EVALUATION MATTERS

A UNIVERSITY OF NEVADA, RENO - EXTENSION PUBLICATION

ISSUE 11 | JAN-FEB 2026

CHRISTOPHER J. COPP
NAJAT ELGEBERI

IN THIS ISSUE

BEFORE AND AFTER THE BALL DROPS

MEASURING CHANGE OVER TIME

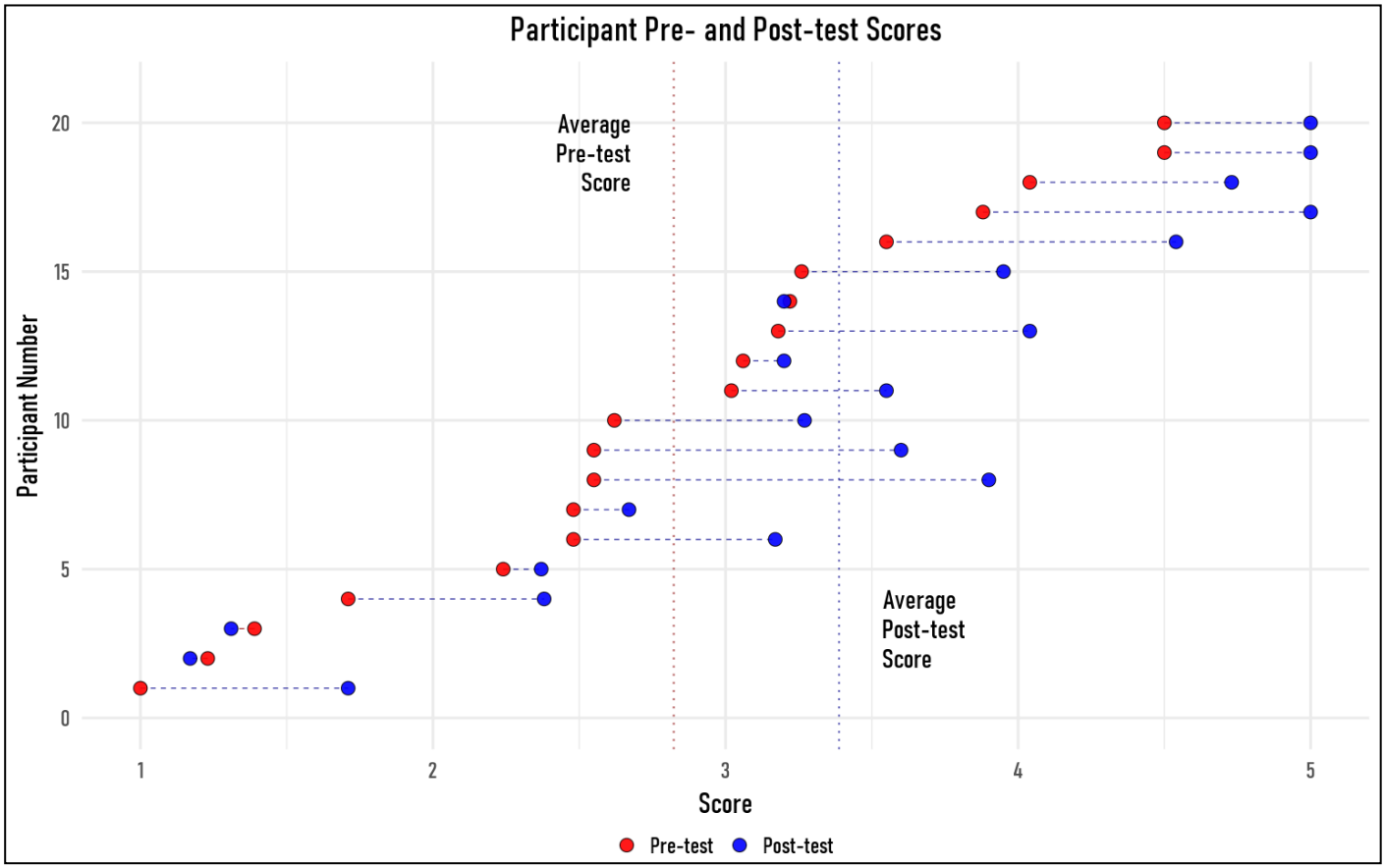
In last month's issue, we discussed how pre- and post-tests capture participant change by measuring the same outcomes before and after a program. Before-and-after surveys are one of the most practical tools extension professionals have, especially when working with trainings, workshops, or other educational programs. At the start of a new year, many teams set goals for strengthening programming and improving outcomes. Pre- and post-data can help show whether those goals are being met, and the paired t-test helps determine whether observed changes are meaningful.

