



TRANSCRIPT

Basic Elements of High-Quality Research: What to Look for in a Research Study

[begin - 0:00:00]

Melissa Josue

Welcome, everyone. Again, welcome to Basic Elements of High-Quality Research: What to Look for in a Research Study.

John Rice

Hello. Good afternoon, everyone. We're really happy you could join us today for this REL West webinar, The Basics of Reviewing a Research Study. I'm John Rice, I'm the Director of the Regional Educational Laboratory West, REL West at WestEd. And I'm joined today by my colleague at REL West, Dr. Lenay Dunn, who is REL West's Deputy Director. And I'm also joined by two individuals from the Arizona Department of Education, who has been our partner in developing and using the materials that this webinar is based on. And they are Dr. Amy Boza, Director of English Language Arts at the Arizona Department of Education, and Sarah Bondy, English Language Arts Specialist at the Arizona Department of Education.

And they are going to provide specific examples of how they've used the concepts in the infographic that we're going to present and the tips for using this infographic as a tool with districts to help inform their selection of evidence-based literacy curriculum and interventions. And finally, Tran Keys will join us as well. She is a senior researcher at REL West, and she will help facilitate our breakout sessions.

Before I begin, I just want to briefly explain REL West. REL West is one of ten regional educational laboratories across the country. And we, like all the rest of the labs, are funded by the Institute of Education Sciences at the US Department of Education. And you can see REL West covers the four western states, and you can see the map there, where the other RELs have their coverage. Now, we help bridge education research and practice in three ways that you see on this slide here. One is by conducting applied research studies with our partners at the state and local levels.

A second way we bridge research and practice is by facilitating the flow of actual credible, up-to-date research evidence from across the field. And finally, we provide technical support around data collection, evidence use, and research to help build the capacities of educators and education policymakers in these areas. Now, the goal of today's webinar and the materials we're going to show you are to help you better understand how to review research articles or evidence summaries in the context of evidence-based strategies as required by the Every Student Succeeds Act, or ESSA.

As you know, ESSA was signed into law in 2015, and its provisions for selecting evidence-based programs are still relevant, particularly with American Rescue Plan and ESSER funding that the US Department of Education has recently provided in response to the COVID-19 pandemic. Now, we know that there's a variety of individuals with different types of research backgrounds on this webinar, so the familiarity, and expertise, and experience with research amongst all of you is going to vary, but we hope that today, after this webinar, no matter your previous experience with conducting or using research, that we've helped you become more critical consumers of education research and have done something to help inform policies and program decisions at whatever level of the system you work in.

Here's what we're going to do today. First, Lenay Dunn and I are going to go over a framework for reviewing research that was recently released by REL West and IES, and we developed this infographic for this very purpose, to help stakeholders like yourselves. And you'll see the link to that infographic in a few moments, and all the slides will be available to you after this webinar. After she and I go over the tool, we're going to have our partners at ADE talk about their experience using this tool. And then, we're going to break into small rooms and use the infographic with a fictitious research study, so you can practice the application of the content and the concepts that we're going to present.

Now, all during this webinar, there's a chat box, so feel free to use it with asking questions or making comments. We will be monitoring it. And one last note before I get into the content is, throughout the webinar today, I and the other presenters are going to use the word program as shorthand for programs, etc. Basically, using the word program for anything that a policymaker, administrator, or educator might do in an attempt to change teacher or student outcomes. As they say, "Let's get on with the show."

When we talk about the basics of reviewing a research study, REL West has lumped these into four broad considerations to focus on, and we're going to cover all of those today in some detail. And again, the infographic that you see the link to at the bottom of the page talks about all these concepts. The first concept is study source. Where is the research coming from? The second concept we'll talk about is study design. How was the study designed and carried out? And third, study findings. What were the measures of the intervention's effectiveness? And what was the strength of the intervention? Finally, and very importantly, study relevance. How the research applies to your specific educational context. These are the four really major considerations we'd like you to think about when you or others you work with are looking at research, trying to make some kind of decision about education practice or policy.

