional agricultural science, the epistemologies and methodologies of the two content areas are often quite similar in that, as participants learn new information, they change their behavior, and that knowledge gain and behavior change can be measured.

This ontological, historical backdrop provides critical context for exploring the epistemological and methodological underpinnings of Extension content areas and their relevance to credible evidence. Several characteristics have long been core to Extension. These include converting science-based knowledge to be accessible to everyday citizens; teaching practical, useful skills that matter to learners; and engaging citizens to be change agents in enhancing their own quality of life. However, these core features are the very things that pose the greatest challenges in meeting the bar set by the university-based academy for rigorous, valid evidence. When working in communities, it is a challenge to incorporate precision, to maintain simple and explicit objectives, to stay time-bound, and to adhere to structured and systematic processes, all of which are critical in pursuing credible and actionable evidence from a traditional academic perspective. In light of this challenge, the following represent the core discussion areas underscored by key informants when discussing Extension and credible, actionable evidence.

### **Key Informant Interviews Core Discussion Areas**

#### **Content (the What) and the Process (the How) of Extension Education Across Programs**

A subtle, but critical difference directly connected to credible and actionable evidence was where program areas stand on the continuum of evidence from content (the “what”) to process (the “how”). Informants across program areas independently articulated credible evidence at the core of their work. The Agricultural informants stated that their credible evidence focus was on the “what” (i.e., the specific research and scientific evidence and knowledge about their diverse topic areas).

On the other hand, a Community Vitality informant led our conversation with a statement, “Our center cares about the ‘how,’ not the ‘what.’” For Community Vitality, the “how” is mobilizing diverse, engaged members of the community through a well-designed, research-based process to facilitate the development and implementation of a shared solution. The Community Vitality program participants themselves work through a collective process to identify the issue of importance (the “what”) for the community.

Similarly, 4-H Youth Development also directs its credible evidence energy towards the “how.” The primary focus for 4-H Youth Development Extension educators is developing young people and, as a result, their research focus is on the effectiveness of the learning setting to develop youth leadership, citizenship, and mastery. The positive youth development philosophy of *youth voice* means the youth choose what to learn. While 4-H relies on existing research across relevant discipline areas, such as science, technology, engineering, and math (STEM), to support youth learning, a 4-H Youth Development informant noted that the exact science of the content is less important than the positive development of young people.

The results from the key informant interviews indicated that Family Development and Natural Resources and Environment were similar in that their evidence needs lie in the critical

intersection between the “what” and the “how.” In essence, the educational process in these areas is directly tied to the content being taught (e.g., nutrition, parenting in Family Development, or invasive species detection, forest restoration in Natural Resources and Environment). There is equal weight given to both content and process relative to establishing credible, actionable evidence. Post-hoc exploration of why these two program areas shared this feature unearthed one common factor that set these programs apart from other program areas, the fact that both Family Development and Natural Resources and Environment are largely grant funded. It should not come as a surprise that the credible and actionable evidence needed to convince funders is qualitatively different from evidence needs considered essential by other constituents.

Interestingly, where the program areas stand on the continuum of evidence from content to process was directly tied to their discussions of challenges and shortcomings. For example, an Agriculture informant noted that: “In our program area, our strength is our research base. Our challenge is the educational methodology. How do people learn?” A 4-H Youth Development informant noted:

What we don't do well, or what we don't have at times is a solid scientific base for our work, and sometimes that makes it difficult for us to obtain grants from places like NSF that require certain requirements for evidence.

### **Connection to Tenured Faculty**

The critical role that tenured faculty with Extension appointments play in building credible and actionable evidence cannot be overstated, based on the feedback from the key informants. In fact, they named no fewer than fifteen campus departments housing tenure-track faculty with Extension appointments. The role of Extension-appointed faculty members is to contribute discipline-based knowledge and research to support Extension work. In some cases, the “how” to deliver research-based knowledge to clientele and communities is delegated to county-level Extension educators. In other cases, campus faculty and Extension educators work hand-in-hand to deliver the education and study its impact. There was general agreement across the key informants that some of the best efforts they have seen in Extension are when campus faculty and Extension educators are able to work closely together on all aspects of program development and dissemination, from research to practice to evaluation.

Whether there is causality cannot be determined by this study, but what is clearly evident is the direct correlation between the level of academic faculty support and the program's position on the continuum between content and process. It is not surprising then that Agriculture leads the programs in terms of their direct connection to campus specialists. Family Development and Natural Resources and Environment followed with faculty-led studies that build evidence of specific educational curricula or partnering with Extension educators on research grants. Community Vitality, the newest Extension unit, started with strategic connections to two

academic departments. 4-H Youth Development was the only program area that has not had a departmental home or tenured faculty connected to their program. They do have loose ties to the Youth Studies Program within the School of Social Work but do not have a formal connection at the time of this article.

In Minnesota, Community Vitality's origin story offers a clear example of the important relationship between the credible and actionable evidence built by campus faculty and subsequent Extension programming. As noted in the earlier historical overview, community development has always been important to Extension. However, it was not until 2001 when University of Minnesota Extension formalized the work around community development. The Center for Community Vitality was established to put structures around what had until then largely been a philosophical or value agreement in Extension that "yes, we are all for community development!" Through a strategic planning process, the formalized Center for Community Vitality agreed on offering Extension programming around three core content areas, including community leadership, community economics, and diversity and inclusion. Prior to 2001, Extension provided some funding to two academic departments, the University of Minnesota School of Public Affairs and the Department of Applied Economics, to develop a research base and educational strategies for Extension community development efforts. The School of Public Affairs had developed a robust, research-based community leadership development model and public participation processes. The early efforts for the Center for Community Vitality focused on taking these complex models and processes developed by campus faculty and applying the knowledge and processes in the communities. It was becoming clear during this critical time that diversity and inclusion could not be sustained as part of their program area. A Community Vitality informant noted, "We realized that diversity and inclusion was a passion area and, at the time, it did not have a solid grounding in education and scholarship. It was difficult, but we learned that you can't run a sustainable Extension program on passion."

### **Realities of Extension Work**

Scientific rigor in building credible evidence requires precision, ideal conditions, and holding key study constructs constant. At times, balancing and maintaining these priorities is a challenge for Extension. As a Family Development key informant stated: "Evidence we build becomes obsolete almost as soon as we build it. People change. Conditions change. The world changes. The context of how families carry out their lives and their realities are constantly shifting."

An Agriculture informant further noted:

In Extension, while we need to bring the science from natural sciences to bear on significant issues important to the ag economy and the environment, we also have to be ready to respond to critical, real-time issues. Hail or natural disasters, for example. Farmers need to know what to do now. We can't sit and wait on the research. The pressure on research is a deeper level of science, but it doesn't allow you to address more

of the immediate needs. In those times, the Extension educator plays a significant role. The local educator has to be credible to address real-time issues. That is the reality of the situation.

Research also has to be translated to people who are not scientists. The challenge becomes controlling the message and ensuring that the Extension educational message is interpreted and used in a way it is intended. The informants discussed the reality of Extension work in the context of competing messages and in the midst of an information age where there is a “great amount of misinformation in the general public.” As one informant noted:

We peddle in education. And getting through to someone with a concept so that it may change their lives and their environment is the core of our work. But we are competing in our messaging with giants in the industries and also people’s prior knowledge and experiences.

These realities mean Extension credibility lies with the educators. The key informants emphasized that to be credible does not always mean research, but it does always mean that the messenger must be believable and trusted. Characteristics that informants noted were critical included, first and foremost, being confident with the content, but additionally being dynamic, good with people, fun, positive, optimistic, intelligent, self-motivated and a good team player. As one informant noted:

You need a unique personality to be in Extension. You need to be able to take complex concepts, make it meaningful to citizens, and apply the concepts uniquely to differing situations and contexts. And you have to be good with people, so they trust you. Only certain people are good at that. If this is their niche, then they stay.

However, above it all, there is a sense among the key informants that Extension education has long-term consequences and that the credibility of both the science and the educator, and the quality of the interaction between these two constructs, matter a great deal in the community. One informant summed up this sobering reality for all our program areas: “We have to live with the consequences of education for a long time. We need to be accurate. It is people’s livelihood.”

### **Who Does Extension Need to Convince Regarding Credible and Actionable Evidence? What Data Convince?**

Given the above discussion by key informants, it should not come as a surprise that there was general agreement across the program areas that the most important constituents Extension must convince regarding credible and actionable evidence are our own staff and volunteers. Convincing our staff of the research evidence of our content as well as specific strategies for



effectively disseminating the knowledge to make a difference leads to effective Extension education as well as staff retention. As one informant noted:

Our staff and volunteers, to continue to live out our program, must come to realize, accept, or believe that we are truly making a difference. Really good Extension work needs heart and passion and that collectively, we make a difference. Ultimately, that is [those] who stay in Extension, those who believe that.

Another informant noted, “Our staff have to believe the evidence. This belief and the skill set to take the research base to form recommendations and figure out how the information can be tailored to help under unique situations of participants.”

Another important set of constituents who need to be convinced that our evidence is relevant and consequential are participants of our education and programs. There was substantial agreement across informants that the evidence that sways our participants are things they can see and hear in real time. In Agriculture, it might be test plots that show improvements in crop growth. In 4-H, it might be a shy, young person presenting in public at the fair or leading a meeting of their peers. In Family Development, it may be a parent sharing with other parents about the benefits of communicating differently with their teenage children. No matter the program area, key informants noted that the evidence that convinces program participants are things people can see and experience for themselves.

A relevant example of offering real-time evidence is 4-H Youth Development. Most informants, including a Youth Development informant, agreed that Youth Development has the least foundation in a research base and the least amount of what might be considered credible evidence by the university academy. However, other program area leaders also acknowledged that, when push comes to shove, “4-H wins - always!” Why is that? The evidence of 4-H Youth Development’s impact is immediate. Key constituents, including youth, their parents, and community members, get to see the value every time they connect with the program. Youth are connected to the program over a longer time span, sometimes over years, so they personally see growth, change, and building of evidence over time and the general feeling of goodwill towards the 4-H program lingers into adulthood. So, for 4-H, its greatest evidence of effectiveness is retention. The youth continue to come, parents continue to support, and community members continue to invest in the program as volunteers and donors.

Interestingly, Community Vitality, the other center that focuses evidence on the “how” also understood that long-term solutions require long-term engagement. Yet they noted their greatest challenge is getting communities to buy into long-term engagement, even though that is what is often needed to show evidence of effectiveness and impact.

Given that, for our program constituents, credible evidence lies where they can observe change, this reality changes how we might go about gathering relevant, consequential evidence. A faculty specialist and Family Development informant discussed the journey she has taken:

My experience in our field about credible evidence and what actually is evidence has shifted to a focus on where we can observe change. I used to focus on quantitative data collection, maybe through pre-post assessments. But I found when we can actually observe change, for example, how people are communicating differently, a specific change along a trajectory, that parents can tell stories to each other, and they are learning how others are changing ... , I am understanding that is convincing evidence. Now I am doing this (gathering stories) that I wasn't doing before. Stories of people are credible evidence. We cannot shortchange this.

Finally, no conversation about who needs to be convinced about the credibility of Extension's programs can be complete without discussing funders. As state- and federally-funded institutions, maintaining Extension's credibility with its participants and the legislative bodies that represent them, is certainly a key feature of Extension's longevity. However, programs funded by grants often have their own unique set of required metrics in order to be determined as credible and worthy of funding. For example, the Center for Family Development receives two major nutrition education grants, the Expanded Food and Nutrition Education Program (EFNEP) and the Supplemental Nutrition Assistance Program - Education (SNAP-ED). Both of these programs require detailed accountability and monitoring systems that include documenting whether participants are achieving targeted behavioral outcomes. These grants allow for the development of robust evaluation systems, setting them apart from what a non-grant funded program's evaluation might look like. It might be surprising to see what evidence might emerge if similar specific funding was set aside at the state and federal levels for evaluation of other Extension programs.

### **Reasonableness in Building Credible Evidence**

A rhetorical, yet critical question was brought up by several informants: "When is enough really enough when it comes to credible and actionable evidence?" Extension has built and honed effective educational and programmatic practices over the years through a wide range of programmatic contexts. Extension educators are skilled at tailoring education to meet the changing needs of learners. Through reflective practice and critical thinking opportunities, educators share their best practices and strategies with other colleagues and community partners. The question being asked is: "Can we reasonably proceed in programming under the precedent that if educators do their jobs well, that expected results can be assumed?". For example, a Community Vitality informant noted:

We know some things. Our model of bringing interested people together to solve issues they are interested in is inherently good. Plus, we have 15 years of evidence that we

increased the degree of participants' leadership. We know this. Couldn't we stop that and spend time studying something more meaningful?

In Natural Resources and Environment, an informant discussed how the Kaplan Reasonable Person Model (Kaplan & Kaplan, 2003) undergirds their work with volunteer naturalists. The Kaplan model assumes that people are more reasonable, cooperative, helpful, and satisfied when the environment supports their basic information needs. So, a lot of up-front work is surveying the audience, empathizing with participants about how and what they want to learn (the Reasonable Person Model). In terms of volunteers, there is a belief that as a reasonable person, if they get certain knowledge and skills they deem important to them, they will take that knowledge and act reasonably to do good. However, the informant did acknowledge:

We want to work more closely with the kinds of metrics that actually convert to impact on the ecosystem. Also, if we can assign volunteers in surgically meaningful positions, with a strong designated volunteer model, and using volunteers more effectively, we may save regulatory dollars and may be more impactful.

Another informant summed up the "when is enough, enough" dilemma this way:

There are lots of moving parts in our work, and we often work in contexts without clear answers or credible processes. We can't be rigorous and credible with everything. So again, at the end of the day, educators have to be credible.

### **The Role of Grants in Building Credible and Actionable Evidence**

There was agreement across program areas that grants are instrumental for building credible and actionable evidence in Extension. Lessons learned as well as products and strategies that get developed through grants are often transferred to Extension's everyday work. One program leader discussed how he/she improved the gathering of credible and actionable evidence through a grant:

It is easy to see the need for understanding about program impacts. It is hard to sell understanding for program improvements. So, we went after grants to study that. This one grant brought in cool evaluation specialists as consultants and thought leaders to help us think about how what we do matters. They helped us get theories into tools to evaluate and developing a data system to quickly track complex efforts. Now, we can describe what we do, our impacts, and most importantly, why it works. It is really surprising how credible evidence supports credibility of our efforts. People buy in, including staff, as they begin to see the benefit.

Other informants also discussed how they use grants to explore evidence where they experience challenges. For example, an Agriculture informant, recognizing that the agriculture program is relatively weak on understanding educational methodology, obtained a grant to work with a

multidisciplinary team to address it. A 4-H Youth Development informant also discussed how that program is able to capitalize on opportunities to study and leverage practice through grants like CYFAR. In Family Development, grants are the gateways to studying educational processes and impacts for new and diverse audiences, testing different educational models, and building data collection systems that can contribute to their ability to build, use and publish around credible and actionable evidence.

### **Lessons Learned Across Extension Program Areas and Implications for Extension**

When we started this project, we expected to find some common type or metric of credible or actionable evidence that some program areas were better at collecting and, potentially, some areas of growth for particular program areas. Instead, what we found is that each program area is thriving in terms of the credible and actionable evidence that is valued by their unique stakeholders. Ranging from youth development, with generations of 4-H volunteers and participants experiencing and observing programmatic impacts every day, to farmers actually seeing their income grow through innovative farming strategies learned from a local Extension educator. Each program area has data, stories, and a research base satisfying their unique group of stakeholders. Evidence, research, and the scientific method matters, and each discipline is informed by emergent data from affiliated academic fields. However, across centers, science looks and is enacted differently. What is common across centers is a need to translate science into practice and communicating with stakeholders. Thus, sharing evidence as well as strategies for gathering evidence relative to effective educational practices across program areas may be a useful endeavor for building credible and actionable evidence in Extension. As alluded to earlier in the article, Minnesota Extension built a mechanism for sharing across centers with ongoing meetings of center-based evaluation and program development specialists. Based on these experiences as well as findings from the informal interviews with program leaders, we offer the following recommendations for building credible and actionable evidence in Extension.

### **Resources for Building Credible and Actionable Evidence**

As noted by the key informants, despite its diverse content areas, one thing common across Extension is the fact that Extension professionals work with people. Therefore, a common area that all program areas could explore is related to the practice of our work. What we often lack is a complex, dynamic system that collectively supports documenting and evaluating how we work with people and what strategies are effective and ineffective, documenting process evidence across programs, and explicitly supporting ways to share innovation and evidence with each other in real time. This includes exploring and understanding how research and practice evidence is interpreted by stakeholders. There is a need for common ways of documenting and sharing the practice of our work with community stakeholders in order to develop an evidence-base of effective strategies. As educators deliver programming to new audiences and within different contexts, the need for resources to build credible and actionable evidence on how to do that work effectively becomes ever more critical.

Additionally, across all programs, there is a need for a clear understanding and capacity for strategic program development and design that allow for real-time decision-making. Using good evaluation practice and data-informed decision-making when designing programs is critical. The Minnesota Extension SNAP-Ed program offers a good example of what is possible when an investment is made in a robust data system. SNAP-Ed grant funds allowed for the development of an evaluation and program management system that is utilized to continuously improve practice in real time. The management system is a relational data system with a capability to connect data across participants, geography, educator, agency partners, curriculum, and time. This system can answer simple questions like:

- Which educators have used a specific curriculum over the past three years?
- How many schools in a region of the state are part of SNAP-Ed?
- Which partners are working together in a community on a specific project?
- What proportion of participants has taken more than one course from SNAP-Ed?
- What are the most popular improvement oriented actions taken by multi-sector collaborations?

Additionally, the system can answer complex questions such as: “What are the interaction effects of multiple programmatic strategies (such as a combination of direct education and environmental changes) on participant behavior change outcomes?”.

The SNAP-Ed data can also be connected to large-scale census or public health data, so programming needs, staffing, and whether a specific curriculum (or interactions of program components) works across geography or participant characteristics to promote program goals and outcomes can be determined in real time. Needless to say, Extension as a whole would be well-served with the resources to build such a system that can be effectively and consistently applied across program areas.

### **Staff Development**

Perhaps the biggest takeaway is that the program leaders in this study believe that credible evidence resides with our staff rather than with data. To meet emergent needs, to address complexity, to make sound judgments in real-time amidst rapidly changing environments, our staff have to be knowledgeable and confident about the evidence behind both the content and the process. It follows then that Extension educators must have a continuous relationship with credible research and a deep understanding of the research behind their messaging. Extension educators must be able to answer the question of what is cutting edge in terms of their content and process. To act as effective interpreters, they need a regular connection to scholars exploring the content of interest as well as a willingness and an interest in continuously learning about what the research community is producing. Extension educators need support to craft and test out their messaging, to share what they learned, and to continuously hone their skills.

## Needs Assessment

In all the focus on credible and actionable evidence, and in terms of data needs, the program leader informants expressed that what is most critical may be more robust needs assessments to understand the learning needs and contexts of our Extension program participants. As two program leader informants shared:

At the end of the day, we want to connect with people. What will make people connect with us? So how do we learn what they value and what they see is value? We have to figure out how we contribute to their lives. Do we have something to offer? Could we contribute to meet their goals?

Our bread and butter is people. We have to understand the learning wishes of our priority audiences. Then we need to feed that knowledge into our planning process. . . . Then figure out within the resources we have right now, what can we do right now to address specific learning wishes of our audiences?

When exploring credible and actionable evidence in Extension, we should be asking: “Are we giving people something of value?” and “Are we helping them meet their needs in ways they can see and find value?”.

## Engaging Citizens to Play a Role in Crafting and Using Research Knowledge

No matter our programs’ impact areas, whether it is healthy and thriving youth, families, communities, agricultural economy or the environment, according to our program leader informants, the common denominator of Extension is people. Given the growing awareness by land-grant universities that technical solutions are not enough to solve complex problems, the University of Minnesota has advanced Grand Challenges priorities that call for reciprocal partnerships with communities to explore research and local impacts. This means we need people, both Extension program staff and program participants, to engage with research and program evidence critically and reflectively. Critical thinking, critical data consumption, and evidence-based decision-making is not just the purview of the expert, it is in everyone’s interest to build this skill set.

Explicit strategies are needed for Extension program participants to engage with research and data so they can become agents of change in their own lives. An innovative solution might lie in Minnesota Extension’s Natural Resources and Environment program area, where they have developed a platform for developing citizen scientists in the content area of aquatic species management and documenting numbers of at-risk species, such as pollinators. What would it look like to develop citizen scientists across all program areas, from leadership to soil quality to parenting to healthy food access and beyond? The next step in the land-grant mission of democratizing high-level community education may be to revisit Extension’s historic charge to

“rouse the people of the land” (Herrera & Hoelting, 2010; Peters, 2002) to organize solutions that are driven by those who are most affected by it. This can only happen if we teach and train in not just content, but also critical thinking, critical data consumption, data-informed decision-making, and direct engagement with science. If Extension embraces this as a core, common goal, then a next logical step would be to evaluate Extension programming with metrics such as changes in individual or community-level efficacy to address issues and concerns about which they care. Other potential outcomes would be increases in citizen involvement and community-led work with indicators like volunteering, coalition formation and membership, and community-led projects.

### **Limitations**

The limitations of this work include its methodology, which describes a particular Extension context and is not intended to describe broader patterns or represent empirical data. Evaluation specialists and content-based program leaders have unique perspectives that may not be representative of all Extension staff. Therefore, any conclusions, recommendations, or suggestions must be considered with this limitation in mind.

### **Conclusion**

When asking the question about what is credible and actionable evidence across Extension program areas, the answer is: That depends on the people and stakeholders Extension serves. What is common is that every Extension educator wants to have a positive impact on how people lead quality lives; however, those people determine quality. A critical step for Extension is to work towards building capacity and developing resources to create a science of delivery and implementation regardless of content area. Extension works with whole people, families, communities, and is as interconnected and interdependent as the people and communities it serves. Yet Extension’s efforts are too often siloed and restricted to specific content areas. No matter the impact of Extension programs, whether it be healthy and thriving youth, families, communities, agricultural economy or the environment, the common denominator of Extension efforts is people. Extension must democratize critical thinking and data-informed decision-making and partner with communities to study and create the solutions to problems they identify as being priorities. Ultimately, that is how Extension develops a broad base of credible and actionable evidence.

### **References**

- Abraham, R. H. (1986). *Helping people help themselves: Agricultural Extension in Minnesota, 1879-1979*. St. Paul, MN: Minnesota Extension Service, University of Minnesota.
- Herrera, M., & Hoelting, J. (2010). To establish an effective community spirit: A land grant Extension and deliberative dialogue. In N. L. Thomas (Ed.), *Educating for deliberative democracy. New Directions for Higher Education*, 152, 43–50. doi:10.1002/he.411

- Kaplan, S., & Kaplan, R. (2003). Health, supportive environments, and the reasonable person model. *American Journal of Public Health, 93*(9), 1484–1489. doi:10.2105/ajph.93.9.1484
- McDowell, G. R. (2003). Engaged universities: Lessons from the land-grant universities and Extension. *The ANNALS of the American Academy of Political and Social Science, 585*(1), 31–50. doi:10.1177/0002716202238565
- Peters, S. J. (2002). Rousing the people on the land: The roots of the educational organizing tradition in Extension work. *Journal of Extension, 40*(3), Article 3FEA1. Retrieved from <https://www.joe.org/joe/2002june/a1.php>
- Scriven, M. (2015). Demythologizing causation and evidence. 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. 115–132). Thousand Oaks, CA: Sage.
- Seevers, B., & Graham, D. (2012). *Education through Cooperative Extension* (3<sup>rd</sup> ed.). Fayetteville, AR: University of Arkansas.
- Snickett, L. (1999). *A series of unfortunate events: The bad beginning*. New York, NY: HarperCollins.

*Mary S. Marczak* is Director of Urban Family Development and Evaluation.

*Emily H. Becher* is an Applied Research and Evaluation Specialist.

*Patricia Olson* is the Associate Dean of the Center for Family Development.



## Credible and Actionable Evidence Across Stakeholder Levels of the Cooperative Extension System

**Nick T. Place**

*University of Florida – Institute of Food and Agriculture Sciences*

**Richard M. Klemme**

*Association of Public and Land Grant Universities*

**M. Ray McKinnie**

*Virginia State University*

**Carrie Baker**

**Jean Parrella**

*Texas A&M University*

**Scott R. Cummings**

*Texas A&M AgriLife Extension Service*

*This article provides a look at the various levels within the Cooperative Extension System and the use of evidence within these levels. The authors examine the factors associated with credible evidence and the various levels. The impact of factors such as politics, science, stakeholder support, and expectations are discussed. The various levels within Extension are summarized in relation to evidence that is routinely requested or required for each. Lastly, the authors use information directly from Extension directors to provide a framework for the discussion.*

*Keywords:* evaluation, stakeholders, politics, science

***“The land grant university system is being built on behalf of the people, who have invested in these public universities their hopes, their support, and their confidence.”***

**—Abraham Lincoln, upon signing the Morrill Act, July 2, 1862**

### Overview

The quote by Abraham Lincoln on the land-grant university system sets the stage for what is to follow. The state and territory members of the Cooperative Extension System (Extension), are integral components of the land-grant universities of the United States and make up a diverse and complex system with stakeholders at many levels. As a component of the land-grant university

---

Direct correspondence to Scott Cummings at [s-cummings@tamu.edu](mailto:s-cummings@tamu.edu)

system, Extension must be accountable to all stakeholders who have invested in the system, as articulated by President Lincoln. As part of this accountability obligation, Extension strives to provide credible and actionable evidence on the quality and effectiveness of Extension efforts. Credible evidence is “information that stakeholders perceive as *trustworthy* and *relevant*” (Donaldson, 2015, p. 5). Numerous factors influence what is credible evidence, including laws and policies, stakeholder and funding requirements, evaluation capacity, and understanding of the evidence itself. Actionable evidence refers to that evidence to which stakeholders can use to make decisions or modify programs or policies.