When pre- and post-tests are given to the same participants before and after a program, the paired t-test is designed to evaluate change directly. Rather than comparing two separate groups of people, this test focuses on the difference within each individual. Each participant serves as their own point of comparison; the initial baseline measurement accounts for where people started. This makes the paired t-test a strong fit for Extension programs because it reflects how participants experience change over time.

This approach is useful because people often do not begin at the same point. Some participants may enter a training with high knowledge or confidence, while others begin with limited experience. If we only compare group averages, we can miss that starting point variation and overlook how uneven program growth may be. By examining how each person changed from pre to post, paired t-tests can provide a

CONTINUED ON PAGE 2

EXTENSION
College of Agriculture, Biotechnology & Natural Resources



clearer view of improvement that respects each participant's baseline.

A paired t-test essentially asks whether the average change across participants is larger than what we would expect to see through normal variation alone. Some people will show large gains, others might shift only slightly, and a few may not change at all. That mix is typical in real-world programming, and it does not necessarily mean the program was ineffective overall. The paired t-test helps determine whether the overall pattern points toward improvement, even when individual outcomes may vary.

Interpreting results from a paired t-test often comes down to the p-value, which reflects how likely it is that the observed change could have occurred randomly. When the p-value is below .05, evaluators treat the change as statistically meaningful. This provides a more confident basis for describing participant progress than relying on averages alone. It also strengthens evaluation findings when results are shared with partners, funders, or program leadership.

It is important to note that statistical significance does not automatically mean the change was large or practically meaningful. A program can show statistically meaningful improvement while still having room to grow, especially if the change is small or limited to certain participants. For this reason, it is helpful to interpret paired t-test results alongside the actual size of the pre- to post change and the context of the program. Taken together, these pieces provide a more balanced picture of what outcomes suggest.

Paired t-tests can support smarter decisions about where to focus effort in the year ahead. When results show meaningful improvement, programs can feel more confident continuing an approach, expanding it, or using it as a model in other settings. When results are weaker or mixed, the same analysis can highlight where adjustments may be needed to strengthen participant experience. Either way, this method supports continuous improvement by grounding reflection in evidence rather than impression.

By pairing strong pre- and post-test design with the right statistical tool, evaluators can move from noticing change to interpreting it with greater clarity. Paired t-tests help teams understand whether progress occurred across participants in a consistent way, and they support stronger communication of program impact over time. This makes evaluation findings easier to interpret, easier to defend, and more useful for guiding programs into the new year.

NEW RESOLUTIONS

USING LAST YEAR'S RESULTS TO GUIDE THE NEXT

For individuals, the start of a new year often comes with resolutions and plans for doing things a little better than before. Extension programs and teams are no different! January is when many people start thinking about what to improve, what to adjust, and where to focus energy in the months ahead. In evaluation, those resolutions are most useful when they are grounded in evidence from the year that just ended.

Annual evaluation results are more than a summary of past activities. They can serve as a practical guide for setting priorities for the year ahead. Participation numbers, outcome trends, and satisfaction patterns all contain information that can shape realistic, focused plans. Instead of starting with guesses about what to change, teams can begin with data that points toward specific areas for improvement.

One way to approach this process is to look for patterns, much like reflecting on habits when setting personal resolutions. For example, you might notice that participants consistently report lower confidence on one specific topic, or that one learning objective shows smaller gains than the others. You might also see the same suggestion appear across multiple feedback forms. Patterns like these help distinguish between one-time comments and consistent signals that deserve attention in the coming year.