The next slide is a little bit about the first consideration, the source of the research, and this really focuses on the credibility of the study itself. These days, as you know, with the internet and easy access to all kinds of information research study, it's easy to find claims of what people might call program effectiveness. But really, to assess the credibility of the research, you have to look first at where it came from. What was the source of the research? Was it reputable and credible? Was it an independent journal? Was it from a developer? Was it something posted on Pinterest? In a few minutes, I'll show you some examples of where there are resources that in education research are usually considered credible resources. One of the key markers of credibility is when research is peer reviewed. Peer review, of course, is not a 100% guarantee that the research is of high quality, but it does increase the odds that it is.

And so, the idea is, in science, of course, you have experts in an area, and they look at each other's research, and they critique it so that once a study is published, you can have greater

confidence that the researchers followed the rules of research and accurately present the findings and their implications. And then, another thing to consider is the independence of the researchers from the thing or product that they're actually studying. This isn't to say that a program developer or someone who has a financial interest in the program is going to lie or cheat on research, but we do know that when you have a stake in something, you can tend to be biased, whether it's conscious or unconscious, about the way you present findings. Often, you could put a higher confidence in research on a program that's been done by an independent third party as opposed to the developers themselves. And if you are looking at research that's been done by a program developer itself, you'll really want to ask some questions about design and findings that my colleague, Lenay Dunn, is going to go over in a few minutes.

These next two slides are some examples of where one can find credible research. Of course, there's the What Works Clearinghouse website, and that has been revamped lately in the past year or two. If you haven't looked at it, I would take another look. ERIC, of course, which is also the Department of Ed, is a great resource for finding original studies. Evidence for ESSA, the third on that list, is a really nice resource from Johns Hopkins University and the late Bob Slavin, where they really go into how well particularly math and literacy programs align with the ESSA tiers. And on the next slide, you can see some examples where you can find some resources on social emotional interventions, particularly coming out of the COVID pandemic and getting students back to school, there's been a real focus on the social emotional wellbeing of our students and our teachers. On the next slide, you can see a few resources that really look at those types of programs specifically.

Lenay Dunn

These were the rest of the resources that John was mentioning. The Best Practices Clearinghouse, there are some strategies from the National Centers on Intervention, and there's a meta-analysis from RAND on social emotional learning interventions that might be particularly helpful to you. Great. Let's move into the next section that really talks about study design. When we think about study design, there are a few key elements that we're going to break down of what you're looking for to have more confidence that it's the intervention itself that's leading to the outcomes that are observed in the study. One is, the research design, what that design is. We'll talk a little bit about that. A key thing that you're looking for is the use of a comparison group.

We're going to talk about the importance of that. We're also going to talk about baseline equivalency, what that means, and where you're looking for it. We're going to talk about sample because you know a large sample gives you more confidence in the generalizability of the study findings. And also, we're going to talk about attrition, or when students or other participants drop out of studies, so we'll be talking through some elements of that. And we're also going to talk about outcome measures. We can have more confidence in a study if the outcome measure is one that is well-established, has been established to be reliable and valid, and is relevant to the intervention being measured. Let's jump in to break some of this down on study design.

There are four basic types of research design, and we highlight some of these in the infographic that really talk through what each of these and what you're looking for. But I want to give you some basics of what you're looking for. An experimental study, these kind of go from the strongest to the weakest, left to right. an experimental study or randomized control trial randomly assigns people, schools, districts to either the intervention or a control group. And

that randomization is really what makes it the strongest research design when there's that random assignment to the different groups the next most rigorous is a quasi-experimental, where it does also have the control group and the intervention group, but people or units aren't assigned to those randomly.

They are assigned to those, and they're assessed on characteristics that are meaningful to make sure that those two groups are not meaningfully different before they receive the intervention. That could be things like the demographics of the group, pretest scores, other things like that. That's quasi-experimental. A correlational study just really looks at the association between an intervention, or program, or strategy and its outcomes. And we'll talk a little bit about what ESSA says in terms of the evidence provisions of what you're looking for in that. But we've all heard correlation, not causation, so a causal study is going to be a randomized control trial, but this is really looking more at the correlation of an association and outcome. And then, finally, the descriptive studies. These are really the least rigorous, and this is often looking at data at one point in time or over a couple points in time. But what really marks this is, it's usually just one group. It's just the group who received the program or intervention, and there's not a control group.