The basic framework of Extension is its three-level funding structure, including funds from the federal government through the U.S. Department of Agriculture’s National Institute of Food and Agriculture (USDA-NIFA), funds from state governments, and local funding through counties/parishes or similar local entities. In addition, in many states, the land-grant universities associated with Extension programs have added another layer of influence by emphasizing the need to acquire grants and other special projects funded from external sources, both private and public. Moreover, Extension is aiming to obtain cost recovery funds that cover the cost of conducting Extension programs. This is necessary as levels of funding from federal, state, and local governments have decreased or remained level over time. The addition of external funding of projects has added to Extension’s stakeholder list and the complexity of the system. Many of these stakeholders have varying requirements for the types of evidence that are deemed credible. For example, the federal partner requires evidence on program outcomes or impacts. Some states require data on number of contacts made within a state. Some universities focus on grants and publications. There will be a more in-depth discussion of these requirements later in this paper.

Mahon and Wartick (2003) wrote that credibility refers to an organization’s history in terms of how it develops reputational expectations, especially among its stakeholders. Therefore, credible evidence is providing something that has been validated (Mahon & Wartick, 2003), whether it be a program, product, consumer satisfaction level, an economic catalyst, or organizational compliance. This does not necessarily mean stakeholders also need evidence that is deemed credible in order to associate trust or value with an organization. For example, a long-time educator in a community may be deemed credible by the fact that they have gained a high level of trust within that community. Anything they recommend could be deemed credible regardless of whether it is based on fact or not.

However, stakeholder perceptions relating to factual or scientific information is another discussion. A credibility transaction is defined by Herbig, Milewicz, and Golden (1995) as “the firm’s comparison between a competitor’s pronouncements or intentions and its true behavior or final actions” (p. 26). They describe four types of transactions: 1) true positive - an organization says it will act and follows through by acting, 2) false positive - an organization says it will act but does not do it, 3) false negative - an organization says it will not act, but changes its mind

and does act, and 4) true negative - an organization says it will not act and follows through by not acting (Herbig et al., 1995).

The focus of this paper is to examine criteria for evidence required or requested by various Extension stakeholders and how evidence is used at these various levels for policy and funding decisions. We will explore the policies, practices, and politics that affect how evidence is used to satisfy the various stakeholder needs. We will also examine what is needed to achieve organizational credibility and the concept of credibility as a multi-dimensional construct. In addition, the authors sought feedback from various Extension directors on the topics covered in this article. Examples from this feedback will be used to support or provide a framework for the discussion.

### **Targets for Credibility**

The building of organizational credibility and trust among stakeholders must become a priority in Extension. The problem, however, is that there has been little conceptual agreement regarding what entails organizational credibility and trust in an organization (Bigley & Pearce, 1998; Young, 2006). Even more problematic is the inconsistency of methods used to assess trust and credibility, which makes generalizability difficult and could lead to errors in measurement and strategic planning decisions (Bigley & Pearce, 1998; Kazoleas & Teven, 2009). Kazoleas and Teven (2009) wrote that in order to reliably measure trust, the relationship between the organization of interest and the public must be clearly understood and must include measurements that disclose the full range of underlying factors that bring about the concept of trust. Failure to do this will result in insufficient measurement models and yield inaccurate results (Kazoleas & Teven, 2009). In addition, strategic planning based on these results will account for a large degree of variance that is not addressed in the measurement model (Kazoleas & Teven, 2009).

Trust is recognized as a “multiple faceted concept that can take on many meanings depending on the perspective from which it is viewed” (Kazoleas & Teven, p. 22). Mayer, Davis, and Schoorman (1995) identified the underlying dimensions of perceptions of trust through a comprehensive review of the literature. Trust regarding organizations can be broken down into three factors: *ability*, *benevolence*, and *integrity* (Mayer et al., 1995). Ability incorporates elements of confidence and reliability as it relates to results of transactional relationships (e.g., interaction with one or more members of an organization?) (Mayer et al., 1995). Benevolence involves the organization's intentions to meet stakeholder needs, and integrity involves accuracy and honesty (Mayer et al., 1995).

It is also important to identify different (and interrelated) targets for credibility and why they are important. While operationalizing credibility within Extension is vital, the complexity of understanding credibility cannot be overstated. Despite the challenges presented, organizations have provided helpful constructs to inform credibility within organizations. The Blandin

Foundation, a private, independent foundation focused on rural communities, developed a philanthropic theory based on applying developmental evaluation and how to establish credibility within an organization and maintain credibility based on program impact results (Annette, Fauth, & Ahcan, 2015). As a funding entity, training organization and educational resource for rural communities, Blandin recognized the importance of evaluative standards, not only for their success but for that of their clientele as well. The Blandin model was built upon prioritizing the relationships held with communities, policymakers, peers, and other stakeholders, which ensured transparency and that the organization's resources best matched the needs of partners (Annette et al., 2015). Through reflective team sessions and deep, evaluative exercises that examined grants, relationships, and foundation work that did and did not manifest relationships, a comprehensive list emerged that included ways in which the foundation could strengthen its role as a connector (Annette et al., 2015). As an organization dedicated to connecting people to networks, knowledge, issues, and resources, this model served to further Blandin's mission and broaden its reach within the communities it serves. These insights formed the basis for the "mountain of accountability" in an effort to deliver and allow replication of their strategic planning (Annette et al., 2015).

In its model, The Blandin Foundation explained the roles of three types of foundation assessments. At the base of the "mountain" sits the basic accountability for management processes, which includes information regarding financial audits and investment returns, an evaluation of human resource performance management, descriptions of basic management-information systems, due diligence, reporting, community indicators for planning, and fulfillment of donor intent and court guidance (Annette et al., 2015). Accountability for impact occupies the middle of the "mountain." This section calls for major program evaluations, an external strategic evaluation, a board survey and feedback, a grantee perception report, a synthesis of grantee's reports, and employee surveys. Finally, to achieve mission fulfillment, the peak of the "mountain" contains accountability for learning, development, and adaptation. This unit includes deep, reflective practice, developmental evaluation, strategic-framework evaluation, and a focus on systems change, innovation, and complexity (Annette et al., 2015).

The "mountain of accountability" is dependent on the ability of individuals and organizations that share a common vision to work together to develop focused, inclusive, and goal-oriented strategies. This system of evaluation ensures credibility within the organization by increasing awareness and nurturing connections among all key players (Annette et al., 2015). The "mountain" further identified the level and criteria for credibility negotiated between program leaders and stakeholders into three sections. The first section, governance philosophy, is the foundation's strategy to lead and direct work. The second section, contextual sensitivity and trend scanning, is how the foundation can ensure its work stays relevant among changing environments. The third section, strategy, is how the foundation implements resources to impact others (Annette et al., 2015).

The Blandin Theory has gained recognition through its detailed and consistent outline to achieving organizational credibility. As mentioned earlier, the many stakeholders at the different levels of Extension require varying types of evidence, adding to the complexity of the Extension system. Next, we will examine the different levels and targets for Extension impact and how these relate to credible evidence.

### **Need for Information and Strength of Evidence**

#### **Good Enough Evaluation**

What is “good enough” evaluation? This topic was addressed and discussed in the first article of this special issue of the *Journal of Human Sciences and Extension* (JHSE). The American Evaluation Association (AEA) has developed standards for evaluation and those conducting evaluation. (AEA, 2018). In some instances, good enough evaluation depends on the stakeholder, and usually refers to those results that meet expectations stated in policies or contracts. However, such expectations are not always explicit, clearly articulated (or measured), consistent with scientific or professional standards, or adequate for specific circumstances they may be intended to address. Examples of these dilemmas are included in the discussion that follows. However, some organizations might be inclined to want to go beyond the concept of good enough evaluation methods. In the first article in this issue, the concepts of integrity, transparency, and adaptability by stakeholders and those conducting an evaluation provide the basis for what evidence contributes to effective and credible evaluation. To contextualize this within Extension, it might be helpful to look at utilization-focused evaluation, which addresses the complexities of program evaluation and accounts for the involvement of multiple stakeholders and decision-makers (Patton, 2003).

According to Patton (2003), utilization-focused evaluation does not operate within the realm of possibility or idealism, instead, this evaluative methodology focuses specifically on whether or not the program effectively and efficiently addresses the needs of actual users. In this way, evaluators work closely with individuals who have the ability to apply findings and implement recommendations in a way that is most useful and relevant to the program’s intended users. Within Extension, this method of evaluation is found most often because of its vast and situational utility. Evaluators deductively develop evaluation models that fit within the intended context and use these models to best address individual program or organizational need (Patton, 2003). While flexible in nature, this approach is also bound to standards and guidelines which increase the credibility and integrity of findings (Patton, 2003). Patton stated:

As a professional, the evaluator(s) has the responsibility to act in accordance with the profession’s adopted principles of conducting systematic, data-based inquiries; performing competently; ensuring the honesty and integrity of the entire evaluation process; respecting the people involved and affected by the evaluation; and being

sensitive to the diversity of interest and values that may be related to the general and public welfare. (2003, p. 224)

In this sense, one might disregard the concept of “good enough” evaluation and instead propose that evaluation is deemed appropriate for the given context and situation in order to provide stakeholders and interested parties with the most relevant and accurate information as possible. Extension exists to serve a plethora of audiences and answers to a variety of stakeholders. Thus, tailored findings and evaluative reports for determining effectiveness is not only helpful but necessary. Patton (2003) stated that “program evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness and/or inform decisions about future programming” (p. 224). If one attaches a utilization focus to these program evaluations, the intent of these evaluations becomes not about distant or unrelated reviewers, but are instead performed with specific, intended audiences and users in mind (e.g., Does this program address the needs of intended users?). While not directly stated or recognized, Extension often uses a utilization-focused approach in program evaluation and reporting not only for credibility with stakeholders, but perhaps more importantly, for its practicality in the field. Specifically, this approach provides formative evaluation that is utilized to make program adjustments in order to ensure effectiveness. In this sense, in evaluation of Extension and development of related-programming, careful consideration is given to how “real people in the real world apply evaluation findings and experience the evaluation process” (Patton, 2003, p. 425). Utilization-focused evaluation considers how evaluation might be used to best impact intended users in the present and inform effective program development in the future. Put simply, Extension not only develops programming with their public audiences in mind, but also continues to be held accountable by these audiences during evaluation (Suvedi, Heinze, & Ruonavaara, 2005). This approach has historically satisfied stakeholders and audiences across multiple organizational levels.

So, where do stakeholders set the bar, and how do we help them place it well? Is the evidence that stakeholders are requesting science credible, or is it credible from a ‘good enough’ perspective? Based on what we know from reviewing evidence required by Extension stakeholders, the authors believe the evidence required is based on past policies or what we have always presented. Evidence may be based on factors such as reputation (e.g., credibility based on past performance or identification with the organization’s mission/people/etc.) and/or tradition (e.g., credibility based on past output or ongoing assessment of needs and performance standards). As a change organization, Extension has provided evidence that is not keeping up with the changes in society. Extension provides research-based information to our clientele. Should we not also be doing the same with regard to evidence of Extension’s value and benefit to those it serves?

Kelsey and Mariger (2002) found through conducting case studies among Extension forestry departments that stakeholders felt they did not receive adequate information from the university for their state. Barriers were discovered that existed in both oral and written communication channels. For example, the use of technical jargon was prominent, which non-science audiences struggled to understand.

### **Influence of Politics on Evidence**

Because the majority of stakeholders that influence Extension funding are political entities and individuals, politics plays a critical role in the expectations for evidence (Larner, 2004). The overarching mission of Extension is to serve clientele with unbiased, research-based content (NIFA, n.d.). While this is the standing goal, given the way funds are allocated and disbursed to certain projects and agencies, political influence is unavoidably a factor. Extension agencies receive funds from government entities and are also expected to report back to these stakeholders. Political influence varies throughout the accountability process depending on reporting procedures required and stakeholder involvement. As McDowell stated:

Extension, as the outreach arm of the land-grant university system, has a primary educational mission. However, it is also expected to collect public political support on behalf of the system, including its research activities, particularly in the case of state government support [funding] for university budgets. (1985, p.718)

While politics may not influence the evidence itself, political influence might affect the type or nature of the evidence to be reported. Increased competition for state and federal funding also creates pressure on Extension agencies to cater to the specific requirements set forth in grant and programming guidelines, outlined by policymakers and legislators, to garner their support. This also plays into how funding is utilized by Extension. There tends to be a tension between traditional vs. nontraditional work expected by clientele and funders. Therefore, the political needs for evidence are not always negative and can often be in stride with the mission of Extension as policymakers and legislators are also stakeholders. McDowell (1985) clarifies that “competition is not only experienced in the politics of the state and federal budget processes, it is also experienced by the Extension staff in terms of competition for audiences, turf, and grants and contract resources” (p. 718).

This interaction challenge is more acute in the political/budget environment. This may, in fact, be even more challenging in an increasing anti-science, anti-higher education environment. In this case, Extension has an advantage over research and formal education in that both agents/educators and specialists enjoy the opportunity to have the kind of interpersonal relationships that provide a chance to overcome the current political winds. Relationships are a key pathway for Extension, and frankly, most political institutions.

Thus, the tension between gaining financial support for the advancement of their mission and the betterment of the publics they serve, while also remaining unbiased in that mission, remains a challenge that Extension educators battle every day. As a public entity, Extension is dependent on state and federal funding in order to provide the public with education, resources, and other services. Thus, the challenge to serve the stakeholders and the public objectively will remain and should be of heightened priority (Voris, 1991). Extension professionals occupy a unique role in bringing university expertise to policymakers and building unbiased relationships with clientele to increase support. In turn, political influence is a factor Extension cannot avoid. Therefore, Extension as an organization and Extension professionals must be comfortable in and aware of their roles (Stoltz, 2002). Acting in both the political and public sectors, Stoltz (2002) stated that Extension must possess acumen in order to effectively participate and serve in both realms.

### **Science and Credibility**

What constitutes good science and good politics at different levels of the Extension system, and how are these priorities compatible? Are decisions based on evidence, or opinions and beliefs? Where do stakeholders set the bar for credible evidence in Extension and how do we help them place it well? Is the evidence that stakeholders are requesting science-credible, or is it credible from other perspectives?

According to the United Nations (n.d.), within the next 15 years, the world population is expected to increase by more than one billion people, nearing 11.2 billion by the year 2100. As the population continues to grow, the agriculture industry is working to meet the needs of consumers and better serve the public in food, fiber, and fuel. With this dynamic change, comes the ever-increasing need to not only provide tangible products but act as a hub for the intangible information regarding the agricultural system. Now reaching into areas such as policy, economics, food systems management, communications and rural development, “agriculture [has become] an information-dependent sector of the economy” (Cash, 2001). Decision-making within the agricultural industry requires an understanding of scientific and technical information that needs to be digestible, not only by those within the industry, but the publics they serve as well (Cash, 2001). In order to fill this need, Extension performs boundary work or serves as what Guston (1996) refers to as a boundary organization. Fundamentally, boundary work fuses connections between science and policy to implement effective, research-based solutions that cater to both sides of the figurative boundary line. As a boundary organization, Extension facilitates both public and political objectives and ensures the protection of scientific credibility through those actions (Guston, 1996). As Cash (2001) stated, “the system has become a partnership between federal, state, and local agencies and educational institutions, with shared responsibilities and funding” (p. 434). As a boundary organization, essentially, Extension bridges the gap between science and policy, linking the two across the different levels of



Extension (local, state and national) and then communicating the implications of that information to the public (Cash, 2001).

While the goal of disseminating research to the public is of critical importance to Extension, at the same time, the organization has “an interest in maintaining independence from the users of the information they produce” (Cash, 2001). According to Cash (2001), “the balance [Extension] seek(s) is to provide useful information but maintain scientific credibility” (p. 440). The credibility of this science often categorizes it as being either “good” or “bad;” thus the need for a distinction of what constitutes good science in regard to Extension. In such a research-driven organization with the goal of breaking down technical information, how do these organizations ensure that the science they are disseminating is deemed as “good” or credible? According to Moss and Edmonds (2005), “good science enables us to understand what we observe” (para. 2.1). As an organization that heavily leans on social science, but is foundationally grounded in the natural sciences, Extension draws evidence from both ends of the scientific spectrum. “Good social science will be in some respects different from good natural science” (Moss & Edmonds, 2005, p. 5). Moss and Edmonds (2005) also said that, “Evidence and observation have priority over theory there -- (in the end) when evidence and theory disagree the theory is changed” (p. 4).

### **Jeopardizing Stakeholder Support**

Stoltz (2002) stressed the importance of Extension administration at the federal level providing Extension faculty at the state level with reliable budget and policy information. He further emphasized the importance of Extension faculty providing accurate information about their work in their conversations with elected officials and clientele to avoid jeopardizing support for the system. When Extension professionals discuss internal conflicts with clients or elected officials, they risk the credibility of not only Extension workers, but they also risk the credibility of the entire Extension organization, in addition to destroying their support base (Stoltz, 2002). According to Stoltz (2002), internal conflicts can range from disputes in the local Extension office or be as widespread as national matters. Stoltz (2002) also wrote that, “Extension administration and field faculty need to understand that effective faculty - highly respected for the job they do - can spearhead change, promote understanding of complex and/or controversial issues, and build political support for Extension.”

In order to gain respect from stakeholders and bolster political support, Extension personnel must build working relationships with policy- and decision-makers, be informed and readily able to provide information on programs and budgets and be responsive to stakeholder concerns. Internally, this begins with the creation of effective programming, based in research and proven through measurable outcomes. According to Stoltz (2002), strong programs earn support, and in order to continue gaining support, Extension professionals are expected to simultaneously serve the public and meet stakeholder expectations through that programming.

### **Expectations, Policy, and Law**

Stevens, Lodl, Rockwell, and Burkhart-Kriesel (1994) explored the different perceptions that federal project directors and state and local level project leaders held about the youth-at-risk grant funds. The study explored not only those perceptions, but also the participants' understanding of project philosophies, goals, and future expectations. In their article, Stevens et al. (1994) attempted to identify differences in project expectations at the various levels of Extension. Data for this project were obtained from various places depending on the level being examined. At the federal and state levels, the researchers analyzed data from requests for proposals (RFPs), project applications and reports, evaluations, and some personal interviews. At the local level, data came from telephone interviews with project leaders (Stevens et al., 1994). All of the data were then cross-analyzed using open coding (Stevens et al., 1994).

Results from the study found that overall expectations at the state level included and addressed expectations at the federal level, with an increased focus on meeting statewide goals and needs (Stevens et al., 1994). Guidelines for program development were put in place at the state level to ensure those federal expectations were met and reported. These guidelines included mechanisms for meeting state goals and addressing clientele needs (Stevens et al., 1994).

In their study, Stevens et al. (1994) found that, while the federal expectations were addressed, emphasis at both the state and the local levels emphasized an increased focus on meeting clientele needs. Stevens et al. (1994) also noted that locally, "the highest priority was given to making 'real' differences in the lives of individuals and their communities" (para. 11). Extension agents and program leaders noted that they addressed specific local needs first, and then once they felt that program was successful and sustainable, they moved on to address other concerns and meet expectations put forth by federal and state entities (Stevens et al., 1994). According to Stevens et al. (1994), "while the federal expectations were global, the state level became more specific and focused on state needs." This exemplifies the tailored, utilization-focused approach of Patton (2003) with regard to evaluation and reporting and also serves to explain if and how differences in reporting exist when addressing federal, state, and local expectations.

As Stevens et al. (1994) stated, "the true challenge for Extension is to help the public understand this [Extension's] mission and how it impacts them as clientele." This aim is further complicated when fluctuations in expectations exist. Thus, priority is often given to addressing local needs first in order to meet federal and state expectations. For example, the basis for Extension work is to help people make sound decisions to improve their lives. This grassroots approach can then be aggregated to show the value of Extension at the state and federal levels. These fluctuations in expectation make Extension's evidence and evaluative reports no less credible, but instead intensifies their utility at all levels, furthering the success of programming and the Extension system as a whole.

### **Credible Evidence and the Success of Programs**

How does credible evidence impact program success with stakeholders? This is a critical question as we move past providing evidence to providing credible evidence. Issues such as trust, relationships, and communication play a role in this movement to credible evidence. We reached out to Extension directors across the country for their input on credible evidence and program success. One director responded: “Our program success is dependent upon providing credible evidence to the individuals and communities we serve.” Also mentioned were the ties Extension has with its Agricultural Experiment Station partners and Extension’s own applied research being critical in continuing to provide credible evidence/solutions. Another director added: “Credible evidence impacts program success directly by strengthening a program’s sustainability and ability to secure ongoing support from stakeholders (financial, personnel, advocacy, marketing, etc.).”

The authors also asked about how non-credible evidence affects program success. An Extension director responded, “Non-credible evidence negatively impacts overall program success. Non-credible evidence will lead to the loss of trust and being viewed as a science-based, unbiased, source of information.” Another Extension director added, “It has less of an impact—and raises questions of value of the organization.” Another director responded, “Non-credible evidence impacts program success by jeopardizing a program’s sustainability efforts and ongoing stakeholder support.”

### **Extension Stakeholder Perspectives and Expectations**

The complex nature of Extension and the differences in funding and stakeholder expectations makes meeting requirements for credible evidence difficult at best. Requirements for evidence vary greatly, based on the stakeholder and their needs. Extension organizations are faced with collecting and reporting different types of evidence to meet the needs of various stakeholders. This complexity poses several dilemmas for Extension and raises the following questions:

- Are the stakeholder’s expectations clear or achievable?
- Are their agendas transparent and consistent with project objectives or capacities?
- Which stakeholder or stakeholders get the most attention when it comes to providing evidence?
- How do politics and policy play roles in providing evidence to stakeholders?

In general, each subgroup of stakeholders sets its own expectations and timetable for evidence based on precedent, laws, policies, or other guidelines. This complexity of expectations requires Extension administrators and staff to balance expectations for evidence that, if not conflicting, may not be complementary. These differences in expectations may include different evaluation targets (e.g., organizational, program, participant behavior), level of focus (e.g., outputs, outcomes), precision of outcomes (e.g., specific vs. general changes), timing (e.g., short- vs.

long-term, frequent vs. one-time reporting), in addition to diverse indicators within and between disciplines and projects. Moreover, expectations for the type of evidence (e.g., quantitative vs. qualitative, or both) and representation of evidence (e.g., local vs. cumulative state or national data) increase the evidence complexity exponentially.

### **Federal Level Requirements and Expectations**

The federal stakeholder for Extension is the United States Department of Agriculture's National Institute of Food and Agriculture (USDA-NIFA). Requirements for program evidence at the federal level are, for the most part, consistent across Extension. Some of the evidence requested is required by the Agriculture Research, Extension, and Education Reform Act of 1998 (AREERA, 1998). This act amended the original Smith-Lever act of 1914 that first established Extension. A more detailed description of the legally mandated evidence and other required evidence is discussed below.

For USDA-NIFA, credibility and credible evidence is provided at the organizational level. Evidence is provided to demonstrate that the Extension organization is utilizing federal dollars wisely and providing evidence to support the mission of USDA-NIFA at the federal level. USDA-NIFA routinely has areas of focus that state Extension organizations provide evidence to support. These have included topics such as climate change, sustainable energy, obesity, and food safety.

Federal legislation requires Extension institutions to submit a 5-year plan of work (POW) and an annual report of accomplishments to receive federal funding. Specific requirements of AREERA (1998) include the following:

- Programmatic overview of the institution;
- Programmatic summary containing research, Extension, and integrated accomplishments;
- Overview of the scientific and merit review process;
- Description of the stakeholder input process;
- Inclusion of all multi-state and integrated components; and
- List of all planned programs.

In addition, documentation is requested within the reporting structure to provide USDA-NIFA with evidence of impact for programs delivered at the state or institutional level. These impacts, based on the planned programs included in each state's POW, are used to fulfill funding strategies and legislative requests.

Federal requirements have remained fairly consistent since the passage of the AREERA in 1998. Changes in administrations and the policies of these administrations have not altered the general requirements for evidence. However, there has been a greater emphasis over the years on the

inclusion of impact-related program objectives and evidence in the federal reporting system. Reviews of submitted reports have focused on the impact of efforts much more than on other evidence data, such as inputs (those items that go into conducting a program, such as resources) outputs (items that derive from a program such as number of participants). For example, in the reviews of federal report submissions, critiques generally focus on the presence or absence of evidence for state-defined program outcomes. These outcome data focus on learning, applications of recommended practices or behaviors, or the effects on an individual or community as a result of those changes of practices or behaviors. Statements including this impact evidence are most commonly used for evidence presented to Congress or other government entities.

In the past, leadership at USDA-NIFA have set priorities for specific interest areas (e.g., sustainable energy, climate change, childhood obesity) and facilitated the development of research bases and performance indicators consistent with Extension's mission, scientific foundations, and stakeholder needs. From the federal perspective, the focus on the specific issue priorities and evidence reported by Extension has changed, although the general criteria for and methods of reporting has not changed.

### **State Level Requirements and Expectations**

For many Extension organizations, the state legislature is also a stakeholder providing funding to Extension. The percentage of a state's total funding that is received from state legislatures varies from state to state, with some states providing a majority of their Extension's total funding to other states that contribute very small percentages to their Extension's total funding. Models also vary on how state funding is appropriated. In some cases, funding is provided to the land-grant university and then to the Extension component. In other cases, Extension funding is provided directly to Extension. For example, in Texas, the Texas A&M AgriLife Extension Service is a state agency under higher education and is a separate line item in the state budget. State funding comes directly to the Texas A&M AgriLife Extension.

At the state level, credibility and credible evidence is also at the organizational level. Evidence is provided to demonstrate that the organization is utilizing state dollars wisely. Typical evidence might include participation or reach numbers, results of programs focusing on the effectiveness of the effort, and in some cases, economic impact of the efforts. Where Extension's state budget is part of the land-grant university's structure, Extension evidence may also support the work of the university's outreach efforts to stakeholders.

In another state, the Extension director noted:

At the state level, the Extension Directors Office provides a quantitative impact report to Central Administration annually, a report on Key Progress Indicators to the Legislative

Finance Committee and Higher Education Department, and qualitative/quantitative impact documents are used during the legislative session.

### **University Level Requirements and Expectations**

Extension institutions vary on how they are connected to the land-grant university. Differences in accountability criteria and structures are typically related to Extension's funding structure. In some cases, Extension is fully integrated within the university, and funding is channeled through the university structure. Funding, reporting, and requirements for evidence come from the university leadership. In other states, Extension is part of a university system. A university system is typically an umbrella administrative structure governing several universities. Although Extension may be integrated into the university, funding, reporting, and requirements for evidence come from the university system or directly from a line-item in a state budget.