One helpful step is deciding which findings merit your attention. Not every data point calls for action, and trying to fix everything usually leads to frustration. Instead, look for one or two areas to implement realistic improvements. This might be a part of the program where participants consistently struggle, a section that runs longer than planned, or a resource that people say they rarely use. Focusing on a few high-impact, manageable improvements makes evaluation-informed resolutions more likely to stick throughout the year.

Another useful tip is to look at when things happened, not just what happened. Program timelines can reveal patterns that are easy to miss during delivery. You might notice that questions spike right after a certain activity, or that satisfaction is higher in sessions held earlier in the day. These timing patterns can inform practical resolutions, such as adjusting program length, reordering content, or building in additional support at key moments.

When translating findings into next steps, it is often best to keep resolutions focused and realistic. Rather than overhauling an entire program, teams might identify one or two specific changes supported by data. This could mean revising a key activity, adjusting outreach timing, adding more opportunities for practice, or providing clearer instructions. Small, targeted adjustments are more manageable and more likely to be carried through consistently.

Finally, using last year's results in this way helps make evaluation feel connected to real decisions. Data collection takes time and effort, and its value increases when findings shape what happens next. Treating evaluation insights as the basis for practical, evidence-informed resolutions helps ensure that the new year begins with direction rather than guesswork. Looking back is not about dwelling on the past, it is about using what was learned to build a stronger year ahead.to look for patterns, much like reflecting on habits when setting personal resolutions. For example, you might notice that participants consistently report lower confidence on one specific topic, or that one learning objective shows smaller gains than the others. You might also see the same suggestion appear across multiple feedback forms. Patterns like these help distinguish between one-time comments and consistent signals that deserve attention in the coming year.

"... look at *when* things happened, not just *what* happened."

HITTING THE SPOT

WHY RELIABILITY MATTERS

Imagine throwing darts at a target. If the darts land in the same spot each time, you have a clear pattern. If they land all over the board, it is much harder to tell what the goal of the game is. Measurement tools work in a similar way. When results are consistent, patterns become easier to see. When results are scattered and unpredictable, it becomes difficult to know whether differences reflect real change or just measurement noise. This idea of consistency is at the heart of a term used often by evaluators: reliability.

Reliability refers to how consistently a tool measures something. If the same person responded to the same question again under similar conditions, would their answer be similar? If multiple participants interpret a question, are they perceiving it in roughly the same way? Reliable measures reduce random variation and make it easier to detect meaningful patterns in outcomes.