It's really hard to isolate if it's the intervention itself that is leading to those outcomes. These are some of the basic research designs. When we think about the ESSA evidence levels and how they associate to these research designs, again, you'll see those four different kinds of research designs, experimental, quasi-experimental, correlational, and the positive evaluation is more like a descriptive study, and those are associated with these different ESSA evidence tiers. The strongest evidence tier, tier one, is really marked by that randomized control trial. A moderate or tier two study would be a quasi-experimental study that's well-designed and well-implemented, which we're going to talk about what that really means. And then, tier three, promising, would be that correlational study, associating outcomes with an intervention or program.

And then, finally, the demonstrates a rationale, or tier four, is really like a study that has a theory of action and ongoing evaluation, showing that it's got some real potential in improving the outcomes that it's hoping to improve through the intervention. The stronger the design, the more confidence we can have in the outcomes of the study. That's really something we want you to take away from this because it's really helpful. Markers, you're looking at all different kinds of research studies and trying to find an intervention or strategy that fits your needs and context.

Another element of research design that we're going to talk about is that baseline equivalency. And you may have heard this term before, or it may be new to you, but baseline equivalency is really trying to establish that in the intervention group and the control group, those two groups were not meaningfully different before the study began. Now, this is really only necessary for a quasi-experimental study because in a randomized control study, the randomization of assignment to groups really helps ensure that those characteristics aren't meaningfully different ahead of time.

You'll see some randomized control groups also do have some measures ahead of time to really establish that those two groups aren't meaningfully different, but randomization itself really helps to ensure that those groups are not going to be too different from each other. But when you can compare those groups and establish that they're not meaningfully different, it just gives you more confidence that it's the treatment itself or the intervention itself that is related

to the outcomes observed in the study. So that's something to look for specifically in quasi-experimental designs.

In the sample, what we're looking for is that it's a sufficient size. Through the non-regulatory guidance from the US Department of Education and the ESSA evidence provisions, it defines a large sample as at least 350 participants from more than one site, like a school or a district. That's really defined for those top two tiers of a randomized control trial or quasi-experimental design, really making sure that the size is large enough to really draw some of those conclusions. A sufficient size is kind of subjective, but if you want to use the ESSA evidence provisions to give you some guidelines, that can help you. Similarly with attrition. Attrition means how many people drop out of the study.

And when people drop out of the study, sometimes you don't know if the outcomes were different for the two different groups because of the dropout issue. But one thing that you're looking for is both overall attrition and then different group attrition. Generally speaking, with 20% or less overall attrition, you can be pretty confident in that. Now, the What Works Clearinghouse did develop some standards to define what high and low study attrition is, but this is just a general guideline if you're looking at a study. Looking for less than 20% of students who drop out of the study. I'm going to pause because I've seen a few questions in the chat. I want to take a minute and see if anyone wants to lift some up for me.

John Rice

Yeah, Lenay, this is John. There's an interesting question here. It's kind of got an complicated answer. But it's an important question. Someone asked, when you're talking about sample size, are you talking about the number of students, or are you always meaning the number of students?

Lenay Dunn

Right. The size at 350 is defined as the people in the study, at least in the provisions. And then, you're also looking for units. If the intervention was a school-level intervention, then that's going to be a little bit different. It depends on what outcome you're measuring. If you're measuring a student outcome, that would be the number of students. If you're measuring a school level outcome, then that would be the schools. Typically, it's going to be the number of students or teachers because that's usually the target of the intervention. Do you want to add to that?

John Rice

Well, I'll just add that it can be confusing because often, in Lenay's example, teachers are the individuals whose behavior you're trying to change. For example, you've got a teacher-level professional development program, but you're measuring that by testing students. In the case where it's a student-level program, the real important subject number is the number of teachers, no matter how many students under them you've tested because you roll up the data. It's a little bit complicated, but that's all to say the main thing to look at is what the sample size of the level you were trying to impact is, the student level, the teacher level, or the entire school level.