As with the state level requirements and expectations, credibility and credible evidence at the university level is at the organizational level. Required and requested evidence is provided to demonstrate that the Extension organization is using state dollars wisely. Where Extension is part of the university structure, evidence may also support the work of the university's outreach efforts to the stakeholders. More of the emphasis here is to document evidence that the public is receiving value from their land grant university at large, via Extension.

University requirements vary by institutional structure and needs. The magnitude of diversity cannot be overstated, and the examples shown in the remainder of this article are only suggestive of the scope. One Extension director noted:

The University requires budget accountability for the state line along with statewide accomplishment information that is used by the University President and Government Relations office. This information is largely based on post-survey impact evaluation data and is required with state budget reports on a yearly basis.

### **Local Level Requirements and Expectations**

Like the state stakeholders, agreements for funding at the local level vary greatly from state to state. In some situations, local entities (e.g., county commissioners or county judges) pay funds directly to Extension and those funds are then used to pay local personnel. In other cases, local Extension professionals are paid by both state Extension funds and local entity funds. In-kind funding, in the form of office space, vehicles, support staff, and other resources are also common ways local entities contribute to the Extension program and support Extension work. As with the state funding partner, evidence required at the local level varies by state and in some cases by the local entity. In Texas, county governments are typically interested in what activities the local staff is implementing. In Kentucky, there is a state agency focused on county-level work.

Local reporting also varies greatly. For instance, Kentucky has a State Department for Local Government which governs reporting for all counties. Other states have no such consistency in requirements for reporting across counties. One Extension director noted:

All county Extension offices enter monthly contacts and yearly provide quantitative program impact documents to their respective County Director. We also use the University's reporting platform (Digital Measures) to capture a variety of reporting variables, including media contacts, publications, and presentations.

Another director noted:

County Extension professionals work in partnership with Extension program field specialists and regional directors to annually complete a County Stakeholder Report. The report highlights outcomes and impact of programs within the four program areas of Agriculture and Natural Resources, 4-H Youth Development, Human Sciences, and Community and Economic Development.

At the local level, credibility and credible evidence may be focused more on specific programs and not on the organization as a whole, given the grassroots nature and expectations of the local stakeholders.

### **Grants and Special Projects Requirements and Expectations**

Reductions in funding from governmental sources at the federal, state, and local levels, as well as partnerships and alternative funding opportunities over the past decades, have led many state Extension organizations to focus more on grants and special funded projects to maintain or increase overall levels of funding. Grant and special project requirements for evidence are specific to the grant/project and the agency or group who provides the funding. The reporting requirements for these grants and projects are often outlined in a request for proposal (RFP). Grant or special project requirements for reporting may be focused on various levels of evidence including inputs (time, FTEs, financial resources), outputs (participation, satisfaction, publications), and/or outcomes (learning, behavior change, impacts). Reporting guidelines for grants and special projects will also delineate deadlines for reports. These deadlines usually take place periodically throughout the program's duration or shortly after the program's termination.

In general, an Extension or university grant or contract approval process reduces conflicts over the types of evidence and access to that evidence, but interests of external stakeholders are sometimes at odds with Extension's mission and evidence criteria. The potential negative impact is lessened when objectives of these external funding stakeholder(s) are congruent with the Extension mission.

Given the diversity of criteria for grants and contracts, it is clear that evaluative measures needed by these stakeholders to make decisions vary extensively. If clear expectations are not in place,

stakeholders might settle for what they consider to be “good enough” evaluation. This might meet the perceived needs of the stakeholder but may not fully provide the full extent of the results seen by the effort.

### **Communicating to the Public**

As the National Institute of Food and Agriculture (NIFA) states,

Extension provides non-formal education and learning activities to people throughout the country—to farmers and other residents of rural communities as well as to people living in urban areas. It emphasizes taking knowledge gained through research and education and bringing it directly to the people to create positive change. (n.d.)

Through a variety of means, Extension provides information, programming, resources, and technology to individuals and the communities in which they live in order to address public need (NIFA, n.d.). In serving these communities, Extension is in the business of building relationships with individuals and communities through Extension’s brand and serving them through the transference of scientific knowledge (NIFA, n.d.). In doing so, an incredible amount of trust must be fostered between Extension and the public. As an information hub, Extension needs to ensure that the information being provided is accurate, relevant, and applicable to the individuals and communities it serves. Thus, a new layer of credibility is added, as another stakeholder, the general public, enters the picture.

The real struggle with communicating to diverse audiences, as Extension does, is meeting the needs of each stakeholder. As a liaison between scientists and researchers, state and federal governments, agriculturalists, health experts, and the general public, Extension needs to be able to adapt its communication strategies, while maintaining a credible core message. In communicating with public audiences, Extension is charged to “translate science for practical application” (NIFA, n.d.). In the dissemination of research, the need arises for Extension professionals to serve as communicators, breaking down scientific concepts and language into digestible dialogue. The question that remains is how does Extension effectively reach an intended audience when there are multiple audiences and/or agendas? Given the role of translational research in Extension, one might argue that the best way to deliver messages to the public is through qualitative, impact-oriented information, rather than quantitative, outcome-related data (NIFA, n.d.). However, understanding that Extension also operates within scientific and governmental landscapes, others might argue that publishing with hard-scientific data and policy jargon is necessary to remain a valid and credible source.

Many Extension organizations have worked to use tailored communication strategies to meet the needs of the audience. When delivering to stakeholders, Extension must also meet accountability and evidence requirements of those stakeholders. Many times, Extension can be taken for granted until it is threatened by real or proposed budget cuts. This is the point at which



documentable impacts become most relevant and important. When serving the public, Extension must be able to offer that information in a practical, understandable way. Not having streamlined reporting guidelines has created difficulty in determining how to best publish and/or present program evidence that is deemed credible across all levels of Extension. However, this perceived fault may also be one of Extension's most valued benefits, as it allows Extension to currently operate within multiple fields, giving Extension a larger platform to reach a wider audience. By not being a one-size-fits-all organization, Extension has been able to reach multiple audiences in diverse ways.

### **Perspectives on Credible and Actionable Evidence from Extension Directors**

Numerous factors shape the generation of credible evidence and actionability of credible evidence within the Extension system. With the multitude of different stakeholders requesting varying degrees of evidence, Extension leadership is faced with having to meet the needs of these stakeholders. Providing credible evidence to meet these needs is challenging at best. One Extension director responded,

Credible evidence, as it relates to Extension, is gained through the long-term trust with Agents and Specialists tied to the Land-Grant University. Evidence that is supported through the scientific process is strengthened by proper evaluation of the educational programs. As defined, credibility has two key components: trustworthiness and expertise. Both are critical for Extension's mission of providing credible information.

As discussed in many of the other articles in this special edition of the JHSE, evaluation strategies and methods are closely linked to the concept of credible evidence. Another Extension director stated, "Evaluation methodology incorporated in the program from beginning to end and post-program is credible evidence. Applied research information is also credible evidence." A challenge in the area of evaluation includes what type of evidence is being requested. One Extension director stated, "Depending on the stakeholder/funder, credible evidence can be inputs, outputs, outcomes, and/or impact."

Internal and external factors also shape expectations for credible evidence and its impact on reporting and funding. Policy, practice, and politics all play a role in these factors. One Extension director summed up the role of these factors:

External and internal factors that shape reporting and funding expectations include, but are not limited to federal mandates and laws; university promotion and tenure systems; annual staff performance reviews; expectations of current university, state, and federal leadership; faculty and staff position descriptions (i.e., expectation to obtain external funding); available funding streams; Extension professionals skill level in securing funding through grants and contracts; and Extension professionals' abilities to plan, design, implement, and measure and report program impact.

This is an area that has vast differences among institutions. University promotion and tenure bring a greater focus on credible impact that meets the academic expectations of a university.

Does credible evidence, in terms of reporting, differ across the various levels (local, university, state, federal) of Extension? Based on the varying levels of stakeholders and their needs, as described in this article, the answer to this question is a resounding yes. As one of the Extension directors stated:

Differences among reporting across local, university, state, and federal levels primarily exist related to types of data requested. For example, at the federal level, civil rights data is requested as is the amount of time and effort Extension professionals spend on federally funded programs. At the federal level, Extension systems also must complete the Combined Research and Extension Annual Report of Accomplishments and Results and the Combined Research and Extension Plan of Work. At the local, university, and state levels, these types of reports are not requested. However, at the local, university, state, and federal levels, all require one or more reports that seek documented program outcomes and impact.”

The differences in evidence that is required by various stakeholders do exist. However, whether reporting to the federal, state, university, or local levels, all evidence must be credible to the stakeholder(s).

### **Conclusion**

Extension as a system has a variety of stakeholders at the federal, state, local, and university levels. Additionally, grants and special projects have added to the diversity of stakeholders to which Extension is responsible for reporting. So how does Extension provide credible evidence to all of these groups? As discussed in this paper, there is no clear answer to this question. If there is no clear answer, how does Extension move forward?

Extension administrators, specialists, agents, and others must recognize the complexity and variety of stakeholders and meet their needs for credible evidence and accountability. In many cases, funding is tied to these expectations. Extension must educate its stakeholders on the evidence that shows the value and effectiveness of Extension’s efforts. Extension must maintain a level of credibility by meeting these needs, while also maintaining its commitment to research-based, unbiased information and evidence. This is no small task, but one that must be accomplished as Extension continues to remain relevant and important in our counties, state, and the nation.

There is a strong emphasis on “no clear answers” in this article. Extension must focus on producing credible and actionable evidence at each level and meeting the challenges and needed efforts to generate and use more credible evidence. Because of the multiple levels of

expectations of local and state funders, there will be large differences among institutions. However, there are areas where Extension can be more consistent in its approaches and how it communicates its value and worth broadly to funders and stakeholders.

### References

- Agricultural Research, Extension, and Education Reform Act [AREERA] of 1998, 112 Stat. 523. (1998). Retrieved from <https://nifa.usda.gov/sites/default/files/resource/Agricultural%20Research%2C%20Extension%2C%20and%20Education%20Reform%20Act%20of%20%201998.pdf>
- American Evaluation Association. (2018). *Guiding principles for evaluators*. Retrieved from <https://www.eval.org/p/cm/ld/fid=51>
- Annette, K., Fauth, W., & Ahcan, A. (2015). The Blandin Foundation: The journey to a theory of philanthropy. *The Foundation Review*, 7(4), 43–53. doi:10.9707/1944-5660.1265
- Bigely, G. A., & Pearce, J. L. (1998). Straining for shared meaning in organizational science: Problems of trust and distrust. *The Academy of Management Review*, 23(3), 405–421.
- Cash, D. W. (2001). “In order to aid in diffusing useful and practical information”: Agricultural Extension and boundary organizations. *Science, Technology, & Human Values*, 26(4), 431–453. doi:10.1177/016224390102600403
- Donaldson, S. I. (2015). Evaluating the backbone of contemporary evaluation practice: Credible and actionable evidence. 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. 3–26). Thousand Oaks, CA: Sage.
- Guston, D. H. (1996). Principal-agent theory and the structure of science policy. *Science and Public Policy*, 23(4), 229–240. doi:10.1093/spp/23.4.229
- Herbig, P., Milewicz, J., & Golden, J. (1995). A model of reputation building and destruction. *Journal of Business Research*, 31(1), 23–31. doi:10.1016/0148-2963(94)90042-6
- Kazoleas, D., & Teven, J. J. (2009). Public relations and organizational credibility: Refining the definition, measurement and assessment of organizational trust. *Human Communication*, 12(1), 19–32.
- Kelsey, K. D., & Mariger, S. C. (2002). A case study of stakeholder needs for Extension education. *Journal of Extension*, 40(2), Article 2RIB2. Retrieved from <https://www.joe.org/joe/2002april/rb2.php>
- Larner, G. (2004). Family therapy and the politics of evidence. *Journal of Family Therapy*, 26(1), 17–39. doi:10.1111/j.1467-6427.2004.00265.x
- Mahon, J. F., & Wartick, S. L. (2003). Dealing with stakeholders: How reputation, credibility and framing influence the game. *Corporate Reputation Review*, 6(1), 19–35. doi:10.1057/palgrave.crr.1540187
- Mayer, R., Davis, J., & Schoorman, F. (1995). An integrative model of organizational trust. *The Academy of Management Review*, 20(3), 709–734. doi:10.2307/258792

- McDowell, G. R. (1985). The political economy of Extension program design: Institutional maintenance issues in the organization and delivery of Extension programs. *American Journal of Agricultural Economics*, 67(4), 717–725. doi:10.2307/1241810
- Moss, S., & Edmonds, B. (2005). Towards good social science. *Journal of Artificial Societies and Social Simulation*, 8(4), 1–14.
- National Institute of Food and Agriculture (NIFA). (n.d.). *Extension*. Retrieved from <https://nifa.usda.gov/extension>
- Patton, M. Q. (2003). Utilization-focused evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation*, 9, 223–244. doi:10.1007/978-94-010-0309-4\_15
- Stevens, G., Lodl, K. A., Rockwell, S. K., & Burkhart-Kriesel, C. (1994). Do local realities clash with federal expectations? *Journal of Extension*, 32(3), Article 3FEA2. Retrieved from <https://www.joe.org/joe/1994october/a2.php>
- Stoltz, M. (2002). Extension faculty and political acumen. *Journal of Extension*, 40(1), Article 1COM1. Retrieved from <https://www.joe.org/joe/2002february/comm1.php>
- Suvedi, M., Heinze, K., & Ruonavaara, D. (2005). *How to conduct evaluation of Extension programs*. Retrieved from [https://msu.edu/user/suvedi/institute/paginas/pdf/Evaluation\\_Manual.pdf](https://msu.edu/user/suvedi/institute/paginas/pdf/Evaluation_Manual.pdf)
- United Nations. (n.d.). *Population*. Retrieved from <https://www.un.org/en/sections/issues-depth/population/>
- Voris, J. C. (1991). Extension-industry-consumer interaction. *Journal of Extension*, 29(2), Article 2F1. Retrieved from <https://www.joe.org/joe/1991summer/f1.php>
- Young, L. (2006). Trust: Looking forward and back. *Journal of Business & Industrial Marketing*, 21(7), 439–445. doi:10.1108/08858620610708920

*Nick T. Place*, Ph.D., is Dean for University of Florida/IFAS Extension and Director, Florida Cooperative Extension Service; and Professor, Agricultural Education and Communication.

*Richard M. Klemme* is Executive Director of the Extension Committee on Organization and Policy at the American Association of Public and Land-Grant Universities.

*M. Ray McKinnie* is Dean and Administrator for Virginia State University.

*Carrie Baker* is a graduate student in the Department of Agricultural Leadership, Education, and Communications at Texas A&M University.

*Jean Parrella* is a graduate student in the Department of Agricultural Leadership, Education, and Communications at Texas A&M University.

*Scott R. Cummings*, Dr.P.H., is head of Organizational Development for the Texas A&M AgriLife Extension Service and Professor in the Department of Agricultural Leadership, Education, and Communications at Texas A&M University.

## Communicating with Data: Telling the Extension Story in Credible and Actionable Ways

**Diane D. Craig**

**Ruth H. Borger**

*University of Florida Institute of Food and Agricultural Science*

*Effective communication requires a good message delivered through an effective channel and received by a receptive individual. When that communication is successful, the result is enhanced credibility and trust between the sender and the receiver. Telling the Extension story effectively requires both relevant, credible data to compose a clear message and appropriate communication channels to deliver the message to various audiences. This article describes the approach taken by Florida Extension to gather better statewide data to improve communication about the impact of its Extension work, primarily through the use of infographics. With credible data, and working together, Extension data analysts and communicators can enhance Extension's reputation, trust, and support with key stakeholders.*

*Keywords:* communication, trust, data visualization, infographics, statewide indicators, impacts, data quality

***“Evidence is evidence, whether words, numbers, images, diagrams, still or moving. It is all information after all. For readers and viewers, the intellectual task remains constant regardless of the particular mode of evidence: to understand and to reason about the materials at hand, and to appraise their quality, relevance, and integrity.”***

—Edward R. Tufte (2006, p. 83)

### Introduction

As discussed in the other articles in this special edition of the Journal of Human Sciences and Extension, much thought and work are required to collect credible and actionable evidence in Extension. With high quality, positive data in hand, it is essential that Extension effectively communicate the value of that effort. What good is credible evidence if not shared and utilized? In this article, we expand the concept of credibility beyond valid evaluation practices, measurement, and context and emphasize the credibility of the message and messenger as equally important when telling the Extension story.

---

Direct correspondence to Diane D. Craig @ [ddcraig@ufl.edu](mailto:ddcraig@ufl.edu)

What happens when data quality is weak or unknown? Often, we still need to tell a story to show there is progress being made on a priority issue or to simply stay relevant. This article describes the University of Florida Institute of Food and Agricultural Science's (UF/IFAS) struggles to obtain credible evidence at a statewide level, its efforts to expand the use and usefulness of available data, and how the organization leverages resources on campus and beyond to tell the Extension story and credibly demonstrate the value of its work. We are highlighting the UF/IFAS experience not as a model approach, but as one path of many toward demonstrating the impact of Extension's efforts.

The UF/IFAS initiative to gather better statewide data, and the increased use of existing evaluation data more generally in Extension communications, has served to increase attention on program outcomes and the quality of its evaluation tools and methodologies. The initiative represents a continuous, non-linear process to make evaluation processes and products more relevant, useful, and adaptable for Extension staff and stakeholders.

Throughout this article we use the term *impact* to broadly describe the demonstrated success of an Extension program or group of programs and do not strictly adhere to the United States Department of Agriculture National Institute of Food and Agriculture (USDA-NIFA) definition of *condition outcomes* as outlined in their Generic Logic Model for NIFA Reporting guidelines (USDA-NIFA, 2015). We believe behavior change data, which is emphasized in our statewide indicators initiative, allows us to make judgments not only about program participants but also about the implications of that collective behavior change to the larger society.

There are many ways to communicate the success of Extension programs or research projects. Our focus here will be on printable, one-page infographics for showing impact because our lobbyists and Extension administrators tell us they are effective when interacting with stakeholders. While Extension data are often shared with the general public through social media and the internet, UF/IFAS has little evidence to suggest its effectiveness in communicating the value of Extension or even raising awareness of Extension. Empirical and marketing research (Evergreen, 2018b; Lankow, Ritchie, & Crooks, 2012; McCue, 2013; Tufte, 2006) indicates that portraying information visually is effective. For this reason, we strongly support the broad use of infographics for telling the Extension story.

### **Communication that Supports Credibility and Trust**

Trust can be an elusive concept and one that is influenced by how an organization presents itself to its stakeholders, including local, state and federal lawmakers; donors and funding agencies; local supporters; partners and organizations; media and digital influencers; clients; and the organization's own faculty and staff. Jiang (2016) contends that two important elements lead to trust between an individual and an organization: *relationships* and *transparency*. People trust their families, friends, and co-workers (relationships) after they have multiple experiences that confirm a level of trust. Transparency involves more analytical thinking involving the individual

being able to see and process information for him- or herself. When combined with the reality that most communication is now conveyed on a digital platform dependent on visuals and that most individuals will find this information by searching online, communicators must ask, “What visuals will increase credibility and nurture trustworthiness in the organization?”

Jiang (2016) tested five design approaches most often used on websites: 1) text and supporting small images, 2) larger photos with minimal text, 3) small infographics to explain operations, 4) text with interactive infographics, and 5) text presented in a flow chart. While there was no one comprehensive finding that one of these methods was more effective than the others, Jiang’s (2016) research indicated that the effectiveness of the communication method (e.g., photo, graphic, infographic, story) depended on the audience for which it was intended. Users preferred text, infographics, and flow chart displays when used for websites, but preferred photo narratives for social media and email. Stakeholders expected to learn different information from different communication channels (i.e., different information was expected from websites than social media). The overall finding for all communication methods was that visuals were very effective in communicating messages regardless of the channel. A significant recommendation for all Extension professionals (i.e., county Extension agents or educators and state faculty with an Extension appointment) would be to think visually and include visuals often in any communication channel.

So, how is a communicator to choose the best method of distributing credible and actionable information to diverse stakeholders? One approach would be to create a visual design of the critical data (e.g., using infographics, photos, charts, drawings) and distribute this information through multiple channels (e.g., website, blog, social media). Communicators could also conduct periodic focus groups with key audiences to assess which methods were most effective. Digital applications can now also provide data, such as the number of views, time on the page, shares, likes, and comments, that can also provide feedback on the effectiveness of different communication methods and channels. These types of information may be used to refine the organization’s communications strategy, thereby increasing the sender’s effectiveness, which leads to the increased potential of the message being received and the information receiver being more receptive to the sender’s credibility.

Specific communications tactics such as logos, website designs, and an organization’s digital presence can also influence an organization’s credibility (Lowry, Wilson, & Haig, 2014). For example, logos are considered an important part of an organization’s identity or reputation. State Extension programs are divisions of the United States’ land-grant universities. The land-grant universities are most likely better recognized and more respected than the actual Extension programs. Combined with the reality that users often make decisions about an organization within the first few seconds of viewing a website or social media post, it would be beneficial for Extension programs to prominently display and connect to their parent land-grant university through use of the university’s logo.

As an example, UF/IFAS went through a branding redesign to enhance its logo to strengthen an intentional connection between the University of Florida and that university's land-grant mission. Branding guidelines instructed Extension professionals, state faculty and staff how and when to use the logo, including emphasizing the use of "UF/IFAS Extension" verbally and in print instead of simply using "Extension" or "IFAS Extension" as they had in the past.

The effectiveness of changes like this should be evaluated over years, not months. At UF/IFAS, our results have been mixed. More Extension county offices and Extension staff are properly referring to the organization according to the new guidelines, yet the institutional application of brand standards has been inconsistent. The only absolute is that attention to and education about branding standards is a constant activity and one that is never completed.

### **The Role of Communication Methods in Establishing Credibility**

Establishing a reputation as a credible source of information is essential to an Extension program's ability to secure the trust of its internal and external stakeholders (Cutlip, Center, & Broom, 1985). The relevancy of Extension hinges on these stakeholders' perceptions that the organization is responding to critical community needs through efficient and effective programming. Understanding how these important audiences best receive information and how they assess it as trusted information is fundamental to effectively advancing Extension's mission.

To be effective, all communication, whether verbal, written, or visual, should have some fundamental elements. These include identification of the target audience(s), development of a concise and clear message, choice of the most desirable "channel" to convey the message, a credible source to deliver the message, an opportunity to be repeated multiple times, consistency of message when delivered, and the capability of the audience to receive the message (Cutlip et al., 1985).

As Extension programs serve multiple and diverse audiences, it is not likely that a single message using a singular method for all audiences will realize a program's communication goals. Therefore, a single message may be altered to be delivered by a variety of "senders" via different "channels" at different times.

Choosing the most effective spokesperson (source or sender), the best channel, clear message, and best communication tactic can be challenging (Table 1). Qualitative and quantitative research methods may be used to identify which communication tools to use. Focus groups with stakeholders, short surveys, and media content analyses are all affordable research methods that may provide valuable insights when developing a communications strategy. Then, matching the communications channel to the audience is driven by data not by personal preference (Wiles, 2017).



**Table 1. Matrix for Determining the Right Communications Channel**

Channel Type	Example	Good Way To	Pros	Cons
Central communications (one to many)	<ul style="list-style-type: none"> <li>• Press release</li> <li>• E-mail</li> <li>• Memo</li> <li>• Intranet post</li> </ul>	<ul style="list-style-type: none"> <li>• Update/inform large group about issues of “big” concern/initiative</li> </ul>	<ul style="list-style-type: none"> <li>• Scalable</li> <li>• Reaches multiple stakeholders</li> </ul>	<ul style="list-style-type: none"> <li>• Difficult to gauge impact</li> <li>• Limited opportunity to clarify</li> </ul>
Leader presentation (one to many)	<ul style="list-style-type: none"> <li>• Media interview</li> <li>• Press conference</li> <li>• Town hall</li> <li>• CEO video/blog</li> </ul>	<ul style="list-style-type: none"> <li>• Motivate and energize audience</li> <li>• Make important announcement</li> </ul>	<ul style="list-style-type: none"> <li>• Good way to address issues</li> <li>• Highly credible source</li> </ul>	<ul style="list-style-type: none"> <li>• One-way communication</li> <li>• Audience often intimidated to ask questions</li> </ul>
Manager cascade (one to few)	<ul style="list-style-type: none"> <li>• Communication in team meetings</li> <li>• E-mail</li> </ul>	<ul style="list-style-type: none"> <li>• Inform/update on team specific and/or sensitive matters</li> </ul>	<ul style="list-style-type: none"> <li>• Trusted source</li> <li>• Personal</li> </ul>	<ul style="list-style-type: none"> <li>• Messages often fail to get through</li> <li>• Manager not familiar/bought-in</li> </ul>
Manager dialogue (interactive)	<ul style="list-style-type: none"> <li>• Group discussion</li> <li>• Manager one-on-one</li> </ul>	<ul style="list-style-type: none"> <li>• Problem solve/gain feedback</li> <li>• Translate strategy into action</li> </ul>	<ul style="list-style-type: none"> <li>• Helps resolve issues</li> <li>• Drives behavioral change</li> </ul>	<ul style="list-style-type: none"> <li>• Time intensive</li> <li>• High variability in manager communication skill</li> </ul>
Mobile and social media updates (one to many)	<ul style="list-style-type: none"> <li>• Company blog</li> <li>• Intranet</li> <li>• Twitter</li> <li>• SMS mobile</li> <li>• Company Facebook</li> </ul>	<ul style="list-style-type: none"> <li>• Time-sensitive information updates/alerts</li> <li>• Humanizing the company</li> </ul>	<ul style="list-style-type: none"> <li>• Ease of access to information</li> <li>• Tracking sentiment</li> </ul>	<ul style="list-style-type: none"> <li>• If not done properly, company can be seen as “phony”</li> </ul>
Social media participation (interactive)	<ul style="list-style-type: none"> <li>• Discussion forum</li> <li>• Blog</li> <li>• Twitter</li> <li>• YouTube</li> </ul>	<ul style="list-style-type: none"> <li>• Sharing viewpoints</li> <li>• Engaging in debates</li> <li>• Creating a dialog</li> </ul>	<ul style="list-style-type: none"> <li>• Builds engagement</li> <li>• Gives a human face to the company</li> </ul>	<ul style="list-style-type: none"> <li>• Little control over the communication</li> <li>• Possible rejection of corporate agenda</li> </ul>
Enabling advocates (many to many)	<ul style="list-style-type: none"> <li>• Employees</li> <li>• Suppliers</li> <li>• Customers</li> </ul>	<ul style="list-style-type: none"> <li>• Reputation management</li> <li>• Reaching out to skeptical audiences</li> <li>• Spread viral message</li> </ul>	<ul style="list-style-type: none"> <li>• Trusted sources</li> <li>• High resonance</li> <li>• High “stickiness”</li> </ul>	<ul style="list-style-type: none"> <li>• Ability to find suitable advocates</li> <li>• Time-sensitive</li> </ul>

*Note:* Adapted from Wiles, J. (2017). *How to choose the right communications channel* [Blog post].