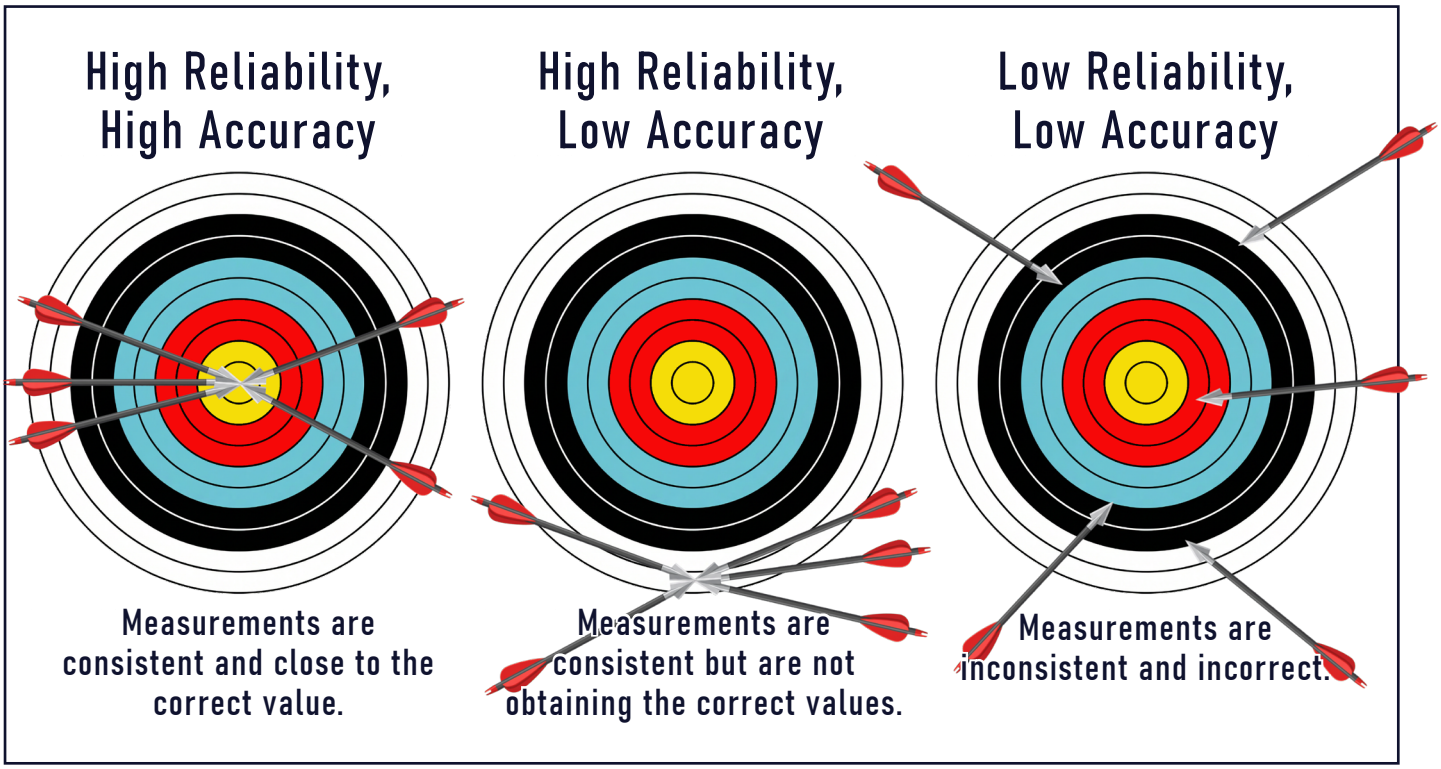
It is important to remember that reliability does not mean accuracy. A tool can be very consistent and still be consistently wrong. Think of a ruler with incorrect markings. Every time you measure a piece of paper, it gives you the same length, but that length is off because the ruler itself is flawed. The measurements are reliable because they are consistent, but they are not accurate because they do not reflect the true value. In evaluation, unclear wording, inconsistent response options, or confusing instructions can create that same kind of instability in the data.

"A tool can be very consistent and still be consistently wrong."

Reliable measurement is especially important when tracking change over time. When tools produce inconsistent results, it becomes difficult to tell whether score differences reflect real growth or just instability in the measurement process. Consistent tools make it easier to see patterns, compare results across groups or time points, and communicate findings with confidence.