Lenay Dunn

I saw another question about the rural school districts. It's a great question, and we're going to get to a little bit more about that and give you some tools to help think about the application of research that's been conducted, let's say, in a large urban area with a large that would apply to your school district. But I think part of the reason why there's less research that's been conducted in rural settings is because of some of this issue where there are smaller numbers of students or teachers to be able to really get to those large sample sizes. You're going to see in some examples we'll share with you later in this webinar that often it's combining from suburban school districts, urban school districts, rural school districts when a study is being conducted. If you're in a rural school district, you need to think about how well the results apply to your setting. But the more you can find studies where rural settings are included, if that's your setting, the better.

Thank you so much for the questions. It's really helpful to know where some of your needs are, and feel free to keep putting them in the chat. We will address them as we can. When we look at outcomes, there are two big things you want to look for. One is how relevant that outcome is to the intervention. If the intervention is looking to improve 3rd grade math scores, and maybe it's a certain dimension of math achievement, then you want to make sure that the measure of that is related to that, that intervention is targeting these particular elements. How well does that outcome actually align to those elements? That's really important because you could have an outcome measure that's not related to the intervention where you see improvement, but that doesn't mean that the intervention itself is really going to be what's creating some of that or contributing to some of that growth.

Similarly, you want to have some established measures. What that means is, sometimes programs might develop measures, so program-embedded measures, where there's an assessment that's built into the intervention. That alone isn't necessarily bad. You just want to make sure that these measures have been established and verified from outside sources to make sure that those have some applicability within your setting. Some established measures or examples could be a state language arts exam, a popular interim assessment, and I know there are several out there that we probably know about, but something that's already been established and verified that is a reliable measure that you can use related to that outcome.

And something that you're really looking for, and the reason this is important, is, sometimes there can be over-alignment as well, if the intervention and the outcome that the developers of the intervention developed are too closely aligned. That really introduces the potential for bias, that you're going to get more positive findings than you would on an outside, more established measure. Just something to have a little bit of caution around. We're going to get to relevance in just a minute about that as well. Thank you for that question. Talking about study findings, there are two components that you're typically looking for, and one is statistical significance, and the other is effect size. Statistical significance is typically defined as a P value, P less than 0.05 is what you're often going to see, and that's pretty much always reported in studies. Where there are quantitative measures, there's going to be a statistical significance reported.

I think something to really look for is digging into those outcomes, so sometimes, you'll only see the outcomes reported that were statistically significant, but it's important to dig a little bit deeper and see all the outcomes that were measured because did some of them have negative effects? Or were other outcomes that were more important to you or more aligned with or

important to the intervention not have statistically significant results? Really digging beyond the overall and digging into what's happening.

Same for particular student groups. Overall results versus what's statistically significant for certain student groups. And then, for effect size, this is sometimes reported. It's not always reported. But the larger the effect size, the more robust or stronger effect that the intervention itself has. And we often like to suggest that when you're looking at a study, it's helpful if the study uses a real-world metric like number of days, how that effect size is different in terms of number of days of instruction or what that looks like on an increase in a particular score. We do want to help really make sure folks understand that while statistical significance and effect size work really well and pair well together, they are different measures, and if you're really trying to ensure that you're meeting the ESSA evidence requirements, effect size is not considered in the ESSA evidence requirements, but statistical significance is, so that's something to look for when you're looking at research studies.

I saw the question about what statistical significance measures, and that's a really good question. I will point you to the definition of it in the infographic, and it's a little bit complicated, so it might take a few times of working with it. And I think Sarah and Amy, when they through how they've applied some of these concepts, can share a little bit about their journey around that. But the P value is about probability, and it's the probability that in the study, the difference that you've observed between the control group and the intervention group, that what was observed would be at least as large as what you're seeing, even if there really was no difference in the groups, the true difference was zero. It really gives you some sense of how much - we don't say chance, but how much this is related to what is just in the sample versus what would be true if everyone in the population was as part of this study.

John Rice

Right. Just riffing on that, on the second question we got, why do we want a P value below 0.05, what 0.05 is saying is that 5 out of 100 times, you would've seen that difference that looks real, even if it weren't because you're picking a sample, and it's out of a larger population. It's really trying to give you a sense of how much confidence you can have that that difference you're seeing is an actual difference. And 0.05 is the convention in social sciences. It's arbitrary in the sense that it's just a cutoff that social sciences and educators have decided on over the years. You'll notice in medicine, where they want much more confidence on more life-and-death types of decisions, you'll want P values of 0.01 or even 0.001, which is 1 in 1,000. We don't need to get into that, but just to say that the 0.05 is just the standard convention in social science.