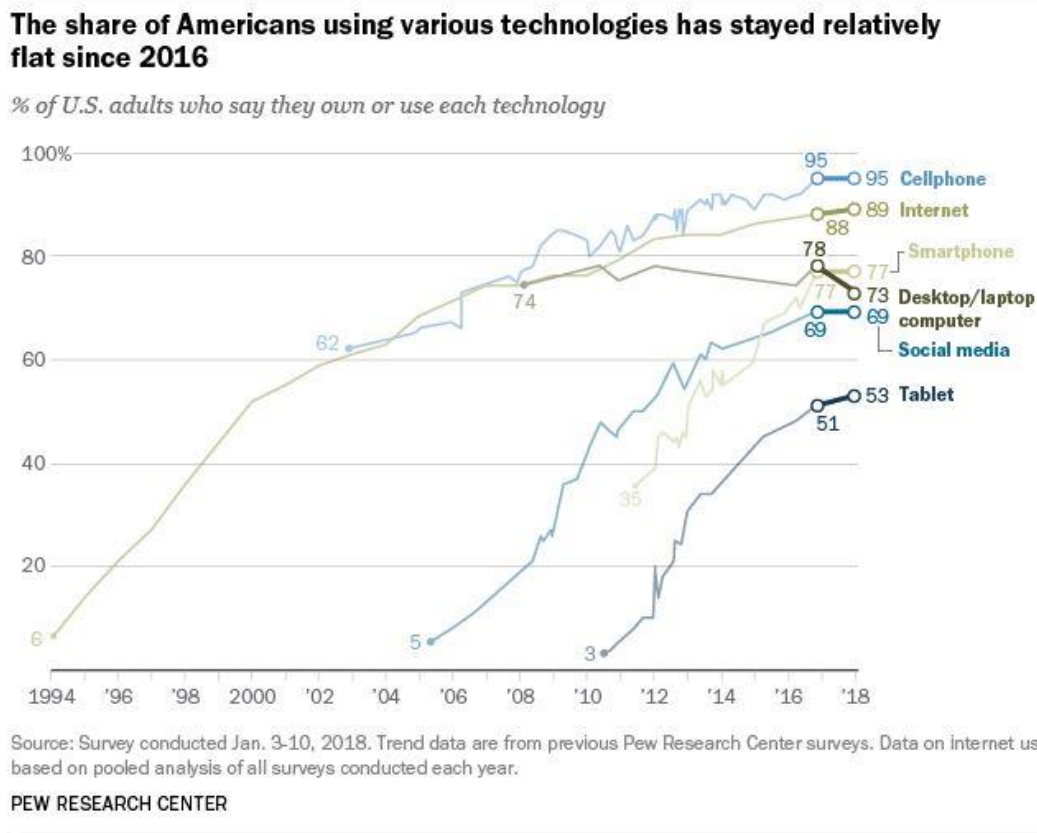
### The Changing Communication Landscape

Extension programs operate in a communication zone that has 24-hour access to data and information through the digital platform. While Extension programming has traditionally relied primarily on print publications, radio, and more recently, websites, information of all and any kind is now available through blogs, Facebook, Twitter, texts, podcasts, live streams, etc. More

importantly, new audiences search for information in an increasing number of new sources on the digital platform. A survey conducted by the Pew Research Center in early 2016 found that about only one-fifth of U.S. adults *often* received news from print newspapers, down from 27 percent in 2013. That decline is also directly related to age with less than ten percent of 18-29-year old individuals getting news from newspapers, while almost half of those 65 and older choosing traditional newspapers as a primary source of information (Mitchell, Gottfried, Barthel, & Shearer, 2016). As traditional print media consolidates and declines, hundreds, if not thousands, of digital outlets publish and distribute information every minute of every day at a global level. These new information outlets provide additional opportunities for Extension programs to distribute evidence of their success.

A more recent Pew Research Center study (Hitlin, 2018) projects that 95 percent of Americans use various technologies to receive information including mobile phones, the Internet, web and social media (Figure 1). According to this Pew study, a majority of U.S. adults now receive their daily news from digital sources, such as the Internet, social media, email, and texts, and more increasingly, are receiving their information on mobile devices.

**Figure 1. Graph of Americans' Use of Various Technologies**



*Note:* From Hitlin, P. (2018). Copyright 2018 by Pew Research Center.

According to a Pew Research Center survey (Hitlin, 2018), about two-thirds of American adults (68 percent) say they at least occasionally get news on social media. While receiving or being exposed to news is increasing online, research surveys indicate that more social interaction about the news is personal, that is, done through word of mouth with someone you know. The 2016 Pew Research Center survey (Mitchell et al., 2016) showed that 85 percent of U.S. adults who most commonly share news with others do so by word of mouth. Seventy-seven percent of U.S. adults had confidence in family, friends, and acquaintances as a trusted source on news. This is compared to 82 percent having trust in local news and 76 percent for national news (Mitchell et al., 2016).

If Extension organizations want to be perceived as credible sources in this dynamic and changing media environment, communicators of Extension information need to understand that how their stakeholders perceive an organization may be more influenced by a third party (a family member or friend) and what those parties know about the organization than by a direct exchange between the stakeholder and the organization. In addition, that perception will most likely be formed by information shared on Facebook, Twitter, and Instagram than through the traditional media of print newspapers.

### **Trusting the Information**

After understanding where and how people receive their news, the foundational issue, again, is trust. Do decision makers, lawmakers, and other stakeholders trust the source of this information? Public trust in higher education institutions is eroding. Surveys of the general public reflect opinions held by some that higher education institutions are too expensive, too political and liberal, do not allow students to think for themselves, are not relevant, and are not well run (Association of Governing Boards, 2018). These surveys call into question Extension's ability to ensure that we are prudent stewards of public and private investment (Edelman, 2018). Extension is a key component of the missions of land-grant universities in the United States and is, therefore, tied to institutions of higher education.

Extension has long prided itself as a trusted source of unbiased information, according to a national branding research study commissioned by the Extension Committee on Organization and Policy (ECOP), conducted in 2008-2010 by Copernicus Marketing and Consulting Research (North, 2011). The study surveyed members of the general public. Being seen as a trustworthy source was the highest-ranking asset that Extension had as part of its brand value (84 percent). This was followed by a credible staff and convenient access to reliable information. In the current fast-changing environment, new trend data are needed.

The ECOP study also found that Extension needed to do a much better job communicating the value it provides to the community. While the researchers concluded that Extension does deliver on its promise to provide science-based programming of relevance to clients, an awareness of Extension, in many places, still remained low. Our own experience shows annual reviews of

Florida county Extension programs regularly cite low awareness of programs as an area of needed improvement (Harder, Moore, Mazurkewicz, & Benge, 2013). In the ECOP national branding study, researchers found that even those who had heard about Extension do not know much about it. The study also found that younger people (18-35) were much less likely to have heard of Extension, posing critical challenges for advancing and securing the future of the organization.

### **Measuring the Value of Extension**

Over the past three decades, funding agencies and stakeholders in general increasingly expect Extension to demonstrate the public value of its programs (Franz, 2013), in part due to increasing competition for public funds (Franz & Townson, 2008). Many large-scale research projects now require an Extension or outreach component (Harder, Lamm, & Galindo, 2018). Thus, Extension must improve evaluation capacity relative to program design (Franz & Archibald, 2018; Rennekamp & Engle, 2008). Moreover, the ECOP national branding study (North, 2011) points to a lack of awareness of Extension's contributions to the community among the general public, particularly young adults, who are Extension's future clients; legislators; and funders.

While, in many cases, it is ideal to show economic impact in terms of cost savings or increased income, due to stakeholders' emphases on return on investment, sometimes the best Extension professionals can say is that a program's outcomes are good for the larger society, and this can be quite effective as well (Franz, 2013). Data collected by land-grant universities for accountability purposes can provide this economic impact and "public value" with some effort and creativity (Franz, Arnold, & Baughman, 2014).

Each year the Extension and Agricultural Experiment Station components of land-grant universities are required to submit a report to the National Institute of Food and Agriculture (NIFA) that demonstrates the impact the universities' work funded through Smith-Lever (Extension) and Hatch (Research) federal funds. To fulfill this requirement, institutions collect data on how many Extension clientele were reached and by what means (e.g., group session, email, phone call) as well as how many clients increased knowledge or awareness, learned a new skill, adopted best practices, or changed their behavior in positive ways. Moreover, NIFA also requires several brief impact narratives that describe a research project or Extension program, its results, and the significance of those results. Beyond this federal requirement, and perhaps other required state and local accountability reports, these valuable data are not often used in ways that convey Extension's successes to stakeholders and even to its own faculty and administrators.

Historically, many land-grant institutions' Extension organizations have struggled with showing statewide impact, focusing instead on smaller projects and programs located in a single county or region within a state. However, this is not ideal for telling a comprehensive Extension story, as evidenced by several recent federal initiatives to develop national indicators. In 2011, NIFA convened a group of 60 Extension and Agricultural Experiment Station professionals from land-

grant universities and 10 NIFA National Program Leaders with a goal of developing indicators that could be used by all states and allow NIFA to more easily provide aggregated data to USDA, the Office of Management and Budget (OMB) and Congress. This working group developed 129 indicators. The panel's final report emphasizes the states' adoption and reporting of these indicators to NIFA is voluntary (USDA, 2011), and this practice continues today.

Similar efforts to develop common measures among states have occurred more recently in other Extension program subject matter areas, such as Family and Consumer Sciences (FCS) in 2013, SNAP-Ed (the nutrition education component of the USDA's Supplemental Nutrition Assistance Program) and Community Resource and Economic Development (CRED) in 2014, and 4-H in 2015. The extent to which states are using these shared measures in conveying Extension's impact to stakeholders is unknown.

### **Demonstrating Statewide Impact – A State Example**

At the University of Florida, Extension professionals are expected to collect evaluation data for their Extension programs. New hires also receive evaluation training and support as part of their onboarding. The primary purpose of Extension program evaluation is to assess the quality, impact, or success of trainings, activities, workshops, and programs. For the purpose of telling the Extension story, the focus is on program impact; that is, "Did the program make a positive difference in the lives of the participants or their community?" Stakeholders and funders want to know if the participants learned something that increased knowledge or awareness, if they developed new skills, if they changed behaviors, or if the program improved the social, economic, or environmental conditions of the individual participants or the communities in which they live. The struggle for UF/IFAS Extension has been how to get from demonstrating success for individual programs to showing the statewide impact of many programs focused on a critical issue. Reasons for this struggle are largely due to four factors: size, job structure, culture, and local influence. Florida Extension employs about more than 600 Extension professionals and state Extension faculty. Most hold tenure-track or permanent status positions, a process which encourages individuality. Florida has a long tradition of autonomy for Extension professionals, and streamlining evaluation represents a significant cultural shift for the organization. In addition, Extension professionals are expected to meet the needs of the counties in which they work, and programs may be adjusted to meet those needs. Florida's 67 counties provide significant funding (ranging from 20-60 percent) for Extension programs.

Encouraged by the 2011 NIFA initiative to create national indicators for priority areas, and after years of evaluation specialists working with issue-based teams of Florida Extension professionals to develop shared surveys and evaluation methods that produced limited results, in 2017 UF/IFAS Extension initiated a new approach to gather more statewide data quickly and efficiently. In six months, new statewide indicators were developed, with input from state faculty, Extension professionals, and Extension program leaders. Many of the NIFA indicators

developed by the NIFA working group in 2011 provided the foundation for the initial UF/IFAS effort.

In designing the Florida Extension indicators, the focus was on creating broadly worded measures that could capture a range of programs with one indicator and provide flexibility to Extension professionals who use multiple methods and/or survey questions to collect evaluation data. The goal was that every Extension professional and state faculty who works with clientele should have data for at least one of the statewide indicators. Importantly, most teams were limited to 2-3 indicators, resulting in 87 statewide indicators for the first year and nearly 100 for the second year of data collection. During this process, it was easy to see how creating indicators can quickly get out of hand. The added burden of having too many statewide indicators on those reporting data as well as those managing the data is equally important.

A key consideration throughout the process of designing and selecting statewide indicators is identifying the critical elements or data needed to effectively communicate impact, and just as important, evaluating whether those data are credible. Are enough faculty using valid instruments to adequately capture the data requested? What is the level of program fidelity (i.e., are the underlying programs being evaluated, strictly implemented as designed in terms of lessons, target audience, and frequency [high fidelity] or are they widely adapted to meet local needs [low fidelity])?" (Olson, Welsh, & Perkins, 2015). Due to Florida Extension's size and decentralized nature of programming, program fidelity is difficult to assess statewide. Moreover, adapting programs to meet local needs may actually increase the data's credibility, so any statewide or large-scale review of program fidelity must be done on a case-by-case basis. Given the significant time and resources it takes to develop valid instruments and assess the degree of program fidelity underlying key indicators, UF/IFAS recognizes this work will take time and must be prioritized, focusing on key measures and evaluation results that the organization needs to communicate to stakeholders.

To limit the number of indicators displayed in the reporting system to the Extension professional and state faculty, UF/IFAS maps each indicator to one or more issue-based teams. When the Extension professional or state faculty member indicates that they work under a particular team, the reporting system displays all the indicators associated with that team. Indicators typically apply to more than one issue-based team, and sometimes more than one program area (e.g., some food safety indicators apply to both the agriculture and family and consumer science program areas). Each indicator is displayed once in the reporting system even if it is tied to several teams associated with the Extension professional or state faculty. Eighty percent of Florida's Extension professionals report under multiple teams so they may be shown several indicators if there is little overlap in those teams' indicators. The difficulty in such an approach is displaying enough indicators to adequately capture the work of the Extension professional or state faculty member while not overwhelming and frustrating them with a long list of indicators.

### **Improving Data Quality for Communications**

One downside to using broad statewide indicators in a reporting system is that much of the underlying evaluation data, methodology, and degree of program fidelity is unknown when looking at the overall results. To address this limitation, the statewide indicator results are shared widely with the Extension teams and administrators. The raw data, listing the individual Extension professional or state faculty member and her or his work unit, are most helpful in assessing data quality by the identification of outliers, inconsistencies, and unexpected results. As one team leader said after a review of the data for his program area, “I don’t think they are interpreting the indicator as we would like.” Ideally, the Extension evaluation specialist uses this information to create better evaluation training for Extension professionals or modifies the indicator wording or both.

Notably, the data quality of the statewide indicators is also weighed in the context of limited time and resources. There is great variability in data collection and quality among programs due to many factors, including external pressure or relevance, program maturity, evaluation skills and capacity of Extension professionals, and the degree of support from Extension evaluation specialists (Rennekamp & Engle, 2008). At UF/IFAS, the more mature programs, with established program theory and research-based outcomes, are more likely to be highlighted using communication devices such as infographics. In this case, the broader indicator or indicators can then be used to supplement the more specific programmatic and credible evidence and provide some statewide context. For example, a 25-year program to develop systems and best practices to help Florida’s watermelon growers has credible data showing greater crop yields with corresponding decreased use of water, fuel, and fertilizer. In Extension communications, these program-specific data can then be supplemented with statewide indicators showing the number of producers statewide that participated in Extension programs and adopted agricultural recommended practices or reduced fertilizer usage. Thus, the story and the potential impact can be broadened beyond a single crop. In sum, whether the statewide indicators are used to tell the Extension story, in whole or in part, there is a constant need to assess and address data quality across all indicators. The main question to be asked is: Do these data credibly represent Extension’s efforts to accurately tell the Extension story and demonstrate impact?

Since launching these statewide indicators in Florida, the number one request from UF/IFAS Extension professionals is a list of the specific survey questions that are tied to each of the statewide indicators. Due to the broadness of the statewide indicators, including the fact that multiple program areas may report under a single indicator, there could be several survey questions created per indicator. The individuals best suited to develop the survey questions are the state faculty and Extension professionals working together on issue-based teams. UF/IFAS is also fortunate to have several evaluation specialists who work with individuals and teams to design measures and evaluation instruments that are reliable and valid. However, their support is limited, given the large number of Extension programs and issue-based teams in their program

area, as well as having to fulfill research, teaching or other Extension-related duties. Even with this dedicated (although limited) support, for some programs it can take years to develop measures widely accepted and adopted by Extension professionals.

The statewide Extension indicators initiative has prompted renewed energy at UF/IFAS toward identifying common measures and building a repository of survey questions that are vetted by evaluation specialists. Taking it a step further, some Extension teams are mapping statewide program objectives to the statewide indicators to better assist Extension professionals with developing their own program plans of work. In short, adding the indicators to the annual reporting system is telling the Extension professional what the organization values as evidence of successful Extension programming. In turn, the Extension professional then asks, “If that’s what I’m supposed to be evaluating, tell me how to measure it.”

In a quest for statewide impact data, a greater commitment to using shared, validated evaluation instruments by some Extension teams is an added, and somewhat unexpected, benefit. While building individual evaluation capacity at UF/IFAS is ongoing and necessary (to, at a minimum, provide a foundation for weighing the credibility of the evidence collected), to truly capture program impact for critical issues facing the state, the authors feel that the greater use of common measures is required. Improved evaluation rigor, along with the responses from a larger number of program participants, will allow for more in-depth analysis to produce results and reports that are both credible and actionable.

### **Focus on Behavior Change**

The focus on knowledge gain or intent to change is common evaluation practice among Extension professionals (Franz & Townson, 2008; Lamm, Israel, & Diehl, 2013). It is difficult and costly (in both time and money) to observe or survey participants multiple times months or years after the conclusion of a program. While this is the best way to assess long-term changes in behaviors and practices, and should be promoted in any Extension organization, the reality is that it will be the exception rather than the rule. More often than not, Extension professionals will conduct a pre/posttest or a single survey at the conclusion of the program asking participants if they anticipate changing or have changed certain behaviors or practices as a result of their participation in a workshop or program. However, a post-program survey provides some evidence of changes in behavior and practices that can be tied to research-based outcomes and allow for estimating the economic impact of those changes. A case in point is the use of participants’ changing watering practices to estimate gallons of water saved rather than participants’ awareness that they *should* change watering practices. The inference from awareness to water savings is not reflected in research or practice. For this reason, the statewide indicators used by UF/IFAS Extension are almost exclusively based on outcomes related to behavior change.



Here is one example from UF/IFAS that demonstrates the commitment it takes to get to the widespread adoption and reporting of behavior changes. For the past five years, a state faculty specialist in urban water conservation, working with other researchers, has educated Extension professionals on best management practices in program planning and evaluation related to water conservation programming through in-service trainings, individual consultations, publications, and webinars. As a result of these efforts, reporting of gallons of water saved by residential participants (one of the new statewide indicators) has increased from 40 million gallons in 2015 to 300 million gallons in 2018. Each year there has been greater participation among Extension professionals in using the recommended tools for capturing change in specific behaviors (e.g., the use of micro-irrigation, reduced irrigation days) and reporting these data in a consistent manner. To date, 100 Extension professionals have been trained.

Gains in awareness and knowledge, while important in assessing the program's quality, will typically not carry as much weight as behavior change with stakeholders looking for program impact. Of course, stakeholders' expectations vary depending on the program and their level of engagement. On the other hand, flexibility is key to ensure that all program areas are captured by the statewide indicators and also to obtain faculty buy-in (i.e., participation). So, some indicators may focus on awareness or intent to change simply because a program is too new or undeveloped to collect evidence of behavior change, but still address a critical need about which we want to inform stakeholders of our progress. Whether a long-standing program or a relatively new one, engaging with stakeholders on statewide indicators is an opportunity to educate everyone involved on what a program designed to produce behavior change looks like and what is required in terms of time and resources to develop an evidence-based program.

By focusing the statewide indicators on behavior change rather than knowledge gain, there may also be some pushback within the organization. Some Extension professionals will not change their evaluation practice to incorporate behavior change and thus, have no data to report for the statewide indicators. Others may feel their individual work is not reflected or that the indicators understate the organization's body of work. The message here is that the statewide indicators are created to show the impact of Extension on key issues important to the state, but not in all areas. Other work is to be reflected in the individual's annual report of accomplishments.

UF/IFAS's recent emphasis on showing statewide impact, primarily through behavior change, is intended to more effectively "tell our story" and improve our data quality. However, the evidence we collect through this effort will not influence public awareness and decision-making unless it is reported in a way that enables stakeholders to understand the impact of Extension programs. In a world of information-overload, telling the story requires effective packaging, but must begin with a substantial and credible package.

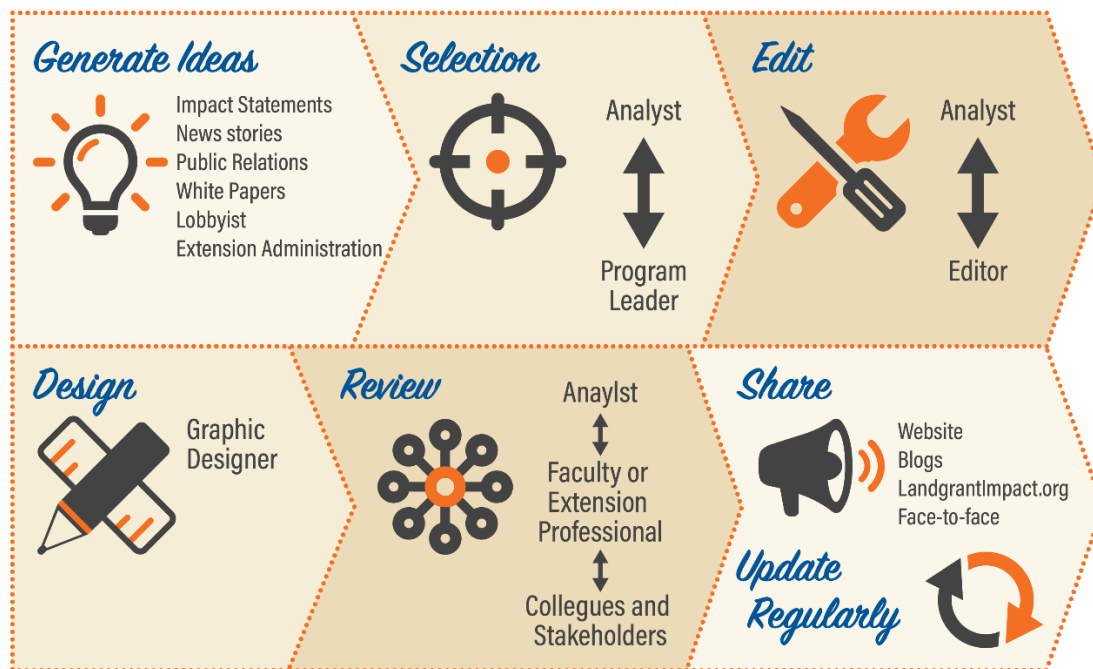
## Infographics

Based on our experience at UF/IFAS, the use of infographics based on well-documented, credible data is one of the best ways to show the impact of Extension. According to Smiciklas' (2012), "[a]n infographic (short for information graphic) is a type of picture that blends data with design, helping individuals and organizations concisely communicate messages to their audience." This definition suits Extension well. According to the Social Science Research Network, about two-thirds of the human population are visual learners (McCue, 2013). With their emphasis on visually appealing layouts and graphics, infographics can quickly bring attention to the key impacts of a program and provide some context. Moreover, studies also suggest that adding visual elements to information improves retention and comprehension (Lankow et al., 2012). Infographics are also easily shared via print or online, and today's digital technology makes infographics well-suited for viewing on mobile devices.

Ideas for infographics can come from a variety of sources that Extension professionals already use to show the impact of their programs, including impact statements typically submitted once a year by Extension professionals for accountability reporting purposes, impact narratives submitted to the national Land Grant Impacts database at <https://landgrantimpacts.org>, white papers describing long-term research projects that address critical state needs, and stories developed by communications staff. At UF/IFAS, the decision to create an infographic is typically made by the data analyst and the program leader as they learn of programs or projects with positive evaluation data that are scientifically sound and document behavior change. This vetting by the program leader ensures that research or program results that are complex, nuanced, or controversial are more carefully reviewed and crafted. Most evidence of this latter nature is better suited for other communication methods, such as white papers, town halls, and strategic communications campaigns (of which, infographics may also be one of many approaches used to communicate the issue at hand).

Establishing a method for developing infographic ideas on a regular basis is instrumental in building up a central repository of infographics. The process for developing a new infographic at UF/IFAS emphasizes the collaboration between the data analyst and the communications team (i.e., editor and graphic designer), and between the analyst and the primary Extension professional or faculty member who is providing the evaluation data. This process is illustrated in Figure 2, developed by the authors. In this set-up, the data analyst serves as the "bridge" between the Extension professional and the communications staff. During the design and review phase, several versions of the infographic are developed as the information and data are refined through continuous dialogue between the Extension professional, who is vetting the infographic with his or her coworkers and the program's advisory board, partners, clients or funders, and the data analyst, who relays information and edits to the graphic designer. Creating new infographics is time-consuming, so the key is to utilize data and design elements, wherever possible, that can be easily updated each year.