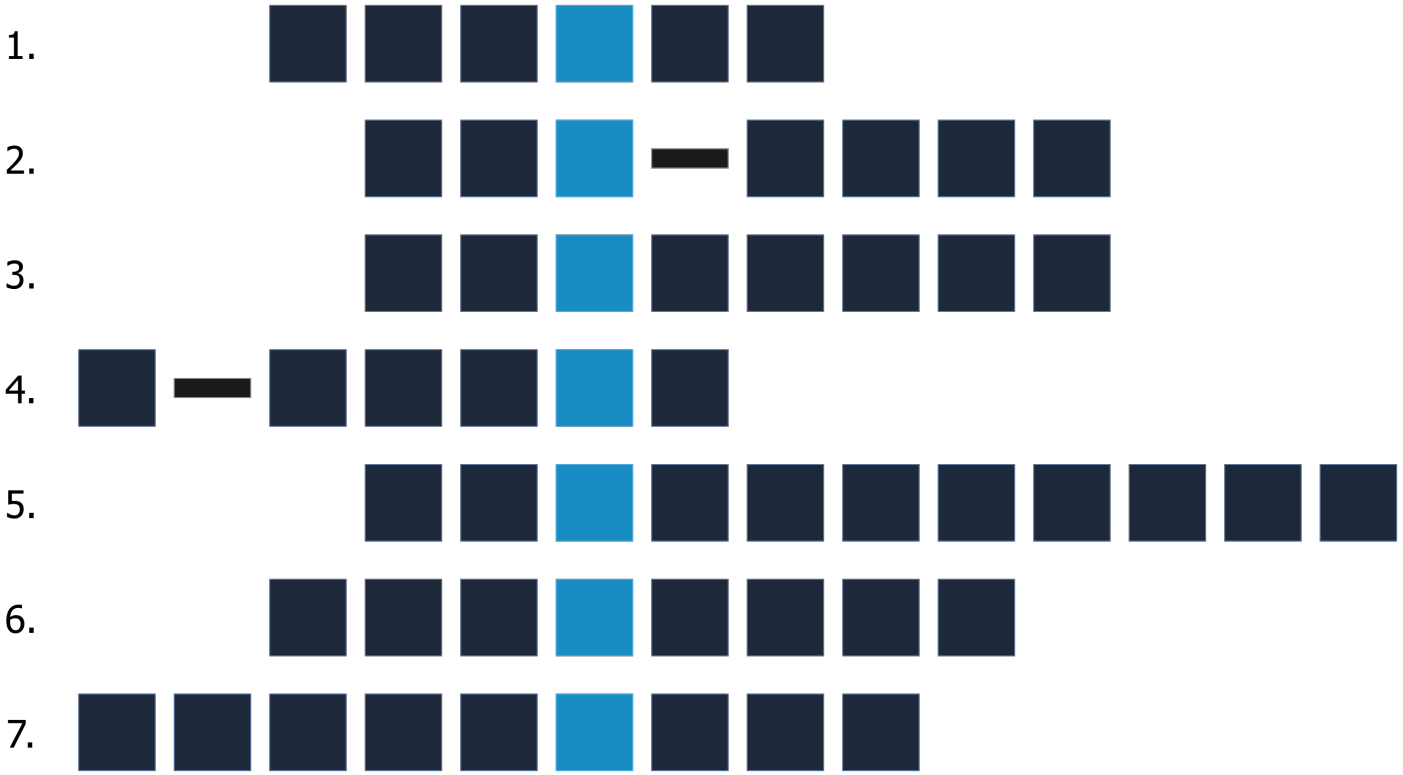
There are several practical ways to strengthen reliability. Clear and simple wording helps participants interpret questions in a similar way. Keeping response scales consistent across items and over time supports more stable results. Pilot testing a survey or assessment with a small group can reveal confusing questions before full implementation. Looking at how sets of items perform together can also help identify questions that do not fit well and may need revision.

Reliability is only one piece of the measurement puzzle, but it is a critical foundation. It tells us whether a tool produces stable results. In our next issue, we will look at validity, which asks a different question: are we measuring the right thing in the first place? Together, reliability and validity help ensure that evaluation data are both consistent and meaningful. genuine differences in experience that are important to understand.



NEW YEAR NEWSLETTER CHALLENGE!!!

Place the correct answers across in the puzzle below. The correct answers spell out a word in the vertical blue squares. Email this answer to ccopp@unr.edu and be entered into a raffle to win a prize!



1. What type of t-test compares the same participants before and after a program?
2. When tests are given to the same participants before and after a program, the 'before' test is called a _____.
3. An initial _____ measurement accounts for where people started before the intervention.
4. What statistical value, often below .05, helps determine if a change is meaningful or whether it could have occurred randomly?
5. This term describes how consistently a measurement tool produces stable results.
6. Reliable measures reduce random variation and make it easier to detect meaningful _____ in outcomes.
7. When improving Extension programming, choose one or two _____ improvements.

NEXT ISSUE:

SPROUTING SUCCESS

Testing whether your program meets the benchmark using one-sample t-tests.

SPRING INTO ACTION

Build healthy data management habits with organization systems that last.

PLANTING ON SOLID GROUND

Check your statistical assumptions to ensure reliable results.