Lenay Dunn

Great. I know someone was very interested in talking about study relevance, so let's talk a little bit about what you're looking for in study relevance. This is going to be really unique to your setting, but what you're looking for is how closely the population in the study is related to the population that you're serving. Back to the rural example earlier, even better if you have a study that included students in a rural setting, that is a consideration. Also, this could be student demographics, teacher demographics, what about your population, how well it relates to the study population, and how well the outcomes match your outcomes of interest. We're going to see some examples where DIBELS is an outcome that's being measured.

Does DIBELS measure what you're hoping to improve? Is DIBELS something that you use? These are questions that you ask. And again, it's not to say that the study isn't important to you or couldn't inform your decisions, but it just gives you a lens to help interpret it. And the reason you really want to think about the relevance is that if you want to get the results that were in the study, the closer that you can have in terms of the population, the setting, the contextual factors, how something was implemented, the more likely you are to get to those outcomes when you actually implement it in your setting. Again, it doesn't mean that the intervention might not be right for you, but it's something to help you think about what it's actually going to take to get the outcomes that were in the study.

Let's look at an example together to apply these big concepts. I'm going to share this scenario with you, and I'm going to take a minute to go through it because we're going to piece-by-piece break this down. This is a preview because when you move into your small groups, you will be doing some of this together with a different scenario and example. This is a completely fictitious study. All of these elements are made up for the purpose of discussion. I did want to note that. In this example, this is a reading intervention that's really for struggling readers in K-2, and the intervention is a 90-day intervention, so keep that 90 days in mind. The study authors are from a research center that's affiliated with a university, and it was published in a peer-reviewed journal. It was designed as a randomized control trial that lasted 180 days. One of the first things you'll do, line up that intervention period with the research study design, just something to look at.

There were 427 student participants, so a pretty large sample size in nine schools in two districts, a rural and a suburban. You can see the demographic breakdown of the students here, so if you were looking at this intervention, you would think about how well it matches your student population. And 66 students dropped out of the study, but the study didn't specify whether that was in the treatment or control group, and they did report baseline scores, so a pretest score for students. In the results, students in kindergarten and 1st grade who received the intervention had statistically significant higher scores on that program assessment that was embedded in the program and in DIBELS compared to the students who were in the control group. 2nd grade students had statistically significant scores in the intervention group, but only on that program-embedded assessment, and it did not affect report effect size. When you look at student group results, all of the subgroup findings were really near what happened in the overall group, but it is important to note that English learners in the treatment group did not make statistically significant achievement gains, and this was true across all of the grades that we studied.

Let's break down what that looks like when we have these different considerations. The source, when you think about its credibility, this was accessed from ERIC.ed.gov, which we know is a credible source. It was published in a peer-reviewed journal, so we know it went through that process. That gives us a little more confidence in the source of the study. And it was a university-affiliated research center that did not appear to potentially be associated with the intervention itself. The study was a 180-day RCT. Now, that's twice as long as the program itself. That's important to note. What does that mean? I'd want to know a little bit more about whether it was repeated, if the students got the intervention twice, and then the outcomes were measured, or if they didn't implement it within a 90-day period, and they implemented it in 180 days. I would want to ask some questions about that to better understand this a little bit and get into that design and implementation of the study to know a little bit more. It wasn't necessary, as we talked about, for a baseline equivalency because this isn't a quasi-

experimental design. Students were randomly assigned to receive the intervention. But the study did report those baseline pretest scores, so that just gives us even more information, even more confidence in the intervention and the results.

It had 427 students, so over that large sample threshold of 350, but 60 students dropped out, and it didn't specify if they dropped out from the control group or the treatment group. That gives us some sense, but we would probably want to dig a little bit deeper to understand if the attrition was more from one group or the other. The outcome—and I did see someone put in the chat what DIBELS is, thank you—itself was relevant to the interventions. The intervention was for struggling K-2 readers and used that established measure that was developed externally and is verified as an assessment or measure, DIBELS. And it also had that program-embedded assessment from the intervention designers. So it had both outcomes. I think that's a really interesting way to think about the outcomes because one is external, and verified, and one that's used pretty widely, and then the other one is just unique to the program. That's a nice way for a study to kind of get around some of those issues and look at multiple outcomes.