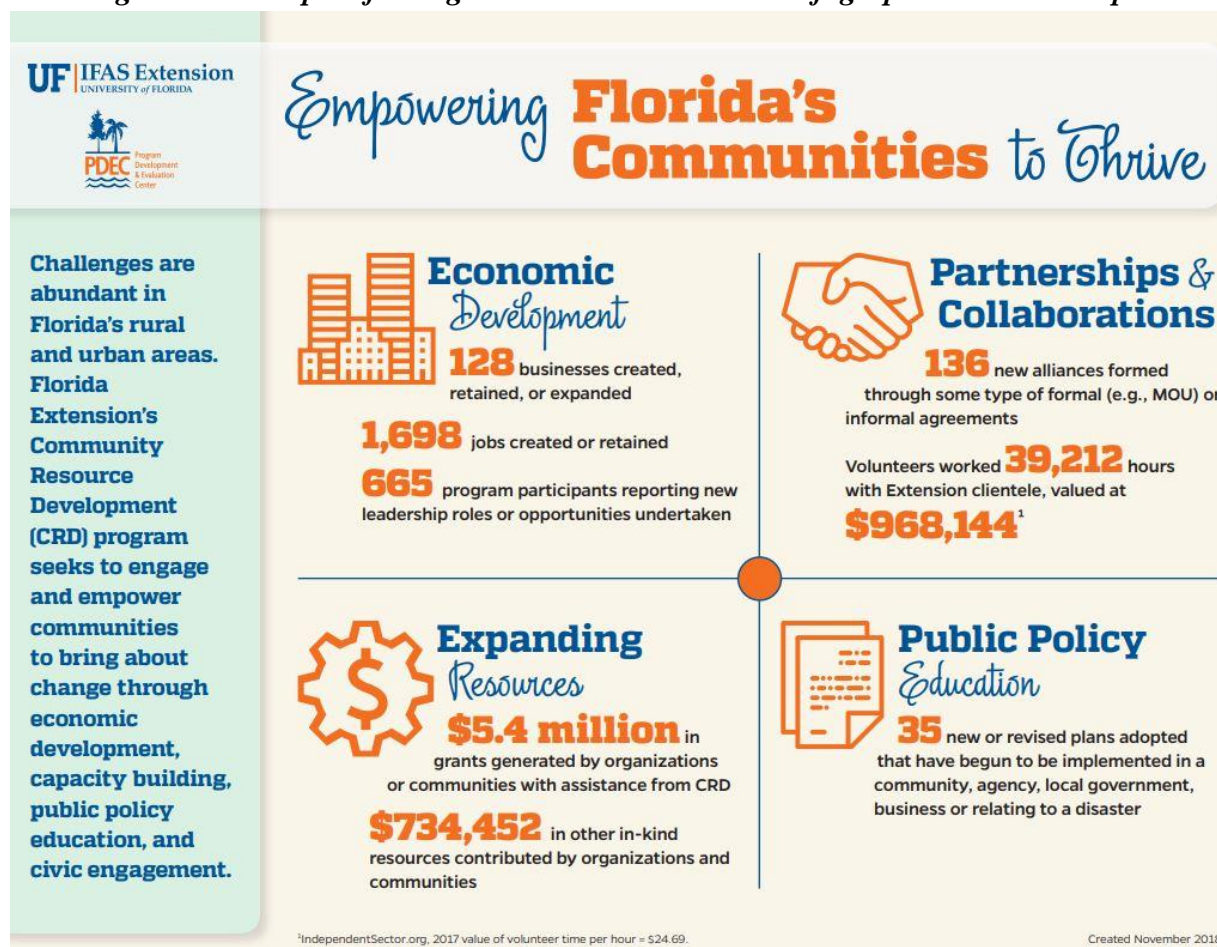
**Figure 2. UF/IFAS Process for Developing an Extension Impact Infographic**



*Note:* From UF/IFAS Communications and Program Development and Evaluation Center (2019).

The process described above is used for infographics focused on the impact of specific programs. At times, UF/IFAS has used the statewide indicators data to quickly produce infographics to meet an immediate or specific need. For instance, during the state legislative session, an Extension- or University-employed lobbyist may request information about a certain program area to jump-start a larger discussion. In a recent example, UF/IFAS was able to quickly produce an infographic based on the Community Resource and Economic Development (CRED) national indicators (and replicated in the UF/IFAS statewide indicators), shown in Figure 3. Due to the statewide indicators' broad, general nature and lack of evaluation details, in terms of credibility, statewide indicators are less desirable when details about specific programs are required.

**Figure 3. Example of Using Statewide Indicators in Infographics to Show Impact**



Note: From Craig, D., Bryant, T. & Palmer, D. (2018a).

### Economic Impacts

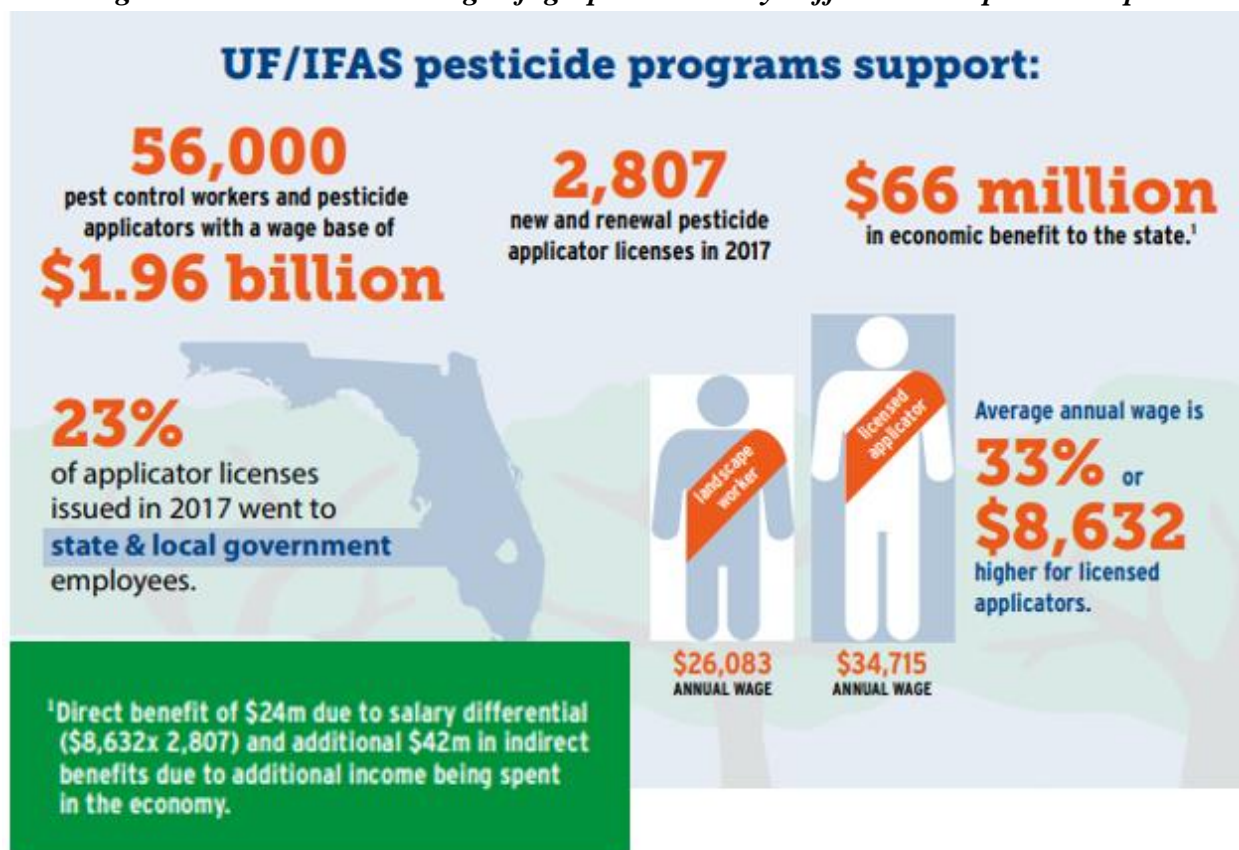
Economic impact studies are a very popular tool for universities to quantify their impact on the local economy. In 2014, the Association of Public and Land-Grant Universities (APLU) and the Association of American Universities (AAU) produced guidelines for doing such analyses, in part to address the lack of credibility of results being touted by universities and colleges (Ambargis, Mead, Rzeznik, Swenson, & Weisenberger, 2014). At the University of Florida, IMPLAN, a widely used and well-regarded economic impact assessment software system, is used to calculate the economic impact of agriculture and natural resources industries across the state at the county and industry sector level. Mulkey and Hodges (2018) provide a more thorough discussion of IMPLAN and its use. A team of 3-4 UF/IFAS researchers and economists replicate this economic impact study each year, producing a 50-page report that, while valuable, will be read by few stakeholders. To highlight this important work and maximize its value, the data contained in the lengthy report are summarized and published in the following formats:

- A county-specific economic impact sheet with other data, including private gifts from county citizens to the university or UF/IFAS, student enrollment, and clientele contacts (<https://ifas.ufl.edu/media/ifasufledu/ifas-dark-blue/docs/orange.pdf>) (Craig & Bryant, 2017);
- The authors' departmental website (<https://fred.ifas.ufl.edu/economicimpactanalysis/publications/2016-ag-natural-resources-and-food-industries/>);
- A "Fast Facts" data-driven brochure focused on the various industries and commodity groups ([https://ifas.ufl.edu/media/ifasufledu/ifas-dark-blue/docs/pdf/impact/ICS\\_FloridaAgFactsBooklet2018.web.pdf](https://ifas.ufl.edu/media/ifasufledu/ifas-dark-blue/docs/pdf/impact/ICS_FloridaAgFactsBooklet2018.web.pdf)) (Hodges, Rahmani, & Court, 2018);
- A Return on Investment (ROI) brochure ([https://ifas.ufl.edu/media/ifasufledu/ifas-dark-blue/docs/ROI\\_Booklet2019.Web.pdf](https://ifas.ufl.edu/media/ifasufledu/ifas-dark-blue/docs/ROI_Booklet2019.Web.pdf)) (UF/IFAS, 2019); and
- An Extension annual report included in Extension calendar ([https://ifas.ufl.edu/media/ifasufledu/ifas-dark-blue/docs/2017ExtensionCalendarAR\\_Final.pdf](https://ifas.ufl.edu/media/ifasufledu/ifas-dark-blue/docs/2017ExtensionCalendarAR_Final.pdf)) (Craig & Bryant, 2018).

This UF/IFAS study is but one approach to an economic impact analysis of Extension programs. Other universities have also developed effective means to convey the economic impact of Extension. For example, Texas A&M AgriLife Extension (<https://agrilifeextension.tamu.edu/about/economic-impact-briefs/>) and North Carolina State Extension (<https://www.ces.ncsu.edu/how-extension-works/extension-impacts/>) have websites devoted to highlighting economic impacts in infographic and other formats.

Economic impact can also be calculated by looking at national studies and data, such as labor statistics, and applying those to the program participant data. For example, in a pesticide licensing infographic (Figure 4) we use the U.S. Bureau of Labor Statistics (BLS) wage estimates for Florida to determine the annual salary differential between being an unlicensed landscape worker and a licensed pesticide applicator. Then we multiply the difference (\$8,632/year) by the number of pesticide workers who received new or renewed licenses in that year. Using IMPLAN, the direct benefit of having those workers with higher income in the state is calculated as well as the indirect benefits that arise when they spend those extra dollars in their communities. While not perfect, using IMPLAN does provide a straightforward way to communicate the economic impact of the program to the state. Importantly, this methodology was developed by an economist, and the infographic is vetted with Extension professionals who teach these certification classes and the Florida governmental agency that provides the licensing data each year to ensure this methodology is a reasonable estimate of the program's economic impact.

**Figure 4. Pesticide Licensing Infographic as Salary Differential Impact Example**



Note: Adapted from Craig, D., Bryant, T. & Palmer, D. (2018b).

### Client Satisfaction Surveys

Customer or client surveys are another means often used by Extension organizations to assess general satisfaction with programs. At UF/IFAS, customer satisfaction data are used in a variety of publications, including county-level infographics. With valid instruments and a sufficient number of respondents, these types of surveys can be used to produce credible evidence that programs are viewed by participants as effective and useful. This information can also show the reach of a program (i.e., are participants sharing the information they learned with others?). Moreover, it is a clear signal to stakeholders and others that the Extension organization is actively and openly measuring its performance. This willingness to circulate participant feedback broadly can, in turn, increase the organization's credibility and trustworthiness. Among Extension professionals, utilizing customer satisfaction data to improve program implementation is another way to make these data more meaningful (Franz & Archibald, 2018).

For some stakeholders, client satisfaction data by itself may be insufficient to assess the success of the programs or to take action (e.g., continue funding). Also collecting basic information about participants' reported behavior changes or practice adoption through a client satisfaction survey can strengthen the evidence of a program's success for many stakeholders. A simple way

to collect behavior change data that can convey such impact in an infographic is to ask the participants whether they saved money, saved water, improved their health, etc., as illustrated in the question shown in Figure 5. Combining satisfaction data with some behavior change impact data elevates the usefulness of these data by demonstrating to stakeholders that our clients received a direct benefit from their interaction with Extension.

**Figure 5. Clientele Benefits Survey Question**

**9. Please think about how receiving Extension services has affected you. How have you benefitted from Extension's services during the last year? Check all that apply.**

Saved me money

Increased my income

Improved my health or well-being

Helped me conserve water or energy better

Developed my skills as a leader or volunteer

Other benefit \_\_\_\_\_

\_\_\_\_\_

I have not benefitted in any way during the last year

*Note:* Adapted from Israel, G. D. (2018).

### **Increasing the Credibility of Infographics**

Assessing the credibility of evidence is more than just looking at data and methodology alone— one must also consider the source, the reasonableness of what is presented, and the believability of inferences or causal relationships claimed (Schwandt, 2015). While the evidence reported must remain credible, different stakeholder groups may require different approaches as they come with varying levels of knowledge, experience, and passion about the issue or program.

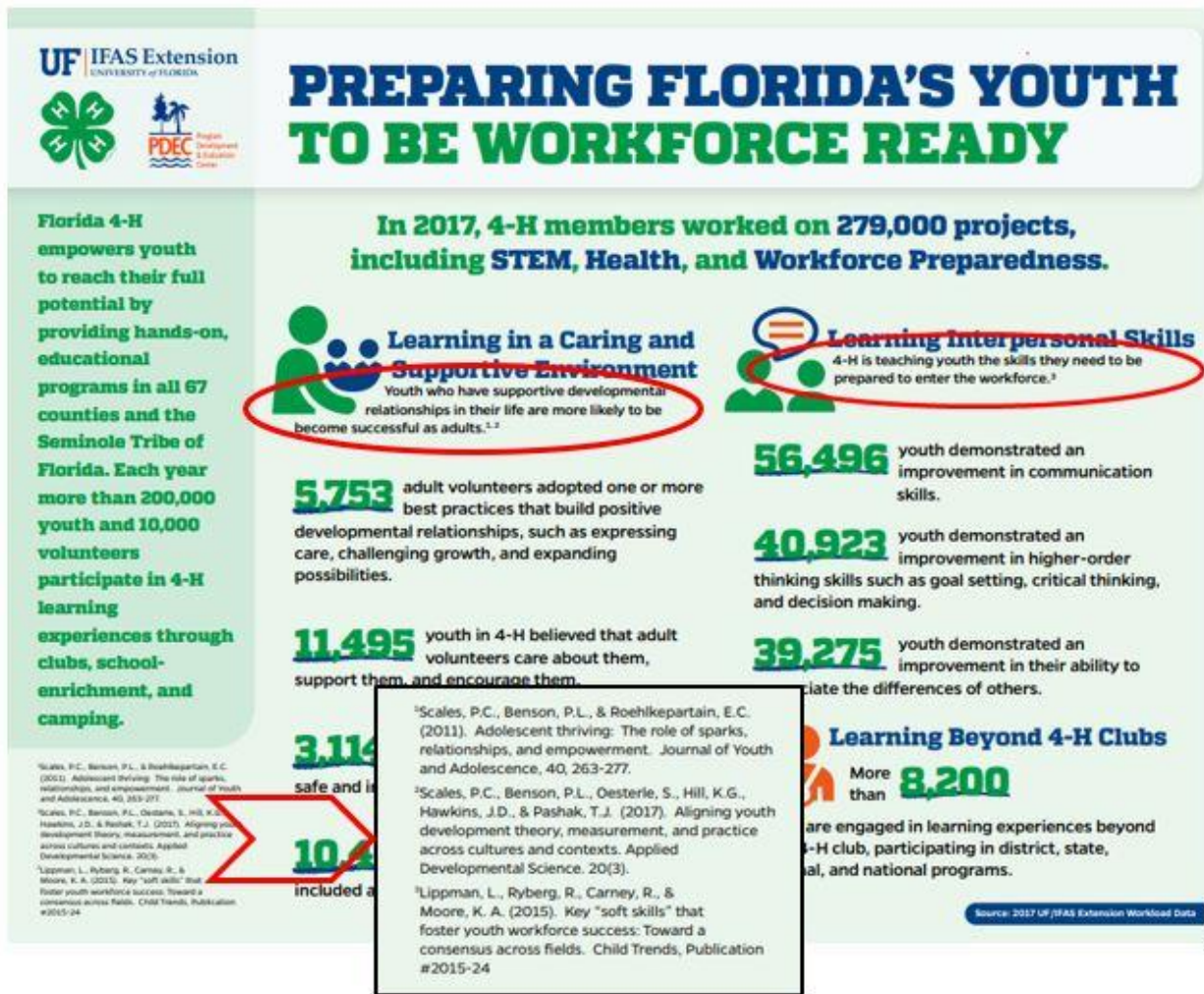
Whether an infographic is demonstrating the impact of a single program or a statewide impact, there are several practical steps one can take to increase the likelihood that the reader will find the evidence credible:

- Provide citations to relevant research;
- Include the methodology used for economic impact estimates;
- Include a logo for the institution and program (when available);
- Use appealing fonts, colors, and images;
- Vet the infographic with state faculty who specialize in the subject area;

- Vet the infographic with Extension professionals who work directly with the clientele; and
- Indicate the data year(s) and update regularly, if possible (Evergreen, 2018b; Krum, 2014).

Based on the UF/IFAS experience, one of the strongest ways to increase credibility is to support the results presented in the infographic with research findings. Using published studies to demonstrate that the behavior change identified in the indicators can reasonably lead to certain long-term outcomes gives the reader additional context and answers the “so what?” questions. This linkage between program evaluation and research literature, described by Urban and Trochim (2009) as the “golden spike,” and illustrated in Figure 6, is important to consider when first developing programs and statewide indicators.

*Figure 6. The “Golden Spike” – Connecting Statewide Indicators to Research*



Note: Adapted from Craig, D., Bryant, T. & Palmer, D. (2018c).



The authors provide this 4-H workforce infographic to illustrate the general concept of linking infographic results with research and recognize that it is not necessarily a strong example of an empirical linkage between evaluation and research. Work continues at UF/IFAS to develop infographics better representing this concept.

Many Extension programs, such as 4-H and Master Gardeners, have positive public perceptions and familiar logos. Connecting lesser-known Extension programs with universities, which are seen by most people as trustworthy, is also beneficial. Adding a well-designed logo for the program or other related images, along with other appealing “surface characteristics” conveys professionalism and improves credibility (Miller, 2015).

Using website and logo designs as the principle communication vehicle for presenting information, Lowry et al. (2014) propose that if a consumer interacts with a website and perceives trustworthiness and expertise in the website, the consumer will likely perceive trustworthiness and expertise in the organization represented by the website. For example, if a website is consistent in design with reassuring colors and stable shapes, the organization increases its perceived trustworthiness. Trusting the source of the information is an important step toward finding the stated evidence itself to be credible (Miller, 2015).

Typically, the infographic is developed with a single state faculty member or Extension professional. Encourage the contributor to circulate the infographic among their peers and their program leader prior to finalizing. Encourage Extension professionals to show it to their advisory board or influential clientele. These individuals, inside and outside the organization, will help ensure that the data, methodology, text, and even choice of images is seen as credible to stakeholders. Mathison (2015) notes the importance of verifying whether the images used in evaluation are viewed similarly by all parties—evaluator, stakeholder, and viewer/reader. This verification is especially important when disseminating infographics broadly on blogs and social media.

Infographics without any dates or with very old information may reduce the credibility of the information for some readers. Timestamps are a shortcut for assessing relevance. A missing date conveys a lack of transparency and may frustrate the reader as they are spending time looking for a date rather than looking at the information itself. Indicating when an infographic was created or updated is an easy way to increase the chances the infographic is viewed as credible.

### **Actionable Evidence**

To get the most out of any infographic, report or presentation showing evidence of the impact of Extension, the key is to use an effective messenger who is familiar with both the stakeholder and the evidence presented. For telling the Extension story, the goal in most cases is to have the stakeholder take positive action (fund or advocate for the program) or *not* take negative action

(defund or oppose the program). The degree to which the evidence presented in an infographic or other communication vehicle is sufficient for a stakeholder to take such action depends on credibility, relevance, and comprehensiveness (Mark, 2015). Asking all of these of a one-page infographic is likely being too optimistic, especially with well-informed and experienced stakeholders. However, university-paid lobbyists and Extension administrators and professionals who work directly with county administrators and other funders can take full advantage of an infographic's pared-down message by using it as a tool to initiate a dialogue about funding levels, new directions to take, or programs to add.

The following are ways that UF/IFAS has strived to increase the actionability of an infographic:

- Select timely and relevant programs or topics,
- Tailor messages to specific stakeholder groups,
- Include several data points or themes to increase chances of “grabbing their interest,”
- Provide a print version in PDF format, and
- Post in multiple places (e.g., website, blogs, news feeds, dean's message).

Beyond this targeted and strategic outward use of infographics, Extension professionals are also often in need of infographics and other demonstrations of impact for responding to ad hoc requests from university administrators or trustees, government officials, lobbyists, etc. Having a repository of infographics on statewide programs or critical issues and promoting them through websites, blogs, and social media encourages easy access by Extension and other land-grant university personnel for immediate use and distribution.

### **Working with Communications Professionals**

Effectively communicating Extension's story requires data that are credible and relevant but also accessible. If the data are not presented in a clear, concise, and consistent way using methods that are familiar or comprehensible to a non-academic audience, then the effort of collecting and interpreting the data is of limited value. Most land-grant universities have professional communications departments to manage the communications efforts needed to distribute information to stakeholders. These units have developed into strategic partners that design and implement integrated communications and marketing programs that identify, establish, and cultivate stakeholder relationships that are mutually beneficial to the Extension program and its stakeholders.

Partnerships between data analysts, who collect and analyze accountability and other data from Extension professionals and Extension or university communicators result in products that not only connect and resonate with audiences but also prioritize the data with the most relevance for the intended stakeholder. Moreover, communications units have access to the software and expertise that allow for information to be created once and then distributed among multiple platforms. This “create once, curate often” strategy is effective in sharing important information

on a regular, consistent schedule. For example, Florida Extension has information on the number of gallons of water saved due to participation in UF/IFAS Extension urban landscape programs. This impact statistic is repeated in the UF/IFAS Extension calendar, displayed in infographics, used on the web, is the topic of social media posts, is featured in a report to the state legislature, and is used in PowerPoint presentations, speeches, op-eds, etc.

Just as with any successful partnership, working with communicators and data analysts takes time, patience, collaboration, and mutual respect. Successful collaborations often start with time to explain the process of creating the work product. Understanding how each other works in this partnership not only streamlines the process but also establishes trust.

Some of the typical questions that are addressed between the communicator and the data analyst include:

- What is involved in creating an infographic?
- How do I distill information into an annual report graph?
- How do you collect the impact and outcome information?

Once there is a basic understanding of each other's role and capabilities, collective brainstorming can result in an agenda of ideas and opportunities that may be developed. Communications divisions often have a rolling calendar of scheduled publications such as annual reports, calendars, briefing reports, and web updates that need content to communicate the impact that the Extension programs deliver. Program impact data and evidence may also be plugged into those content calendars for consistent exposure.

### **Options for Low Support Environments**

Many of the ideas described thus far come from the perspective of a large, major research university with dedicated communications and marketing staff. Some organizations do not have ready access to such staff and fewer resources in general. For those with limited access to graphic designers, there are several free or low-cost infographic templates available online, including Piktochart, Canva, Visme, and Vennage. For more customization and the ability to use data-merge to create multiple versions of the same infographic (e.g., region or county level), Scribus (and the add-on ScribusGenerator) is a free desktop publishing tool similar to Adobe InDesign, software commonly used by professional graphic designers. The learning curve for Scribus is higher, but worth it if the infographic is one that will be updated on a regular basis. To get started quickly, without learning new software, one may want to consider hiring a graphic designer through a low-cost, freelance site like fiverr or guru. However, freelance graphic design services only cover the layout of the infographic and not the content.

While good graphic design is vital to producing a credible infographic or other communication vehicle, the writing and fine-tuning that is required to get to the essence of a program's outcomes

and impacts are equally important. For those who lack these skills and cannot recruit or hire a student or someone in the organization to help, there are freelance content writers available online, but it may take a lot of searching and several hires to find the right person. Because the content may need to be vetted and refined with several people in the organization, it is unlikely that freelancing is the best solution for creating infographic content. Regardless of whether this work is done in-house or not, it is important to review the content with program leaders, specialists, and Extension professionals before starting the design process.

### **Interactive Data Visualization**

Thus far, our article has focused primarily on communicating impact through infographics, a static form of data visualization. However, Extension can also display data in more dynamic, interactive ways (Lankow et al., 2012). Online data dashboards that display key metrics and allow the user to filter and easily compare data are becoming increasingly popular in today's technology-savvy society. Businesses have used this type of data visualization internally for decades, but widespread use began in the early 2000s as Microsoft Excel charting improved and new desktop software designed for a non-technical user became available. Today, many universities and colleges are using interactive data dashboards for their "university factbook" (i.e., student enrollment, faculty and staff counts, degrees awarded, etc.) and graduate exit or other student-focused surveys. Among these higher education users, the most popular data visualization software is Tableau, followed by Microsoft Power BI.

Data dashboards are typically more exploratory in nature than infographics, which are explanatory (Knafllic, 2015; Lankow et al., 2012). A designer can create a dashboard that "walks" the user through a storyline, but the fixed infographic, which is highlighting specific data and content, is a more direct route to communicating impact. In a dashboard that features a lot of interactivity, the user can get sidetracked as they explore the data and possibly miss the story the designer was trying to tell.