If we can look at the next piece, we have the findings. As we noted kindergarten and 1st grade students had that statistically significant positive outcome on both the program-embedded assessment and that external measure, but 2nd grade only made that improvement on the program-embedded assessment, so there might be something there to further look into or think about in your setting. Are you targeting more kindergarten and 1st grade? Would that be OK? Or was there something about the way that the study was carried out that 2nd grade didn't receive the intervention as fully? But it is important that English learners did not make statistically significant gains, which would be something I would definitely want to look into, especially if my school district had a large English learner population. Effect size was not reported, but that doesn't mean that it's not a rigorous, or important, or helpful outcome. Because we saw the statistical significance.

It would be even better if we knew some measure of the program's impact or strength. In the relevance, you'd have to think about your own setting. You'd look at if your setting was rural or suburban, if you see yourself reflected there, and you'd compare your student population. And one of the reasons we think about the student population is, would there be something about the setting or the population that you'd have to modify the intervention in order to implement it well or implement it to meet your student needs? And if you modify that, will that make it so different than what the intervention was in the core of the intervention that you wouldn't get the same or similar outcomes from what was in the study? It's just a consideration. It doesn't mean it's not necessarily the right fit for you, but it's something to think about. And then, finally, the match to your needs.

And I would say this is probably one of the most important pieces, to think about the match to your needs. Is DIBELS and the program assessment, what it's measuring, what you're hoping to improve? Are these pieces that, if you have your struggling readers, and this is an intervention you're considering, really fit what you're looking for? And if you have a large number of English learners, I really would want to think about this in my setting. I'd think, "Well, this might not be the right intervention for us to really target for all students. Maybe we need to think through that a little bit more." I see some questions have popped up, John. Are there any in there that you wanted to address at this point?

John Rice

Folks are asking questions, but they're also making some very good observations. For example, someone earlier mentioned that statistical significance is more likely if you have a really, really large sample, which is true. And so, that's why it's also important, if you can, to look at effect size because with a really large sample, you could find statistically significant differences, but the magnitude or effect of the program could be really small. That's why it's always nice to have effect size as that companion to look at alongside statistical significance, so that was a great point. One other point someone made was, with the study you were looking at, this individual said, "EL's were only 4% of the study population, which is 17 students. So I think it would be hard to draw any conclusions on that small of a sample." And that's a great point. If your needs were to find an intervention specifically for an English learner population, and you were considering that program, you might want to see if there was evidence beyond that study, given, like you said, there were only 17 students involved in the research. That's a great skeptical hat observation.

Lenay Dunn

Also, thinking if this was an intervention you really were interested in pursuing, and you have a large English learner population, not necessarily ruling it out because that was such a small study sample of English learners, but saying, "Well, maybe we look for another study to see if it has a larger English learner population or study sample." Because we wouldn't want to necessarily knock it out based on that.

John Rice

Right. Good point. I think the last thing I'll bring up here is, someone asked, "Are there tools for measuring or calculating effect sizes?" Effect size is really based on means and standard deviation, so the actual components of it aren't complicated, but there are some different ways of calculating it using those statistics. We'll probably include those as one of the resources that we send as a follow-up to this webinar, how to calculate it and any kind of maybe free software that's out there because it's a pretty common thing to calculate. That's a great question.

Lenay Dunn

Thanks, John. To step back and think about some questions you might ask as you look at all this information, there are a few key questions and elements. One is, which evidence level you think the study would need and why. That would be something you would probably discuss with some of your colleagues. Because it was a randomized control trial, it seemed to fit within the well-designed and well-implemented parameters to most degrees. The attrition would be something I'd look into a little bit more. But it might be likely to meet a tier one. That would be something you could have a discussion about with your colleagues. Thinking about potential subjectivity, bias considerations or questions, we don't really have to dig too much into this, and you're going to have an example in your small groups where you will do this, but I didn't see any major red flags in terms of that.