Data visualization software such as Tableau is ideal for analyzing "big data," which are very large data sets with a lot of detailed information. These data may be real-time (i.e., coming directly from a reporting system or web server) or a fixed data set (e.g., an Excel spreadsheet). Many types of collected assessment, monitoring, and evaluation data can be visualized to identify trends, patterns, and programmatic impact. For example, the Program for Resource Efficient Communities at the University of Florida collects energy use data from thousands of Florida homes and uses statistical packages and data visualization to identify effective energy efficiency measures. Utilizing these "big data" together with Extension clientele contacts could provide some interesting dashboards for stakeholders and the general public. Further, by generating charts and data at the county level, one could produce an infographic for every county in the state. By using some of these techniques and playing the role of "data broker," Extension can increase its relevance and worth to county governments and others.

To the best of our knowledge, very few land-grant universities use these interactive data visualization tools for their Extension efforts. The examples we discovered through a search of Tableau Public (free version of Tableau, <http://public.tableau.com>) and Google, or via word-of-mouth, focus mainly on reporting clientele contacts or demographics rather than evaluation data or impacts and are provided in Table 2.

**Table 2. Examples of Extension-related Data Dashboards and Visualizations**

Institution, Organization, or System	Description of Data Visualization Usage
University of New Hampshire's Logic Model Planning and Reporting System (LMPRS)	LMPRS is shared by six institutions and uses Tableau Public for some limited data visualization of accountability data such as clientele contacts and demographics.
Kansas State University's Program Evaluation and Reporting System (PEARS)	PEARS is used by 25 states for SNAP-Ed evaluation and includes a data dashboard (i.e., an organized display of data on a page or pages) and some geographic data mapping built into the system.
North Carolina State University Extension	Their "County Profile" approach nicely packages a lot of demographic, economic, and other data helpful in needs assessments into a set of interactive dashboards housed on Tableau Public.
UF/IFAS Program Development and Evaluation Center (PDEC)	PDEC has a profile on Tableau Public and has created several data visualizations over the years for various purposes including sharing survey results and data exploration.
National Institute of Food and Agriculture (NIFA)	The Leadership Management Dashboard is available through the NIFA Reporting Portal at <a href="https://portal.nifa.usda.gov">https://portal.nifa.usda.gov</a> , but it is limited to a single display of competitive and capacity grants by congressional district.
USDA Economic Research Service (ERS)	ERS has several visualizations, primarily related to ag economics, available online at <a href="https://www.ers.usda.gov/data-products/data-visualizations">https://www.ers.usda.gov/data-products/data-visualizations</a>

For those interested in building charts and dashboards with Extension data, both Tableau and Power BI offer free versions of their software and YouTube overflows with how-to videos for building dashboards using Microsoft Excel. The learning curve is high in using data visualization tools, especially in getting the data in the right format (known as data shaping and

modeling). When working with data visualization, the vast majority of effort will be collecting the data, cleaning it, and setting it up so that the tools work properly and efficiently.

The free versions of Tableau and Power BI have limited sharing capability, so wide use of the tools for your organization would likely require purchasing an upgraded version of the software. Besides the costs, which can be quite high in rolling out to a large university, data visualization requires much more participation from the users. For this reason, administrators and stakeholders often do not fully engage with the tools or use them regularly. The best way to increase stakeholder use of these tools is to make sure you deliver data that meets their needs, is simple to understand, and runs efficiently on the system (i.e., no long lag times to generate the charts).

Regardless of the software used to visualize data, there are many books and online resources available for advice and best practices in creating effective charts and graphs for reports, infographics, presentations, etc. (Evergreen, 2017; Knafllic, 2015). Stephanie Evergreen's website hosts a popular blog and a useful, interactive tool to rate the quality of a chart or graph based on her data visualization checklist (Evergreen, 2018a).

### **Conclusion**

Collecting and packaging data to show impact is no small task. Often, the work required is seen as too large a task, so it never gets done. However, the demand for demonstrating impacts is growing, and there is no evidence that this demand will subside anytime soon. Given this trend for more accountability, and in an environment known for tight budgets, the Extension components of land-grant universities need to work more efficiently to tell the Extension story. Many Extension organizations already collect data, such as clientele contacts, customer satisfaction surveys, and 4-H Common Measures, and can begin creating infographics or data dashboards using those data. The use of statewide Extension indicators, such as those created in Florida, offers a practical and relatively low-cost way to show impact across a state on critical issues. The continuous vetting and assessment of the data being reported by Extension professionals will help ensure that data quality will improve over time. Efforts in creating and distributing credible evidence can be leveraged by displaying data in multiple infographics, distributed through multiple channels. Extension organizations are encouraged to identify key programs or imminent legislative priorities to be promoted through the use of infographics.

Technology is rapidly changing, and the need for expertise in communication methods is more important than ever. Working with communications professionals to produce high quality, visual materials (including websites and infographics) is ideal, but there are several free tools available to help you get started even without this support. While the technical expertise that a communications staff can provide is invaluable, even more important is their proficiency in promoting and distributing your work. If no one sees the infographic, it does not matter how effective or beautiful it is. So also seek out those individuals in your organization who can help

you get your message out. Most likely, those same individuals also feel the pressure to provide the answers to the “So what?” questions, need data, and will be receptive to a collaborative effort.

This article highlights work being done at the University of Florida. Many land-grant institutions are also telling the Extension story through infographics, websites, data visualization, etc. To promote the sharing of infographics, the authors have created a website at <http://pdec.ifas.ufl.edu/credibleinfographics/> that houses examples of the different ways in which the Extension story may be told using credible and actionable evidence. We encourage you to send infographics you have developed and want to share with your peers to the contact email listed on the website. All of us can learn from each other—and what better way than visually.

### References

- Edelman. (2018). *2018 Edelman trust barometer*. Chicago, IL: Edelman. Retrieved from <http://www.edelman.com/trust-barometer>
- Ambargis, Z., Mead, C. I., Rzeznik, S. J.; Swenson, D., & Weisenberger, J. (2014). *Economic engagement framework: Economic impact guidelines*. Washington, DC: Association of Public and Land-grant Universities and Association of American Universities. Retrieved from <https://eric.ed.gov/?id=ED555635>
- Association of Governing Boards of Universities and Colleges. (2018). *Public confidence in higher education*. Washington, DC: Association of Governing Boards of Universities and Colleges. Retrieved from <https://www.agb.org/reports/2018/public-confidence-in-higher-education>
- Craig, D. & Bryant, T. (2017). *Orange County* [Infographic]. Gainesville, FL: University of Florida Institute of Food and Agricultural Sciences. Retrieved from <https://ifas.ufl.edu/media/ifasufledu/ifas-dark-blue/docs/orange.pdf>
- Craig, D. & Bryant, T. (2018). *2017 UF/IFAS Extension impacts and budget annual report* [Infographic]. Gainesville, FL: University of Florida Institute of Food and Agricultural Sciences. Retrieved from [https://ifas.ufl.edu/media/ifasufledu/ifas-dark-blue/docs/2017ExtensionCalendarAR\\_Final.pdf](https://ifas.ufl.edu/media/ifasufledu/ifas-dark-blue/docs/2017ExtensionCalendarAR_Final.pdf)
- Craig, D., Bryant, T., & Palmer, D. (2018a). *Empowering Florida’s communities to thrive* [Infographic]. Gainesville, FL: University of Florida Institute of Food and Agricultural Sciences. Retrieved from <https://pdec.ifas.ufl.edu/impacts/crd.pdf>
- Craig, D., Bryant, T., & Palmer, D. (2018b). *Improving Florida’s workforce, economy and natural resources* [Infographic]. Gainesville, FL: University of Florida Institute of Food and Agricultural Sciences. Retrieved from <https://pdec.ifas.ufl.edu/impacts/pesticidetraining.pdf>
- Craig, D., Bryant, T., & Palmer, D. (2018c). *Preparing Florida’s youth to be workforce ready* [Infographic]. Gainesville, FL: University of Florida Institute of Food and Agricultural Sciences. Retrieved from <https://pdec.ifas.ufl.edu/impacts/4hworkforce.pdf>

- Cutlip, S. M., Center, A. H., & Broom, G. M. (1985). *Effective public relations*. Englewood Cliffs, NJ: Prentice-Hall.
- Evergreen, S. D. H. (2017). *Effective data visualization: The right chart for the right data*. Thousand Oaks, CA: Sage.
- Evergreen, S. D. H. (2018a). *Data visualization checklist* [Interactive tool]. Retrieved from <https://datavizchecklist.stephanieevergreen.com/rate>
- Evergreen, S. D. H. (2018b). *Presenting data effectively: Communicating your findings for maximum impact* (2nd ed.). Los Angeles, 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://www.joe.org/joe/2013june/tt1.php>
- Franz, N., & Archibald, T. (2018). Four approaches to building Extension program evaluation capacity. *Journal of Extension*, 56(4), Article 4TOT5. Retrieved from <https://joe.org/joe/2018august/tt5.php>
- Franz, N. K., 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 <https://joe.org/joe/2014august/comm3.php>
- Franz, N., & Townson, L. (2008). The nature of complex organizations: The case of Cooperative Extension. In M. T. Braverman, M. Engle, M. E. Arnold, & R. A. Rennekamp (Eds.), *Program evaluation in a complex organizational system: Lessons from Cooperative Extension. New Directions for Evaluation*, 120, 5–14. San Francisco, CA: Jossey-Bass. doi:10.1002/ev.272
- Harder, A., Lamm, A., & Galindo, S. (2018). *Finding grant opportunities to support county Extension programs*. Gainesville, FL: University of Florida Institute of Food and Agricultural Sciences. Retrieved from <https://edis.ifas.ufl.edu/wc134>
- Harder, A., Moore, A., Mazurkewicz, M., & Bengé, M. (2013). Problems impacting Extension program quality at the county level: Results from an analysis of county program reviews conducted in Florida. *Journal of Extension*, 51(1), Article 1RIB2. Retrieved from <https://joe.org/joe/2013february/rb2.php>
- Hitlin, P. (2018). *Internet, social media use and device ownership in U.S. have plateaued after years of growth* [Blog post]. Washington, DC: Pew Research Center. Retrieved from <http://www.pewresearch.org/fact-tank/2018/09/28/internet-social-media-use-and-device-ownership-in-u-s-have-plateaued-after-years-of-growth>
- Hodges, A., Rahmani, M., & Court, C. (2018). *Florida agriculture: Fast facts 2018* [Brochure]. Gainesville, FL: University of Florida Institute for Food and Agricultural Sciences. Retrieved from [https://ifas.ufl.edu/media/ifasufledu/ifas-dark-blue/docs/pdf/impact/ICS\\_FloridaAgFactsBooklet2018.web.pdf](https://ifas.ufl.edu/media/ifasufledu/ifas-dark-blue/docs/pdf/impact/ICS_FloridaAgFactsBooklet2018.web.pdf)



- Israel, G. D. (2018). *Florida Cooperative Extension Service's customer satisfaction survey protocol*. Gainesville, FL: University of Florida Institute for Food and Agricultural Sciences. Retrieved from <https://pdec.ifas.ufl.edu/satisfaction/CSS%20Protocol%20-%202019.pdf>
- Jiang, X. (2016). *Designing and communicating trust: How nonprofits can use design to better communicate their trustworthiness* (Master's thesis). Retrieved from <https://pdfs.semanticscholar.org/cfbe/525b5a6b4d02a0bee2ac33837ad84992452a.pdf>
- Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. Hoboken, NJ: John Wiley & Sons.
- Krum, R. (2014). *Cool infographics: Effective communication with data visualization and design*. Indianapolis, IN: John Wiley & Sons.
- 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 <https://joe.org/joe/2013february/a1.php>
- Lankow, J., Ritchie, J., & Crooks, R. (2012). *Infographics: The power of visual storytelling*. Hoboken, NJ: John Wiley & Sons.
- Lowry, P. B., Wilson, D. W., & Haig, W. L. (2014). A picture is worth a thousand words: Source credibility theory applied to logo and website design for heightened credibility and consumer trust. *International Journal of Human-Computer Interaction*, 30(1), 63–93. doi:10.1080/10447318.2013.839899
- 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.
- Mathison, S. (2015). Seeing is believing: Using images as evidence in evaluation. 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. 157–176). Thousand Oaks, CA: Sage.
- McCue, T. J. (2013, January 8). Why infographics rule. *Forbes*. Retrieved from <https://www.forbes.com/sites/tjmccue/2013/01/08/what-is-an-infographic-and-ways-to-make-it-go-viral/#791758417272>
- Miller, R. L. (2015). How people judge the credibility of information. 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. 39–58). Thousand Oaks, CA: Sage.
- Mitchell, A., Gottfried, J., Barthel, M., & Shearer, E. (2016). *The modern news consumer: News attitudes and practices in the digital era*. Washington, DC: Pew Research Center. Retrieved from <http://www.journalism.org/2016/07/07/the-modern-news-consumer>
- Mulkey, D., & Hodges, A. (2018). *Using Implan to assess local economic impacts*. Gainesville, FL: University of Florida Institute of Food and Agricultural Sciences. Retrieved from <https://edis.ifas.ufl.edu/fe168>

- North, E. G. (2011). *Extension brand value: Lessons learned from the Copernicus study*. Presentation at the Southern Region Program Leaders Network, Fort Worth, TX. Retrieved from <https://slideplayer.com/slide/6935305>
- Olson, J. R., Welsh, J. A., & Perkins, D. F. (2015). Evidence-based programming within Cooperative Extension: How can we maintain program fidelity while adapting to meet local needs? *Journal of Extension*, 53(3), Article 3FEA3. Retrieved from <https://www.joe.org/joe/2015june/a3.php>
- Rennekamp, R. A., & Engle, M. (2008). A case study in organizational change: Evaluation in Cooperative Extension. In M. T. Braverman, M. Engle, M. E. Arnold, & R. A. Rennekamp (Eds.), *Program evaluation in a complex organizational system: Lessons from Cooperative Extension. New Directions for Evaluation*, 120, 15–26. doi:10.1002/ev.273
- Schwandt, T. A. (2015). Credible evidence of effectiveness: Necessary but not sufficient. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *Credible and actionable evidence: The foundation for rigorous and influential evaluations* (2nd ed., pp. 259–273). Thousand Oaks, CA: Sage.
- Smiciklas, M. (2012). *The power of infographics: Using pictures to communicate and connect with your audiences*. Indianapolis, IN: Que Publishing.
- Tufte, E. R. (2006). *Beautiful evidence*. Cheshire, CT: Graphics Press.
- University of Florida Institute of Food and Agricultural Sciences (UF/IFAS). (2019). *Public investment in UF/IFAS yields significant economic benefits and jobs, 2018* [Brochure]. Retrieved from [https://ifas.ufl.edu/media/ifasufledu/ifas-dark-blue/docs/ROI\\_Booklet2019.Web.pdf](https://ifas.ufl.edu/media/ifasufledu/ifas-dark-blue/docs/ROI_Booklet2019.Web.pdf)
- Urban, J. B., & Trochim, W. (2009). The role of evaluation in research-practice integration: Working toward the “Golden Spike.” *American Journal of Evaluation*, 30(4), 538–553. doi:10.1177/1098214009348327
- U.S. Department of Agriculture, National Institute of Food and Agriculture (USDA-NIFA). (2011). *National outcomes and indicators: AREERA plan of work reporting system*. Retrieved from <https://nifa.usda.gov/sites/default/files/resource/Natl%20Outcomes%20%26%20Indicator%20Listing.pdf>
- U.S. Department of Agriculture, National Institute of Food and Agriculture (USDA-NIFA). (2015). *Generic logic model for NIFA reporting*. Retrieved from <https://nifa.usda.gov/resource/generic-logic-model-nifa-reporting>
- Wiles, J. (2017). *How to choose the right communications channel* [Blog post]. Stamford, CT: Gartner. Retrieved from <https://www.gartner.com/smarterwithgartner/corporate-communications-four-steps-to-choosing-the-right-communication-channel>

*Diane D. Craig* is a research/data analyst with the Program Development and Evaluation Center in the department of Agricultural Education and Communication at the University of Florida

Institute of Food and Agricultural Sciences. Her areas of expertise include accountability reporting and higher education benchmarking.

*Ruth Hohl Borger* is the Assistant Vice president for UF/IFAS Communications at the University of Florida. She provides leadership in strategic communications and marketing, including reputation management, media relations, stakeholder communications, digital communications video, publications, and production support.

## How Evaluation Capacity Building Grows Credible and Actionable Evidence for Cooperative Extension Programs

Chelsea Hetherington

Cheryl Eschbach

Courtney Cuthbertson

Michigan State University Extension

*Evaluation capacity building (ECB) is an essential element for generating credible and actionable evidence on Extension programs. This paper offers a discussion of ECB efforts in Cooperative Extension and how such efforts enable Extension professionals to collect and use credible and actionable evidence on the quality and impacts of programs. Sufficient investments in ECB, both at the individual and organizational levels, can better equip Extension to advocate for and make changes to programs, advance as a learning organization, and have a more powerful impact on communities. Furthermore, as Extension program stakeholders often have varying perspectives on the credibility of evidence, these perspectives must also be accounted for in efforts to build Extension's evaluation capacity. Intentional investments in ECB efforts provide an opportunity for Extension to further deepen and expand impact, positioning programs to most effectively and positively benefit individuals and communities.*

**Keywords:** evaluation capacity building, credible evidence, evaluation use, evaluation stakeholders

***“For apart from inquiry, apart from the praxis, individuals cannot be truly human. Knowledge emerges only through invention and re-invention, through the restless, impatient, continuing, hopeful inquiry human beings pursue in the world, with the world, and with each other.”***

—Paulo Freire (1970)

### Introduction

The Cooperative Extension Service (Extension) is charged with delivering research-based educational programs that positively benefit individuals and communities. As a result, Extension is expected to provide credible, actionable evidence on the quality and impacts of its programs (Franz & Archibald, 2018; Taylor-Powell & Boyd, 2008). Because stakeholders hold diverse expectations on what constitutes credible and actionable evidence, because Extension staff are not typically experts in program evaluation, and because contexts, conditions, and criteria for

---

Direct correspondence to Chelsea Hetherington at [hether13@msu.edu](mailto:hether13@msu.edu)

demonstrating credible program evidence are complex and changing, evaluation capacity building (ECB) is critical to Extension's organizational and professional development efforts (Taylor-Powell & Boyd, 2008). ECB refers to the intentional efforts to both build and sustain an organization's ability to conduct quality, credible evaluations, including factors such as instrumental evaluation support, evaluation resources, and a broader organizational context that supports meaningful program evaluation (Preskill & Boyle, 2008; Stockdill, Baizerman, & Compton, 2002; Taylor-Powell & Boyd, 2008).

In this paper, we discuss how ECB contributes to the collection and use of credible and actionable evidence of Extension program quality and impacts. By directing efforts towards building organizational evaluation capacity across individual and organizational levels (Taylor-Powell & Boyd, 2008), Extension professionals in all roles can be equipped with the skills needed to collect and interpret *credible* data—that is, data that provide trustworthy, compelling evidence of a program's quality or impact (Donaldson, Christie, & Mark, 2015). Moreover, different aspects of and perspectives about credibility can play a role in impacting ECB efforts (e.g., credibility to communities served, to external funders and policymakers, internally to Extension professionals themselves and program administrators, to the broader university or scientific community). Such differing aspects of credibility should be considered in ECB efforts. Finally, the credibility of evidence collected on Extension programs can be impacted, both positively and negatively, by individual ECB approaches and organizational factors (Preskill & Boyle, 2008; Taylor-Powell & Boyd, 2008).

### **Background on Evaluation Capacity Building and Credibility in Extension**

Different pressures engender the need for credible evidence about Extension programs. The broad mission of Cooperative Extension is to provide research-based educational opportunities to individuals and communities, supporting these individuals and communities in meeting the unique challenges that they experience (Franz & Townson, 2008). This mission underlies the activities and expected outcomes of Extension programming. Extension professionals face the challenge of translating scientific evidence in ways that individuals can use, often with limited resources or tools to assess community needs and program impacts.

Nevertheless, as a public-serving organization that receives funding from federal, state, county, and other sources, having accountability to external stakeholders is a key component of Extension work (Franz & Townson, 2008). The collection and use of credible and actionable evidence are crucial in establishing accountability with these stakeholders and in demonstrating the impact of Extension programs.

At a basic level, organizational reporting and evaluation requirements typically demand that Extension professionals demonstrate to stakeholders that audiences were reached and impacted in the ways designated by the program (Baughman, Boyd, & Franz, 2012). More broadly, Extension programs must demonstrate the relevance and impact of Extension work in ways that

will be credible to stakeholders with distinct needs, interests, and perspectives. Indeed, stakeholders often have varying criteria for what they deem to be credible, trustworthy evidence of a program's impact or quality. Individual standards and interests can drive differences in what is seen as credible by different stakeholder groups. Though these criteria can overlap, Extension programs must balance satisfying demands of credibility to communities served, internal Extension administrators, the broader professional and scientific community, those stakeholders providing funding for programs, and in the delivery of programs to participants, among others. Examples of criteria for each of these program stakeholder groups are shown in Table 1.

**Table 1. Criteria for Credibility Among Extension Stakeholder Groups**

<b>Stakeholder Group</b>	<b>Sample Criteria for Credible Evidence</b>
Communities served	<ul style="list-style-type: none"> <li>• Programming addresses a need or gap</li> <li>• Satisfactory program delivery</li> <li>• Evidence of program effectiveness (quantitative or qualitative, may vary depending on community)</li> </ul>
Internal Extension administrators	<ul style="list-style-type: none"> <li>• Outputs, or numbers served</li> <li>• Quantitative evidence of program outcomes</li> <li>• Data-driven program planning</li> </ul>
Scientific community	<ul style="list-style-type: none"> <li>• Evidence-based programming</li> <li>• Experimental or quasi-experimental designs</li> <li>• Rigorous methods</li> <li>• Scholarly publications</li> <li>• Peer-reviewed conference presentations</li> </ul>
Program funders	<ul style="list-style-type: none"> <li>• Outputs, or numbers served</li> <li>• Quantitative evidence of program outcomes</li> <li>• Rigorous evaluation methods (may vary depending on the funder)</li> </ul>

Groups may also have differing standards for the varied aspects of credibility that can impact data quality, such as utility, relevance, generalizability, and objectivity (Radhakrishna, Tobin, Brennan, & Thomson, 2012). There may even be variations in what individuals within a specific group will deem credible or trustworthy. Because Extension professionals are charged with translating and disseminating the scientific work of land-grant universities to communities (Franz & Townson, 2008), they must balance meeting the needs of communities in addressing local challenges and the needs of other program stakeholders, while remaining grounded in research- and evidence-based programming (Olson, Welsh, & Perkins, 2015). Indeed, the Extension mission is best served when programs bridge the gap between implementing rigorous research models and meeting local community needs (Fetsch, MacPhee, & Boyer, 2012).

### **Credibility in Evaluation Capacity Building Efforts**

Professional development efforts centered around ECB should serve to better facilitate Extension professionals' understanding of best practices in serving clients, delivering and evaluating educational content, and sharing program quality and outcome metrics with stakeholders. Appropriately delivering effective Extension programs is contingent upon Extension

professionals' capacity to both use and generate credible evidence of the quality and effectiveness of those programs. Extension systems are complex with differing levels of faculty and staff knowledge and experience (Franz & Townson, 2008), with many staff being experts in specific content areas, rather than experts in research methodologies or evaluation processes (Arnold, 2006). As Extension professionals are charged with planning, delivering, and evaluating research-based educational programs that benefit individuals and communities, ensuring that programs are adequately serving those individuals and communities requires a continuous investment in organizational and professional development efforts that support this aim (Taylor-Powell & Boyd, 2008). The size and complexity of Extension work require ECB to design, deliver, and evaluate evidence-based, impactful educational programming, making ECB a key building block in establishing the credibility of evidence supporting Extension programs.

As shown in Table 1, stakeholders' varied notions of credibility relate to ECB in several ways. ECB approaches should help Extension professionals understand these different aspects of credibility in evaluation data. ECB initiatives should incorporate efforts to support Extension professionals to consider and understand perspectives about the credibility of evidence for their specific program stakeholders in planning programs and evaluations. For instance, if an individual Extension professional plans a rigorous, pre-/post-test evaluation of a program to satisfy grant requirements, yet their local county stakeholders are more interested in qualitative narratives around a program's impact, the credibility of those findings will fail to meet the expectations of that local stakeholder group. Given that having evaluation findings that meet stakeholders' evidence needs can be highly motivating for Extension professionals (Guion, Boyd, & Rennekamp, 2007), this is an important dimension to consider and incorporate into ECB initiatives. Extension evaluation specialists can support Extension professionals in generating the most appropriate credible evidence for their specific stakeholder group(s). For instance, in the previous example, an evaluation specialist might advise collecting both pre-/post-test data and qualitative narratives to satisfy the needs and interests of both stakeholder groups.

### **Evaluation Use and Credibility**

The meaningful use of evaluation data can bolster credibility in several ways and should be a focus of efforts to build evaluation capacity. Use of evaluation data can increase the credibility of the evaluation process with Extension professionals (Lamm & Israel, 2013). Use of evaluation data helps Extension professionals see the value of collecting credible evidence and can improve the quality of data that is collected, such that staff become more invested in the quality of the data that they are collecting and using (Baughman et al., 2012). Collecting and using evidence that is methodologically sound enables staff to be confident in using data that reports program impacts or quality, whether using such data to advocate for programs with key stakeholders or making changes to improve programs. Through the use of evaluation data, Extension professionals can better see the value of collecting such evidence, which can improve the organization's future capacity for collecting and using such evidence (Baughman et al., 2012).

Use of evaluation data can also empower Extension professionals to be more fully engaged in the programming process (Patton, 2008). Rather than simply collecting data that are aggregated into statewide outcomes, Extension professionals' own use of evaluation data engages them in intentionally thinking about the information they need to advocate for or make changes to their programs. Such efforts engage Extension professionals in evaluative thinking and can support the creation of an organizational culture of learning (Buckley, Archibald, Hargraves, & Trochim, 2015). ECB efforts will take root in creating a learning culture when staff see evaluation as part of their daily work and a critical component of the program planning and implementation process (Fetterman, 2003; King, 2007).

Meaningful evaluation use can also bolster credibility with stakeholders, establishing the integrity of Extension as an organization that has the capacity to make a difference in the lives of youth, families, and the broader public (Franz, 2015). Collecting and using evaluation data to change and improve programs establishes Extension's integrity as an organization that values continuous improvement and works to refine its educational offerings. When findings are actually used to make changes to or improve a program, program participants see that their voices have been heard on a deeper level. This can increase their stake in the program as well as their view of Extension as a credible provider of educational programs. Thus, use of evaluation data can be a clear asset in using ECB to generate credible evidence about Extension programs.

### **Evaluation Competencies for Extension Professionals**

In seeking to build any organization's capacity to generate and use credible and actionable evidence, one must hone in on the specific competencies required and expected of individual staff. What does an Extension professional need to know in order to credibly deliver and evaluate educational programming? The level and type of competencies required of individual Extension professionals will vary greatly depending on the type of program being evaluated (Franz & Archibald, 2018), the overall organizational structure (Taylor-Powell & Boyd, 2008; Preskill & Boyle, 2008), and Extension stakeholders' perspectives on credibility (Bryson, Patton, & Bowman, 2011; Johnson et al., 2009). For instance, within the framework proposed by Franz and Archibald (2018), ECB efforts will be most successful when they align with the specific educational initiative that the Extension professional is seeking to evaluate. That is, if an Extension professional is engaging in educational programming that entails content transmission, competencies should focus on knowing how to collect credible evidence that measures the program's effectiveness in increasing participants' knowledge and changing their behavior. Similarly, for service-focused efforts, competencies would align more closely with knowing how to assess participant satisfaction with the program.

Competencies expected of individual Extension professionals will also depend greatly on the expectations for their roles within the specific Extension system. For instance, in some Extension systems, Extension professionals with academic appointments are expected to engage in scholarship and publish research-based evidence of their work (Taylor-Powell & Boyd, 2008).



In other Extension systems, Extension professionals meet scholarship expectations in different ways, such as through curriculum creation or securing external grant funding (Franz & Townson, 2008). The presence of expectations for Extension professionals to engage in scholarly research around their programs will shape what ECB efforts look like within that particular Extension system (e.g., the level of rigor needed in data collection plans, the size and scope of projects, the level of data analysis skills needed).

### Specific Evaluation Competencies for Extension Work

Extension professionals need a number of general evaluation competencies in order to generate and use credible, actionable evidence. Extension professionals should be able to: 1) use data to assess the needs of the communities they serve, whether through data they themselves collect or data collected by others; 2) develop and implement credible programs to fidelity (i.e., develop and implement programs with a basis in research and/or with some evidence of impact and attend to the necessary implementation criteria); 3) collect credible data (i.e., the types of data that stakeholders see as credible) on the impact, quality, and fidelity of the program; and 4) use acquired data to inform changes to the program and convince stakeholders of the program's value. Specific evaluation competencies within these areas are shown in Table 2.

**Table 2. Evaluation Competencies for Generating and Using Credible, Actionable Evidence**

Step in Program Process	Specific Evaluation Competencies
1) Needs assessment	<ul style="list-style-type: none"> <li>• Finding existing data</li> <li>• Collecting new data (e.g., via surveys or focus groups)</li> <li>• Analyzing quantitative and/or qualitative data</li> <li>• Learning and understanding stakeholder perspectives</li> <li>• Interpreting data</li> <li>• Using data to inform decisions</li> </ul>
2) Program design and implementation	<ul style="list-style-type: none"> <li>• Developing programs based on stakeholder needs</li> <li>• Creating logic models</li> <li>• Writing program objectives</li> <li>• Assessing a topic's research base</li> <li>• Interpreting the evidence base for existing programs</li> </ul>
3) Evaluation data collection	<ul style="list-style-type: none"> <li>• Designing program evaluations</li> <li>• Selecting and designing data collection methods (e.g., survey design)</li> <li>• Collecting data (e.g., via surveys or focus groups)</li> <li>• Considering ethics of evaluation data collection</li> </ul>
4) Evaluation use	<ul style="list-style-type: none"> <li>• Analyzing quantitative and/or qualitative data</li> <li>• Interpreting data</li> <li>• Using data to inform decisions</li> <li>• Learning and understanding stakeholder perspectives</li> <li>• Creating and sharing evaluation reports</li> <li>• Interpreting and presenting results to stakeholders</li> </ul>

Admittedly, this ideal list of competencies will not always translate to the realities of Extension programming overall or to every individual Extension program. As noted earlier, the specific competencies most crucial to an individual Extension professional will be directed by the goals of the specific educational program being evaluated (Franz & Archibald, 2018). Still, these remain key competencies for Extension professionals. Though the specific responsibilities of Extension professionals can vary across systems, the demands of Extension work, where Extension professionals should know how to adapt to the needs of communities, deliver credible, quality programs, collect credible program data, and subsequently use that data, require that these competencies be present.

The nature of Extension programs is such that evaluation data is frequently collected and used for reporting, accountability, or persuasive purposes; that is to inform supervisors, funding agencies, legislators, and other stakeholders about the impact of a program (Baughman et al., 2012). As a result, ECB efforts may increasingly focus on competencies that serve this need, such as collecting impact data, while neglecting other competencies, such as collecting data on community needs, strengths, and opportunities.

Extension professionals should be sufficiently trained in all aspects of the program process (Arnold, 2006). This can include trainings around the development and use of logic models to meet community needs, including supporting Extension professionals in identifying specific program outcomes and indicators (Arnold, 2006). This has the added benefit of grounding a subsequent program evaluation in the program's logic model. By including the program development process in ECB efforts, we continue to advance Extension's capacity to deliver credible educational efforts and collect credible evidence of program quality and impacts.

ECB efforts also present an opportunity for administrators to build their own evaluation competencies. Individual Extension administrators can benefit from increased evaluation capacity by advancing their use of data to inform decisions and advocate for programs with key stakeholders. Extension administrators should also understand what is required for Extension professionals to deliver and evaluate quality programs if they are to support Extension program staff in accomplishing this goal. Extension administrators play a key role in making decisions about Extension programs, setting organizational priorities, creating expectations for staff members' programming and evaluation efforts, and advocating for Extension programs with stakeholders. Extension administrators have the ability to influence policies and practices around evaluation at an organizational level, such as through rewarding good evaluation practice or providing staff sufficient time and training to collect and use credible evidence (Boyd, 2009; Silliman, Crinion, & Archibald, 2016). As such, it is essential that those who serve in administrative roles understand the components of the program delivery and evaluation process and how crucial these competencies are to implementing quality programming.

Most crucially, for Extension professionals across all levels and responsibilities, evaluative thinking is an overarching, critical competency (Buckley et al., 2015). Evaluative thinking refers to one's ability to critically reflect on programs, people, and processes for effective change (Buckley et al., 2015). A crucial aspect of ECB in Extension is that it can support Extension professionals in engaging in such critical thinking about evaluation, and particularly in interpreting and weighing the *quality* of evaluation evidence (Preskill & Russ-Eft, 2016). In discussing programs and program evidence data with stakeholders, Extension professionals should be able to adequately interpret program evidence, making evaluative thinking key in advocating for Extension programs.

### **Assessing and Meeting Community Needs**

As noted previously, a key competency in the program development and delivery process is using data to assess the needs of communities and then delivering programming to meet those needs. Beyond assessing and meeting the needs of communities traditionally served by Extension programs, Extension professionals must also be equipped to meet the needs of communities that have traditionally been disenfranchised (and at times, outright excluded) from Extension programs—for instance, black and indigenous communities (e.g., Harris, 2008; Schor, 1986). Extension educational programs cannot take a “one size fits all” approach, assuming that existing programs will meet the needs of, have an impact on, or be credible to all communities. We cannot assume that what is effective or credible in communities we traditionally serve will also be effective or credible in others. Indeed, understanding the cultural or social contexts in an evaluation (e.g., stakeholders' perspectives on credibility, culturally responsive methodologies) is increasingly being recognized as a critical component of the program planning and evaluation process (Centers for Disease Control and Prevention; 2014). Failing to take a cultural or contextual lens in delivering or evaluating programs can threaten the utility, quality, and credibility of said programming and data, and troublingly, can even cause harm to the communities being served (Bowman, Dodge Francis, & Tyndall, 2015). Extension professionals must be equipped with the knowledge and skills to competently serve all of the residents of their communities, not just those who have historically or traditionally been served by Extension. As such, ECB should incorporate efforts focused on developing the skills needed to meet community needs (e.g., training in needs assessments, attending to diverse stakeholder perspectives, culturally relevant evaluation practices).

### **Barriers to Evaluation Capacity Building on the Path to Credible Evidence**

Many factors and the realities of Extension work can compete with ECB efforts and hinder the generation of credible evidence. Extension professionals are frequently asked to accomplish more with less time, fewer resources, and smaller teams; the challenges of finding time to devote to professional development on any topic presents a challenge for Cooperative Extension as a whole (Arnold, 2006). Further, many program staff across disciplines find evaluation to be an intimidating or anxiety-provoking topic (Arnold, 2006; Donaldson, Gooler, & Scriven, 2002),

which may increase their reluctance to learn more about it. Still, when they do engage in professional development opportunities around evaluation, this fear can be ameliorated (Kelsey, 2008).

Some Extension professionals may see the development of skills in program evaluation as a lower priority than other professional development needs (Arnold, 2006). For example, they may see evaluation as someone else's responsibility, might feel they do not have time to engage in evaluation, may not see any personal value to themselves (i.e., intrinsic or extrinsic rewards) for evaluating programs, or might have a certain level of anxiety about engaging in evaluation. Extension professionals might also experience a lack of relative frequency with which they have the opportunity to exercise a diversity of evaluation skills, which might further inhibit learning. For instance, Extension professionals may have frequent opportunities to exercise survey data collection skills using existing survey instruments, while opportunities to actually write and create survey instruments may occur less frequently. Furthermore, the prospect of engaging in a rigorous program evaluation may seem daunting to someone with minimal background in research or evaluation methods, while conducting more cursory evaluations might seem tedious with little payoff. This list of barriers is included not to dissuade from ECB efforts in Extension, but rather to recognize and address these factors when implementing successful ECB initiatives.

### **Organizational Context and Expectations**

For Extension organizations seeking to collect and use credible and actionable evidence of program quality and impacts, ECB is a critical component of organizational development. Beyond individual Extension professionals, the Extension organization as a whole must also provide an organizational context that supports the collection and use of credible evidence. This means that Extension organizational leadership should support, encourage, and reward efforts made by Extension professionals to implement and evaluate programs with credibility (Preskill & Boyle, 2008). Though individual ECB efforts may be successful in developing an individual Extension professional's competencies in the short term (e.g., providing technical assistance or in-service training to a small group of Extension professionals), having a sustained culture that generates credible and actionable evidence hinges on an organizational context that adequately supports ECB and credible, quality program evaluation (Preskill & Boyle, 2008; Taylor-Powell & Boyd, 2008).

An organizational culture that supports learning from credible evidence can subsequently bolster evaluative thinking on an organizational level (Taylor-Powell & Boyd, 2008). Administrative buy-in and organizational support are critical factors in order for ECB efforts with program staff to take root in supporting Extension as a learning organization that both generates and values credible evidence (Boyd, 2009; Preskill & Boyle, 2008; Taylor-Powell & Boyd, 2008). In an organization where administrators endorse and model using evaluation findings, credible and high-quality data will be seen as an organizational priority (Preskill & Boyle, 2008). Even if Extension professionals themselves do not directly use the data they collect, seeing such data

used by the organization (e.g., by colleagues for program improvement or organizationally in marketing programs) can improve morale and increase evaluation capacity.

### **How Organizations Can Support Evaluation Capacity Building**

Extension professionals and programs must have adequate instrumental support for evaluation efforts, including adequate staff devoted to evaluation efforts and adequate financial support for data collection, entry, and analysis (Taylor-Powell & Boyd, 2008). Individual Extension professionals should also have adequate time to devote to planning, collecting, and using credible evidence (Preskill & Boyle, 2008). As in any organization, effectively building evaluation capacity in Extension requires clear expectations for evaluation among Extension professionals (Preskill & Boyle, 2008; Volkov & King, 2007). Individual Extension professionals should be given clear requirements, guidelines, and expectations for evaluation efforts. In order for Extension to successfully build and use quality, credible, and actionable evidence about programs, Extension must remain dedicated to devoting ample time, staff, money, and resources to evaluation capacity building.

Organizationally, evaluation capacity can also be built through the identification, training, and fostering of evaluation champions, or Extension professionals who show an interest in conducting and supporting quality evaluation work (Silliman et al., 2016; Taylor-Powell & Boyd, 2008). By creating a community of evaluation champions, where groups of Extension professionals engage in evaluative thinking, Extension evaluation specialists can create opportunities for professional development with individuals who can share what they have learned with their peers. Extension administrators can also be leveraged as evaluation champions through the support of organizational efforts to build evaluation capacity and generate credible evidence (Boyd, 2009).

Organizations can also support ECB through adequate staffing of Extension evaluation specialists (Taylor-Powell & Boyd, 2008). Evaluation specialists can provide trainings and technical assistance in evaluation, either through face-to-face support or online trainings (Taylor-Powell & Boyd, 2008). If a state Extension organization does not have an evaluation specialist on staff, those organizations can support ECB by having an external evaluator, either an Extension evaluation specialist from another state or from outside of Extension, lead professional development trainings about program evaluation.

Organizations can also develop written and electronic evaluation resources for staff (Taylor-Powell & Boyd, 2008). Online professional development opportunities are especially of interest to Extension professionals (Senyurekli, Dworkin, & Dickinson, 2006), suggesting that Extension would benefit from the use of technology in ECB efforts. In Michigan, Extension evaluation specialists recently collaborated with an Extension instructional design specialist to develop a self-paced, online course on evaluation (Hetherington, Eschbach, Cuthbertson, & Shelle, 2018). With the support of the instructional design specialist, the course was designed to follow best

practices in digital education, including the use of intentionally brief video lectures, interactive activities, and badging to reward participation. Creating a standardized set of modules that can introduce staff to evaluation concepts forms a purposeful, structured socialization into the evaluation process. This is beneficial not only because it saves time that might be taken up by individual staff consultations, but also because it allows incoming staff to be introduced to evaluation concepts in a standardized manner (King, 2007).

Organizations can also support ECB by housing evaluation specialists with content expertise in specific program areas (e.g., an evaluator with expertise in child and youth development being designated to work with 4-H staff). Having an evaluator with content expertise can be an asset to ECB, in that this can increase Extension professionals' own trust and willingness to engage in the evaluation as well as provide expertise in methodologies or measures specific to that content area. Evaluation specialists with content area expertise can evaluate relevant programs with a greater degree of depth than those with content expertise in other areas, which can further create opportunities to engage in scholarship (e.g., publishing evaluating findings in peer-reviewed journals). By designating specific evaluators to work within specific content areas and program teams, this can further support the collection of credible evidence as Extension evaluators provide instrumental support on specific evaluation projects.

### Conclusion

Evaluation capacity building (ECB) is a foundational aspect of building credible and actionable evidence about the quality and effectiveness of Extension programs. ECB efforts can bolster Extension professionals' understanding of the program development, implementation, and evaluation processes, thus advancing Extension's ability to generate and use credible evidence. Extension professionals must be equipped to understand how to collect credible evidence about program impacts and to consider varying stakeholder perspectives on what constitutes credible evidence. Complex organizations only reap the benefits of ECB efforts as far as what they put into it, and Extension is no different. When Extension builds its evaluation capacity, it not only builds the capacity to collect high quality data, it also builds the ability to use such data in advocating for and making changes to improve programs, increases the ability to advance as a learning organization, and supports Extension's ability to have a positive impact on individuals and the communities in which they live.

### References

- Arnold, M. E. (2006). Developing evaluation capacity in Extension 4-H field faculty: A framework for success. *American Journal of Evaluation*, 27(2), 257–269.  
doi:10.1177/1098214006287989

- Baughman, S., Boyd, H. H., & Franz, N. K. (2012). Non-formal educator use of evaluation results. *Evaluation and Program Planning*, 35(3), 329–336. doi:10.1016/j.evalprogplan.2011.11.008.
- Bowman, N. R., Dodge Francis, C., & Tyndall, M. (2015). Culturally responsive indigenous evaluation: A practical approach for evaluating indigenous projects in tribal reservation contexts. In S. Hood, R. Hopson, & H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 335–359). Charlotte NC: Information Age Publishing.
- Boyd, H. H. (2009). Practical tips for evaluators and administrators to work together in building evaluation capacity. *Journal of Extension*, 47(2), Article 2IAW1. Retrieved from <https://www.joe.org/joe/2009april/iw1.php>
- Bryson, J. M., Patton, M. Q., & Bowman, R. A. (2011). Working with evaluation stakeholders: A rationale, step-wise approach and toolkit. *Evaluation and Program Planning*, 34(1), 1–12. doi:10.1016/j.evalprogplan.2010.07.001.
- Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation*, 36(3), 375–388. doi:10.1177/1098214015581706
- Centers for Disease Control and Prevention. (2014). *Practical strategies for culturally competent evaluation*. Atlanta, GA: US Dept of Health and Human Services. Retrieved from [https://www.cdc.gov/dhdsp/docs/cultural\\_competence\\_guide.pdf](https://www.cdc.gov/dhdsp/docs/cultural_competence_guide.pdf)
- 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.
- Donaldson, S. I., Gooler, L. E., & Scriven, M. (2002). Strategies for managing evaluation anxiety: Toward a psychology of program evaluation. *American Journal of Evaluation*, 23(3), 261–273. doi:10.1177/109821400202300303
- Fetsch, R. J., MacPhee, D., & Boyer, L. K. (2012). Evidence-based programming: What is a process an Extension agent can use to evaluate a program's effectiveness. *Journal of Extension*, 50(5), Article 5FEA2. Retrieved from <https://www.joe.org/joe/2012october/a2.php>
- Fetterman, D. (2003). Fetterman-House: A process use distinction and a theory. *New Directions for Evaluation*, 97, 47–52. doi:10.1002/ev.74
- Franz, N. K. (2015). Programming for the public good: Ensuring public value through the Cooperative Extension program development model. *Journal of Human Sciences and Extension*, 3(2), 13–25.
- Franz, N., & Archibald, T. (2018). Four approaches to building Extension program evaluation capacity. *Journal of Extension*, 56(4), Article 4TOT5. Retrieved from <https://joe.org/joe/2018august/tt5.php>

- Franz, N., & Townson, L. (2008). The nature of complex organizations: The case of Cooperative Extension. In M. T. Braverman, M. Engle, M. E. Arnold, & R. A. Rennekamp (Eds.), *Program evaluation in a complex organizational system: Lessons from Cooperative Extension. New Directions for Evaluation*, 120, 5–14. doi:10.1002/ev.272
- Freire, P. (1970). *Pedagogy of the oppressed*. New York, NY: Continuum Books.
- Guion, L., Boyd, H., & Rennekamp, R. (2007). An exploratory profile of Extension evaluation professionals. *Journal of Extension*, 45(4), Article 4FEA5. Retrieved from <https://www.joe.org/joe/2007august/a5.php>
- Harris, C. V. (2008). "The Extension Service is not an integration agency": The idea of race in the Cooperative Extension Service. *Agricultural History*, 82(2), 193–219. doi:10.3098/ah.2008.82.2.193
- Hetherington, C., Eschbach, C., Cuthbertson, C., & Shelle, G. (2018, November). *Evaluation capacity building in Extension using online instructional technology*. Roundtable presented at the Annual American Evaluation Association Conference, Cleveland, Ohio.
- Johnson, K., Greenseid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, 30(3), 377–410. doi:10.1177/1098214009341660
- Kelsey, K. D. (2008). Do workshops work for building evaluation capacity among Cooperative Extension Service faculty? *Journal of Extension*, 46(6), Article 6RIB4. Retrieved from <https://www.joe.org/joe/2008december/rb4.php>
- King, J. A. (2007). Developing evaluation capacity through process use. *New Directions for Evaluation*, 116, 45–59. doi:10.1002/ev.242
- 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.
- Olson, J. R., Welsh, J. A., & Perkins, D. F. (2015). Evidence-based programming within Cooperative Extension: How can we maintain program fidelity while adapting to meet local needs? *Journal of Extension*, 53(3), Article 3FEA3. Retrieved from <https://www.joe.org/joe/2015june/a3.php>
- Patton, M. Q. (2008). *Utilization-focused evaluation*. Thousand Oaks, CA: Sage.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443–459. doi:10.1177/1098214008324182
- Preskill, H., & Russ-Eft, D. (2016). *Building evaluation capacity: 72 activities for teaching and training*. Thousand Oaks, CA: Sage.
- Radhakrishna, R., Tobin, D., Brennan, M., & Thomson, J. (2012). Ensuring data quality in Extension research and evaluation studies. *Journal of Extension*, 50(3), Article 3TOT1. Retrieved from <https://www.joe.org/joe/2012june/tt1.php>
- Schor, J. (1986). The black presence in the US Cooperative Extension Service since 1945: An American quest for service and equity. *Agricultural History*, 60(2), 137–153.



- Senyurekli, A. R., Dworkin, J., & Dickinson, J. (2006). On-line professional development for Extension educators. *Journal of Extension*, 44(3), Article 3RIB1. Retrieved from <https://www.joe.org/joe/2006june/rb1.php>
- Silliman, B., Crinion, P., & Archibald, T. (2016). Evaluation champions: What they need and where they fit in organizational learning. *Journal of Human Sciences and Extension*, 4(3), 22–44.
- Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation*, 93, 7–26. doi:10.1002/ev.39
- Taylor-Powell, E., & Boyd, H. H. (2008). Evaluation capacity building in complex organizations. *New Directions for Evaluation*, 120, 55–69. doi:10.1002/ev.276
- Volkov, B. B., & King, J. A. (2007). *A checklist for building organizational evaluation capacity*. Retrieved from [http://dmeforpeace.org/sites/default/files/Volkov%20and%20King\\_Checklist%20for%20Building%20Organizational%20Evaluation%20Capacity.pdf](http://dmeforpeace.org/sites/default/files/Volkov%20and%20King_Checklist%20for%20Building%20Organizational%20Evaluation%20Capacity.pdf)

*Chelsea Hetherington* is an Evaluation and Research Specialist at Michigan State University Extension, specializing in research and evaluation of child and youth development programs.

*Cheryl Eschbach* is the Director of the Health and Nutrition Institute at Michigan State University Extension. She previously worked as an Evaluation Specialist for MSU Extension programs.

*Courtney Cuthbertson* is the Community Behavioral Health Specialist at Michigan State University Extension. They previously worked as an Evaluation Specialist for MSU Extension programs.

## **Where Do We Go from Here?: Credible and Actionable Evidence in Extension**

**Scott R. Cummings**

*Texas A&M AgriLife Extension Service*

**Benjamin Silliman**

*North Carolina State University*

*The evolution of evaluation in the Cooperative Extension System (Extension) has gone through many changes over the years, from focusing on participation, to the measurement of outcomes, and then impacts. Now, the new evolution in Extension is the use of credible and actionable evidence. This special edition of the Journal of Human Sciences and Extension (JHSE) explored the theme, “What is credible and actionable evidence in Extension programs?” The authors of the articles in this issue wrote about the important concepts ahead of us as we begin on the road to more credible and actionable evidence. This article provides some closing thoughts on this special issue and sets forth challenges as we move forward.*

*Keywords:* credible evidence, actionable evidence, evaluation, Cooperative Extension, stakeholders

***“Trust is built on credibility, and credibility comes from acting in others’ interests before your own.”***

**—Stephen Denny**

### **Prelude**

Our professional colleagues, especially those in program evaluation, provide a rich dialogue on the dimensions and implications of credible and actionable evidence in policy, programs, and personal decisions (Donaldson, Christie, & Mark, 2015). Among consumers, court jurors, or citizens in general, critical and conscientious decision-makers generally affirm the need for “the best available evidence,” whether pursuing their own interests or contributing to deliberative democracy. As pragmatists—and there are both opportunities and dangers in pragmatism—Americans are interested not just in trustworthy and relevant evidence, but in “what now?” or “where does this lead?” evidence of action-ability. This special edition provides a bridge between the deep reflection of evaluators and philosophers and the everyday questions that are implicit—but need to be explicit—in Cooperative Extension (Extension) work. Hopefully, reading and reflecting on these themes will help practitioners, administrators, and scholars understand, engage, and contribute to that work more effectively.

---

Direct correspondence to Scott R. Cummings at [s-cummings@tamu.edu](mailto:s-cummings@tamu.edu)

Much of the debate about what is credible has focused on the quality of *scientific* evidence that undergirds or affirms the efficacy of programs or policies. Utilizing and documenting best practice evidence employing appropriate scientific or systematic methods has been the primary focus of this special edition. Evaluation experts (Greene, 2015; Schwandt, 2015; Shadish, Cook, & Leviton, 1991) also acknowledge the significance of *practical* credibility (e.g., understanding where and how knowledge works for individuals or policy decision-makers) and *relational* credibility (e.g., whether the source of information is trusted). Several authors in this special edition highlight these broader meanings of credibility in the context of discussing program development, implementation, and evaluation.

The diversity of Extension disciplines, stakeholders, and delivery models preclude a single criterion for credibility. Indeed, Extension's place at the nexus of rapidly changing scientific discoveries, social, technical, and economic conditions and contexts precludes a single, static definition of credible evidence. Thus, Extension professionals must understand not only the current best evidence but the assumptions on which their teaching and listening rest. Likewise, their first and most important work in applying best practices is to interpret how and why those practices are effective, but not before they themselves reflect on the evidence behind a program and understand stakeholders' needs and concerns, then work with stakeholders to make an appropriate match of credible evidence with immediate and sustainable needs.

Hopefully, readers will find our work a relevant and trustworthy, or credible, source of insight and guidance for thinking and doing for the public good. This special edition is designed to start the conversation, not provide an encyclopedia or practice manual. We believe the future of Extension rests on continuing this conversation at all levels, in all disciplines, in every context, and, as appropriate, with every stakeholder and partner.

### **Background/History**

Understanding where we go often involves understanding where we have been and where we are today. There is a long history of the use of evaluation in Extension. In fact, Seaman Knapp and other Extension pioneers used demonstrations to provide concrete evidence of the efficacy of new hybrids or methods, encouraged dialogue to help producers understand not just what works, but how, where, and why it works (Seevers, Graham, Gamon, & Conklin, 1997). Different conditions required different evidence for action: larger farms might produce more efficiently, but insect pests plagued those using monocultural practices.