It had that program-embedded assessment, but it also had another established measure. And those design or outcome considerations, again, we talked through some of those things that would make me want to dig a little bit deeper. The four elements, the source, design, findings, and relevance, this is just kind of a reminder for you of the components that we looked at

under each of those. And I'm going to hand it over to my colleagues at the Arizona Department of Education because they're going to talk a little bit about how they've used these four components. And as John mentioned early on, we developed this based off of work that we've been doing with the Arizona Department of Education and this team. So I'm going to hand it over to my colleagues, Dr. Amy Boza and Sarah Bondy, to tell us a little bit about how they've used some of these components.

Sarah Bondy

Perfect. Thank you, Lenay. I'm going to kick us off just to set the stage a little bit and how we do utilize these tools and some of the work we've done with REL West. You can see on your screen, we are with the Department of Education, directly with Move On When Reading, which here in Arizona is our K-2, primarily, our early intervention policy so that we can set all of our students up to read on grade level by 3rd grade, and we lead this team of four. There are four of us, one of the smallest teams in the nation, and we support over 1,400 schools in our state, which impacts about 300,000 K-3 students each year. And we're really fortunate that we have this legislation that is funded with \$45 million annually to support our educators, all of our LEAs, which of course, trickles to our students. And we have quite a big lift in the state to support our schools in regards to Move On When Reading, and that comes with curricular decisions, core interventions, supplemental, all of their assessments. We help to guide and provide support in that way.

Professional development, supplemental staff, support staff in reading, and overall, their RTI systems. And one of the biggest tasks and lifts is really that curricular decision-making process. We are very much a local control state, so districts do have the autonomy to choose what they feel is best, however, we do have a set criteria, so we do provide a lot of support year-round to our schools. And one of the key criteria is that the materials meet one of the top three tiers of ESSA. Through that process, and as you've heard since 2015, really impact our state in 2016, we've been doing this for some time, yet still are starting from square one in really understanding what ESSA's all about and the intent of it. Amy's going to speak a little bit more about the resources we've been fortunate to be able to create with REL West and how we use them in the state.

Amy Boza

Thank you, Sarah, for just sharing the landscape of what we're working with. She's absolutely right. ESSA is not a new thing in Arizona, it's been on the stage for a little bit, but it's taken some time for all parties involved to kind of get a grasp and understand what it includes and what is expected in that realm. We have been super fortunate to have support and coaching over the past year and a half with our colleagues, Drs. Lenay Dunn and Eric Ambroso. They have been working hard with us, just doing a lot of coaching and training, much like today's session, where we have been able to go through these different elements and review real research to say, "What do we think, and where are we moving?" So through their support and guidance, we've been able to really hone our skills so that we can support LEAs working through the process. I've written down a few things, so I keep checking my notes just to make sure that I capture things.

We've worked through not only the intent and the legislative components, the non-regulatory guidance, all of those documents, but we've also worked, what does that mean moving down? What do programs and interventions look like from one space to another? Like I said, we've

evaluated lots and lots of studies. Even now, when we get stuck looking at studies, we call them right up and say, "Hey, can you take a look and help us out with this?" And you don't give us the answer, they make us work for it a little bit, so that's good. [laugh] And then, most importantly, I think, and part of this conversation is really about how this impacts your continuous improvement, and that's a big message, too, that we've been signaling. We're thinking about the ESSA levels, but why are we choosing these things, and how do they meet our specific needs? And that kind of fits into some of this criteria in this tool.

And that leads also to many of the tools that are available, tool five, tool six, and then this newest infographic, there are lots of tools that have really helped us, and then we're using those tools to support others. Some things that have come out, also, are of course, the increased use of those tools. Most recently, this infographic that we're using today. Also, through this process, we have been able to create a state-vetted list of [0:46:51] materials and intervention materials. The information is provided by vendors, but we're able to say, "These tools, based on the research, meet the criteria of ESSA." We've been able to create these vetted lists so that LEAs or schools that are looking for a curricular tool can say, "OK, well, at least we know that they have this one piece. Now, we have to do some work to figure out if it meets our criteria or need."