In 1959, the Division of Extension Research and Training, Federal Extension Service at the USDA published a document entitled "Evaluation in Extension" (Bryn et al., 1959) in which they discuss several aspects of evaluation still debated today. More recently, the General Accounting Office (GAO) released a report on the mission of Extension (GAO, 1981). In this report, the GAO stated that Extension must use resources as efficiently as possible and that improved performance and assessment of impact were needed to determine the effectiveness of

programs. One thing we know for sure is that change and innovation have been a constant in Extension. Although the USDA document discussed evaluation concepts still used today, change and innovation have occurred with regard to evaluation in Extension over the past three decades.

As late as the 1990s, reporting on participation or customer satisfaction was considered to be sufficient evaluation in Extension. With the passing of the Government Performance and Results Act of 1993 (GPRA, Office of Management and Budget, 1993) and the Agriculture Research, Extension, and Education Reform Act of 1998 (AREERA, 1998), the requirements for evidence began the shift from reporting outputs, such as participation to the reporting of outcomes such as knowledge, skills, and practices. The GPRA put an expectation on agencies receiving federal funds to collect evidence on items such as goal setting, measuring results, and reporting these results. AREERA amended the original Smith-Lever Act of 1914 and required additional information from land-grant institutions, including programmatic summaries, scientific review processes, stakeholder input processes, multi-state and integrated work, and planned programs. Current program reporting requirements of the National Institute of Food and Agriculture (NIFA) include evidence of impact for programs delivered at the state or institutional level. These impacts, based on the planned programs, are used to fulfill funding strategies and legislative requests. During this time, there has also been an increased emphasis on demonstrating the value of Extension to our stakeholders (Franz & Townson, 2008).

During the timeframe from the 1990s to the present, Extension professionals have invested time and energy to the areas of evaluation and accountability. Communities of Extension evaluators and program staff have met formally and informally to discuss issues, identify challenges, and implement strategies to meet those challenges at the state, regional, and national levels.

To illustrate this, let's look at the recent evolution of evaluation of Extension programs in Texas. In the 1980s and 1990s, Texas focused on the reporting on both direct and indirect program clientele contacts. This evidence provided stakeholders with information on the extent of the reach of Extension programs. As we moved into the early 2000s, evaluation efforts moved to a structured and comprehensive assessment of program clientele satisfaction. Efforts were conducted to assess satisfaction of Extension clientele, volunteers, and elected officials. These efforts led to a formalized performance measure legislated by the State of Texas.

As we moved further into the 2000s, a shift was made to increase the assessment of knowledge, attitudes, skills, and application. The focus on application was primarily assessed by the measurement of the intention of clientele to adopt a behavior or a best management practice. In some cases, these data were collected to meet the required laws or policy, such as GPRA or AREERA. As we entered the 2010s, impact evaluation became critical. The primary focus of these efforts was on economic impact. This was a major shift in thinking and the culture. We were being asked about return on investment and economic benefits of our efforts. Public value

also became an important addition to our evaluation processes. Public value added another way to tell the story of the value of Extension. Public value allowed us to show the value of our efforts beyond just our program participants, expanding the distribution of that information to the general public (Kalambokidis, 2004). In addition, a move was made to use new tools for telling the story such as infographics, video, and social media.

This brings us to the newest evolution of Extension evaluation: the collection and use of credible and actionable evidence. As described in other articles in this special issue of the JHSE, credible evidence is defined as “information that stakeholders perceive as *trustworthy* and *relevant*” (Donaldson, 2015, p. 5). In Extension, as well as other community education programs, we have collected evidence for decades on our efforts. We have generally focused on data being scientifically valid. However, have we ever determined or focused on the credibility of our evidence to our stakeholders? Is our evidence credible just because it comes from Extension? Is it credible because it is based on research? How do we move past just collecting evidence to collecting evidence that is deemed credible and actionable?

This special issue has addressed issues, challenges, and opportunities with regard to credible and actionable evidence and how these concepts can be used in Extension. The authors raised questions, identified concepts and practices, and hopefully sparked discussion and debate for this topic. As publicly accountable institutions, land-grant university programs like Extension must utilize these concepts and practices to maintain and increase its credibility in educational delivery and the evidence that supports its value to stakeholders.

### **Reflecting on Special Edition Insights**

Now that the concept of credibility and actionability has been raised, how does Extension address these concepts and the issues raised in this special edition? Let’s review the concepts and conclusions discussed in this special issue.

In the article, “Whose Extension Counts? A Plurality of Extensions and Their Implications for Credible Evidence Debates,” Tom Archibald focused on how we define Extension and the implications of these definitions. Archibald argued for an ontological plurality of Extensions and that this prevents a one-size-fits-all approach to credible evidence. He went on to summarize that we need to use the best-suited methods to obtain credible evidence based on the ontology of Extension being followed. This provides a strong argument for knowing who we are and what we do. Extension professionals must reflect not only on their methods but their underlying assumptions about what is valued by stakeholders for their communities or organizations. We cannot assume that what is good for one stakeholder will be sufficient for another.

Archibald also discussed the general debate over the concept of credible evidence and the use of randomized control trials (RCT) as the ‘gold standard’ versus other methods for gathering credible evidence. So how are RCTs used in Extension? One might also ask are RCTs used in

Extension or should they be used in Extension. As we think through these questions, we must be guided by what we are trying to measure and for whom. RCTs are one method we might use to obtain credible evidence. As noted in his article and the article on quantitative and qualitative methods by Jones, Gwynn, and Teeter, there are other methods available to be used. For example, how do qualitative methods play a role in credible evidence if RCTs are the gold standard? The methods and measures must fit the needs of the program and the stakeholders.

In their article, Chazdon and Grant focused on the relationship between situational complexity and credible evidence. Situational complexity by definition from Chazdon and Grant refers to the differences between simple, complicated (both technically and socially), and complex situations. These descriptions most definitely fit the Extension model. So how do we approach these issues? We must build relationships with our stakeholders. This includes knowing their needs, how they use evidence, and what story they are trying to tell. As with programming, evaluations require different types and amounts of evidence for different stakeholders.

Chazdon and Grant concluded that relationships with stakeholders are crucial to building credibility with stakeholders and that, in some cases, credibility is more about the program and the people than the evaluation. “One-size-fits-all” credible and actionable evidence may sound efficient and fair but rarely meets the real-time needs of program partners. Thus, evaluative thinking and dialogue are critical to both program management and accountability.

In the article on measurement, Marc Braverman discussed the relationship between measurement and credible evidence. Braverman concludes that there are a number of factors (recommendations) that must be taken into account with regard to credible evidence. These include quality, rigor, engagement, education, communication, and resources. There is a saying that “What gets measured, gets done,” but if programs are not measured accurately, evidence may be neither credible nor useful. Whether Extension educators are using common measures, adapting measures from other contexts, or developing their own measures, rigorous reflection and pilot-testing are critical to generating quality data and insights for program accountability, improvement, and partner education.

Related to this article is the paper on methods authored by Jones, Gwynn, and Teeter. They discussed the use of methods, both quantitative and qualitative, for credible evidence. As with measurement, methods must be used to collect evidence that fits the needs deemed as credible by stakeholders. Research methods are required in most graduate and some undergraduate programs that prepare Extension professionals. The press for credible and actionable evidence makes understanding and mastery of these approaches a necessity for all Extension professionals. Measurement and method issues in regard to collecting evidence, and more importantly, credible evidence, is not just creating and implementing a survey. Thought must go into what we are trying to measure, how we collect that evidence, and for whom the evidence is being collected. We must be deliberate and focused.

In the papers by Marczak et al. and Place et al., the concept of credible evidence was discussed across Extension program areas and amongst the various stakeholders within Extension. Both of these papers discussed the vast challenges facing Extension with regard to credible evidence and meeting stakeholder needs. Making evidence trustworthy and relevant across a wide range of programs and contexts begins with understanding why and how programs and evaluations are conducted and what, how, and why stakeholders will understand those programs and the evaluations that follow. At best, making program development, implementation, and evaluation more transparent for Extension professionals will clarify program and evaluation reasoning. Anticipating and knowing the needs of stakeholders and investing more in collaborative planning and analysis with decision makers will help Extension professionals become more credible and actionable into the future.

Closely connected to these articles was the paper on telling your story by Craig and Borger. Communication strategies to improve credibility were discussed, along with highlighting the use of infographics. For credibility, how the story is told is an important step and one that is often diminished in importance or ignored. We too often leave data sitting on the shelf. When we do tell our story, we often try to tell everything that happened regardless of the needs of the stakeholder. Credibility begins with solid evidence but requires equally sound—and diverse—ways of “telling the story” to be useful to stakeholders. New technologies such as blogs, websites, and infographics offer more user-friendly options to reach different audiences. Before Extension can get to the right “packaging” of its evidence, it often must invest significant effort, with trial-and-error, innovation-and-resistance, aligning diverse partners to plan, program, collect and analyze data together. Credibility and actionability, even given their diverse criteria and contexts, remain critical concepts to guide Extension evaluation and communication into the future.

Finally, the paper on evaluation capacity looked at how Extension organizations can build capacity for evaluation and credible evidence. The paper highlighted strategies and models for creating knowledge, skills, and a culture of evaluation. Each of the preceding papers highlights the theme that generating credible and actionable evidence is not a spectator sport. Nor is it “someone else’s job.” The credibility of “what goes out” (e.g., evidence-based programming) and “what comes in” (e.g., high-quality evaluation data) depends principally on each Extension professional’s evaluative thinking and doing. For individuals, programs, and systems, this requires continuous and creative investments in building Extension professionals’ capacities at all stages in the evaluative cycle. There are a multitude of strategies for improving capacity and credibility, but only one good time to start doing something: today.

## Dilemmas and Challenges Facing Extension

***“We're blind to our blindness. We have very little idea of how little we know. We're not designed to know how little we know.”***

**—Daniel Kahneman (2011)**

Psychologist Daniel Kahneman explored many dimensions of human judgment and decision-making, including human propensity to substitute intuition or bias for reasoned conclusions. He noted tendencies, even among experts, for making faulty observations and using inconsistent reasoning, and for explaining either the wealth or dearth of evidence as supportive of their theory. His work (Kahneman, 2011), in part, inspired the summary we offered in the introductory article of this special issue and gives us pause in closing to consider the challenges of credible evidence.

Credible evidence is not implicit in Extension work. What “makes sense” in Extension practice may be a product of habit or an improvement on the process. Business analysts Jeffrey Pfeffer and Robert Sutton (2006) examined the emerging concept of evidence-based practice and noted that studies in medicine, where the concept began, found that physicians followed evidence-based practice about 15% of the time. Was this because they retained old habits not quite up to current research standards or because the wisdom of experience guided them beyond particular procedures to the best interest of the patient? Reflecting on those patterns, how do we, as Extension professionals, know and use our evidence base? When we depart from a curriculum guide or practice standard, do we enhance outcomes or conveniently maintain our routine? Alternatively, do our clientele listen to us because we have research-based information or because we have a long history of service?

Not all areas of Extension work have the same breadth and depth of research to guide practice. In some cases, traditional practices may be assumed to be sufficient evidence of “what works.” Credibility needs a referent, an indicator (Rockwell & Bennett, 2004), and a process for evaluating programs against that criterion. Extension professionals can then develop a theory of change (Weiss, 1997), especially one that is sensitive to the complexity of learning objectives and settings (Douthwaite & Hoffecker, 2017). However, sometimes research evidence is not as reliable as was thought (Cohen, 1990), more nuanced and complex (Fry et al., 2016; McNamara, 2015; Mitloehner, 2016), and requires a change of theory.

Credible evidence is not always easy to attain. Scientific discoveries, program applications, and gathering of evaluative evidence take time, money, expertise, and, sometimes, just the right conditions. In some cases, precision is critical, as in measuring toxic particles-per-million in water quality or vaccination dosage. In other settings, accurate though not exact estimates of dollars saved due to energy efficiency or conservation practice may be acceptable. Behavior change may be a leader- or sponsor-valued priority for a Master Gardener program but may not



be a goal for all participants. In fact, among participants who value behavior change, goals may include growing a healthy, sustainable garden as well as saving money on food, eating healthy and losing weight, expanding social networks, or increasing community service. Evidence of program impact Extension programs such as 4-H clubs is generally greater when long-term involvement and outcomes are measured. However, where participation is inconsistent, program quality may be the best proxy for potential benefit (Arnold & Cater, 2015).

Credible evidence is increasingly difficult to attain in a complex, fast-moving environment. Yet across most disciplines, innovation for getting products and services to market and problem-solving responses to changing conditions is increasingly necessary for remaining competitive or sustainable. Smarter use of technology may be part of the picture (Milla, Lorenzo, & Brown, 2005), but changes in human systems are likely a larger part (Boteler, 2007; Warren, 2018). Extension is and has been a critical catalyst for technology transfer, but perhaps is most effective as a promoter of process skills (e.g., facilitation, teamwork, leadership, problem solving, communication skills) and community ownership (Colasanti, Wright, & Reau, 2009). What program participants learn about taking perspective, thinking critically and creatively, engaging in collaborative problem solving, and evaluating the outcomes may be more valuable and transferable than any specific solution or product produced in an Extension program.

Credible evidence is not universally acclaimed. Diversity in credibility criteria is not simply a matter of personal tastes and opinions. Evidence viewed from short- or long-term perspectives may seem more or less credible. Not all stakeholders, or evaluators for that matter, are equally capable of discerning credible evidence (Miller, 2015) and are likely to approach evidence from different social inquiry paradigms (Christie & Fleischer, 2015). Different disciplines take different approaches (Moon & Blackman, 2014) and, as shown in the Place et al. article in this special issue, criteria for evidence varies from the federal to the local level. Although these different perspectives may impede or delay consensus on evidence, they may actually enhance breadth and depth for understanding complex evidence.

So how do Extension professionals deal with these dilemmas? First, by seeking to understand issues of credibility and actionability, becoming evaluative thinkers. Second, by becoming “catalysts for critical reflection,” listening empathically but encouraging citizens and partners to question assumptions and strategies, sometimes to overturn them but often to devise more actionable solutions. Finally, to become a “community of practice,” sharing problems needing evidence and effective approaches that are appropriate for a wide range of stakeholders.

### **Final Thoughts**

What can you do to examine the credibility of evidence that informs your program or evidence that represents your program? What are you doing to help stakeholders understand and embrace higher standards for credible evidence? How do you need to interpret the same program evidence to different stakeholders? In how many ways, or with how many different groups or

projects can you apply evidence? How have you been challenged to learn new ways of gathering evidence?

Although much of this issue has focused on credibility, we must not forget about actionable evidence. Actionable evidence refers to that evidence which can be used to make decisions, including in the areas of policy and programming (Julnes & Rog, 2015, p. 221). How will you use the evidence collected? How do the stakeholders plan to use the evidence collected? If we do not consider how the results will be used, programming and evaluation just does not work. We encourage you to not over-complicate the process. Be innovative on how to use measures and methods. Don't waste your efforts, and most importantly, don't collect evidence that won't be used.

We hope you have enjoyed this special issue of the JHSE about credible and actionable evidence. Let's look back at where we began and ended the opening article.

- “I know I am making a difference,” a confident young county Extension agent declares. “Our nutrition education program served 4,500 people last year.” “OK,” the county director replies, “So, how many of those participants and their families are eating healthy meals or saving money on food or medical bills?”
- “I know I am making a difference,” explains an experienced field crop Extension agent, “Producers are implementing conservation practices, trying drought-resistant varieties, and recognizing early-on when they have disease problems.” “Great,” replies a state Extension specialist, “But did producers “check off” those items on a list, or describe what they actually do? Have you been in the field with them to observe these changes?”
- “I know I am making a difference, an Extension program leader notes. “Three counties with long-standing financial management programs saw an increase of ten percent in families becoming self-sufficient. In three counties where there was never an interest in those programs, at least five participating families became self-sufficient and recommended the program to their friends.”
- “I know I am making a difference,” an Extension volunteer youth leader insists, “Our programs teach life skills, so they will be productive citizens in the future.” An interested county commissioner replies, “What exactly are those skills, and how do you know it is your program that turns youth into productive citizens?”

Using the information and insight from the articles in this special issue, how would you now work to provide credible and actionable evidence to address these issues? What measures and methods would be needed? How would you determine if these are the right kinds of credible and actionable evidence to be used to tell Extension's story?

The next steps are ours. There is no one-size-fits-all answer to the topic of credible and actionable evidence. Different stakeholders will have different expectations for credibility. Different disciplines and different issues within a discipline will require different measures and methods. Organizations will have different levels of resources and support to address these issues. Finally, there are also different expectations at all levels within the Extension organization.

The bar for collecting and using credible and actionable evidence has now been raised. Let us all set a goal to meet and exceed this new challenge.

### References

- Arnold, M. E., & Cater, M. (2016). Program theory and quality matter: Changing the course of Extension program evaluation. *Journal of Extension*, 54(1), Article 1FEA1. Retrieved from <https://joe.org/joe/2016february/a1.php>
- Agricultural Research, Extension, and Education Reform Act [AREERA] of 1998, 112 Stat. 523. (1998). Retrieved from <https://nifa.usda.gov/sites/default/files/resource/Agricultural%20Research%2C%20Extension%2C%20and%20Education%20Reform%20Act%20of%20201998.pdf>
- Boteler, F. E. (2007). Building disaster-resilient families, communities, and businesses. *Journal of Extension*, 45(6), Article 6FEA1. Retrieved from <https://www.joe.org/joe/2007december/a1.php>
- Bryn, D., Fessenden, J. G., Frutchey, F. P., Gallup, G., Matthews, J. L., Porter, W. F., Raudabaugh, J. N., & Sabrosky, L. K. (1959). *Evaluation in Extension*. Washington, DC: U.S. Department of Agriculture.
- Christie, C. A., & Fleischer, D. (2015). Social inquiry paradigms as a frame for the debate on credible evidence. 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. 27–38). Thousand Oaks, CA: Sage.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12), 1304–1312. doi:10.1037/0003-066X.45.12.1304
- Colasanti, K., Wright, W., & Reau, B. (2009). Extension, the land-grant mission, and civic agriculture: Cultivating change. *Journal of Extension*, 47(4), Article 4FEA1. Retrieved from <https://www.joe.org/joe/2009august/a1.php>
- Donaldson, S. I. (2015). Examining the backbone of contemporary evaluation practice: Credible and actionable evidence. 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. 3–26). 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.

- Douthwaite, B., & Hoeffcker, E. (2017). Towards a complexity-aware theory of change for participatory research programs working within agricultural innovation systems. *Agricultural Systems*, 155, 88–102. doi:10.1016/j.agsy.2017.04.002
- Franz, N. K., & Townson, L. (2008). The nature of complex organizations: The case of Cooperative Extension. In M. T. Braverman, M. Engle, M. E. Arnold, & R. A. Rennekamp (Eds.), *Program evaluation in a complex organizational system: Lessons from Cooperative Extension*. *New Directions for Evaluation*, 120, 5–14. doi:10.1002/ev.272
- Fry, J., Neff, R., Martin, B., Ramsing, R., Fitch, C., Kim, B., Biehl, E., & Santo, R. (2016). A response to Dr. Frank Mitloehner's White Paper, "Livestock contributions to climate change: Facts and fiction". Retrieved from [https://www.jhsph.edu/research/centers-and-institutes/johns-hopkins-center-for-a-livable-future/\\_pdf/about\\_us/FSPP/letter-policymakers/20160512\\_Mitloehner\\_Response12.pdf](https://www.jhsph.edu/research/centers-and-institutes/johns-hopkins-center-for-a-livable-future/_pdf/about_us/FSPP/letter-policymakers/20160512_Mitloehner_Response12.pdf)
- General Accounting Office. (1981). Cooperative Extension Service's mission and federal role need Congressional clarification. Retrieved from <https://www.gao.gov/products/CED-81-119>
- Greene, J. C. (2015). How evidence earns credibility in evaluation. 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. 205–220). Thousand Oaks, CA: Sage.
- Julnes, G., & Rog, D. (2015). Actionable evidence in context: Contextual influences on adequacy and appropriateness of method choice in evaluation. 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. 221–258). Thousand Oaks, CA: Sage.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus, and Giroux.
- Kalambokidis, L. (2004). Identifying the public value in Extension programs. *Journal of Extension*, 42(2), Article 2FEA1. Retrieved from <https://www.joe.org/joe/2004april/a1.php>
- McNamara, D. J. (2015) The fifty-year rehabilitation of the egg. *Nutrients*, 7(10), 8716–8722. doi:10.3390/nu7105429
- Milla, K. A., Lorenzo, A., & Brown, C. (2005). GIS, GPS, and remote sensing technologies in Extension services: Where to start, what to know. *Journal of Extension*, 43(3), Article 3FEA6. Retrieved from <https://www.joe.org/joe/2005june/a6.php>
- Miller, R. L. (2015). How people judge the credibility of information: Lessons for evaluation from cognitive and information sciences. 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. 39–61). Thousand Oaks, CA: Sage.
- Mitloehner, F. (2016). *Livestock and climate change: Facts and fiction*. Retrieved from <https://caes.ucdavis.edu/news/articles/2016/04/livestock-and-climate-change-facts-and-fiction>

- Moon, K., & Blackman, D. (2014). A guide to understanding social science research for natural scientists. *Conservation Biology*, 28(5), 1167–1177. doi:10.1111/cobi.12326
- Office of Management and Budget. (1993). Government Performance and Results Act of 1993, 107 Stat. 285. Retrieved from <https://www.govinfo.gov/content/pkg/STATUTE-107/pdf/STATUTE-107-Pg285.pdf>
- Pfeffer, J., & Sutton, R. I. (2006). Evidence-based management. *Harvard Business Review*, 84(1), 62–74, 133.
- Rockwell, S. K., & Bennett, C. F. (2004). *Targeting outcomes of programs (TOP): A hierarchy for targeting outcomes and evaluating their achievement*. Retrieved from <http://digitalcommons.unl.edu/aglecfacpub/48>
- Schwandt, T. A. (2015). Credible evidence of effectiveness: Necessary but not sufficient. 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. 259–273). Thousand Oaks, CA: Sage.
- Seevers, B., Graham, J., Gamon, J., & Conklin, N. (1997). *Education through Cooperative Extension*. Albany, NY: Delmar Publishers.
- Shadish, W. R. Jr., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Thousand Oaks, CA: Sage.
- Warren, W. A. (2018). Developing a socio-ecological approach to Extension natural resources programming. *Journal of Extension*, 56(7), Article 7COM1. Retrieved from <https://www.joe.org/joe/2018december/comm1.php>
- Weiss, C. H. (1997). Theory-based evaluation: Past, present and future. *New Directions for Evaluation*, 76, 41–55.

*Scott R. Cummings*, Dr.P.H., is head of Organizational Development for the Texas A&M AgriLife Extension Service and Professor in the Department of Agricultural Leadership, Education, and Communications at Texas A&M University.

*Benjamin Silliman*, Professor and Extension Specialist, Department of Agricultural and Human Sciences, North Carolina State University, specializes in program evaluation and youth development. He received his Ph.D. in Human Ecology from Kansas State University and Evaluation Practice Certificate from The Evaluators Institute.

### Acknowledgment

The authors acknowledge Dr. Joseph Donaldson for his insights on the role of evaluation in Extension history.

Thanks to reviewers for this special edition, who responded quickly, then invested insightfully and sacrificially. Their investments made us all better, and we were blessed to work with them.

–Ben & Scott

Jason Altman, TerraLuna Collaborative  
Karen Ballard, University of Arkansas  
Sarah Baughman, Virginia Tech University  
Gary Briers, Texas A&M University  
Herb Byrd, University of Tennessee  
Jan Carroll, Colorado State University  
Melissa Cater, Louisiana State University  
Anil Chaudhary, Penn State University  
Jerry Chizek, Iowa State University  
Lisa Diaz, University of Illinois  
Shannon Dill, University of Maryland Extension,  
Mark Faries, Texas A&M AgriLife Extension Service  
Karen Franck, University of Tennessee  
Nancy Franz, Iowa State University (Emeritus)  
Doreen Hauser-Lindstrom, Washington State University  
Nigel Gannon, Cornell University  
Nancy Grudens-Schuck, Iowa State University  
Glenn Israel, University of Florida  
K.S.U. Jay Jayaratne, North Carolina State University  
Kimberly Keller, University of Missouri  
Vikram Koudrina, University of California Agriculture and Natural Resources  
Mike Lambur, Virginia Tech University  
Kendra Lewis, University of California Agriculture and Natural Resources  
Teresa McCoy, University of Maryland  
Bill Miller, University of Massachusetts  
Alda Norris, University of Alaska  
Janine Parker, North Carolina A&T State University  
Mitch Owen, Mitchen Leadership and Organizational Development  
Michelle Rodgers, University of Delaware  
Christian Schmeider, University of Wisconsin

## **Journal of Human Sciences and Extension**

### **Aims and Scope**

The Journal of Human Sciences and Extension is a peer-reviewed, open-access, online journal focused on disseminating knowledge and information to academicians, educators, and practitioners. Topics addressed include human development (e.g., early care and education, youth development); family studies; agricultural education; leadership development; extension; health and wellness; apparel, textiles, and merchandising; agricultural economics; nutrition and dietetics; family resource management; and program and staff development, planning, and evaluation. The journal seeks to bridge research and practice, thus all manuscripts must give attention to practical implications of the work. The journal is sponsored by the School of Human Sciences at Mississippi State University.

### **Types of Articles Published**

Several types of articles in the content areas listed above are considered appropriate for the journal: Original Research, Brief Reports, Practice and Pedagogy, Theory, Emerging Scholarship, To the Point, Book and Media Reviews. From time to time, the journal may publish a special issue on a specific topic. In these circumstances, a call for papers may be issued with details. Additionally, the Editorial Board may directly invite individuals to submit a manuscript.

### **Frequency of Publication**

The Journal of Human Sciences and Extension is published online three times a year in October, February, and June.

### **Open Access Policy**

The Journal of Human Sciences and Extension is a fully open-access journal, meaning that all works published in the journal are freely available to read, download, copy, print, and share/transmit.

### **ISSN**

ISSN 2325-5226

### **Publication Agreement**

JHSE requires all authors to sign a publication agreement prior to online publication of an accepted manuscript.

### **For more information**

Visit <http://jhseonline.org>