That has been a helpful product or outcome of our time together. Also, through that process, just reinforcing the message to be a good consumer. Like John mentioned, there's sometimes, whether conscious or unconscious, bias from developers and publishers. We just want to be sure that we're looking at real research and not their summaries or those kinds of things. And then, one of our most exciting things is that we've created a video series with our team and our coaches from WestEd, and Sarah's been able to put that in the links. So we have a mini video series to support our LEAs in each of the steps of the process. We have been super excited to grow our own understanding and to be able to reciprocate that learning to our LEAs across the state. And so, we hope to continue our partnership with our special friends, and we're going to continue to grow and provide as many tools and resources as we can for LEAs, including adding some supports for using this newest tool that we have. Thank you for having us.

Lenay Dunn

Well, thank you, and I really appreciate the partnership and that we've been able to really develop these ideas together, really crystallize some of this to help schools and districts in their work to find the best strategies and interventions for their needs and contexts. So thank you, we really value that.

Tran Keys

Sarah, if you could just give me one share out from your group. I know you talked about a lot of things, but could you just give us one share out from your group in terms of what you were discussing under consideration of source and design? Maybe something that was discussed a lot in your group.

Sarah Bondy

Yeah, absolutely. For the source of our fictitious study, knowing that it was published internally, there was some conversation back and forth of some who would automatically discount this and some who said they wouldn't automatically use that as a determination to discount the study, so that was a good kind of back and forth. Just because they published the

intervention and published the study, what direction do we take that? So that was great. As for design, we all know there were a lot of unknowns. For it to land at that lowest level, most likely descriptive, because we didn't have the information needed, there weren't those groups and that baseline equivalency established with that missing pretest information.

Tran Keys

Great. Thank you, Sarah. When we came up with this fictitious example-normally, you would be reading a whole article. We didn't want to lose folks, so we didn't make you read an article for an hour. But we wanted to just give you key points. And the idea, and what we were seeing in the note-taking also, is that you're asking questions like, "Wait a minute, I can't really answer this," or, "I'm not really confident to make a conclusion because there's not enough information." And that's just the reality as you review a study, so we purposely did want you to have those conversations. So thank you, Sarah. One more share out. Amy, would you mind sharing out for your group? Since we covered the considerations of source and design with Sarah's group, would you talk a little bit about the considerations of study findings and relevance for us?

Amy Boza

Yes. We had great conversations through all of them, but the statistical significance-there was such a wide berth of different groups. While it was noted that was there, we kind of landed on, yes, it would be more confident in both that and the effect size, but we would want to see some more consistent information across subgroups perhaps. Relevance, this was more of a conversation, particularly the EL population. Lots of our settings and lots of people represented in my group don't teach in places that have only 3% EL population. Having something like that was very-not concerning, but our confidence in the study as it aligns to the populations that we see, it probably wouldn't be what we would need. Also, Lenay and REL West team, I have dubbed those teeter-totters, the little confidence scales, so we did a lot of higher on one side, and lower on the teeter-totter, just in case we want to give them a name.

Tran Keys

I love it, Amy. I never thought about it that way. Now, that's going to be in my mind. John, sorry to put you on the spot, but folks are just really interested in the effect size. People have heard about statistical significance for years and years. Effect size is not new to statistics and methodology, it's just newer for education research. Am I correct? And could you say a little bit more about the translation of this point 3-5 for education?

John Rice

Sure, sure. Let me just preface this by saying conveniently, or coincidentally, maybe, REL West is about to release a two- to three-page infographic just about effect size, what it means, what influences effect size in terms of study designs, and then how to interpret effect size. It's by no means comprehensive, you could read all the statistical manuals about effect size, but that will be coming out in the next week or two, and so we will make sure to forward that to this group. But just to say briefly, because we don't have much time, effect size is basically the measurement of the strength or magnitude of a program against business as usual or the group that didn't get the program.

And I think the most important thing to say in just the few seconds that I have is that unlike statistical significance, where we have this convention of 0.05 or less, there's no hard and fast rule for effect size. It just depends. That 0.35 is considered a pretty sizable effect and something you might want to take into consideration. But it also depends on the cost of the program. If it's a really cheap program, 0.35 is great. If it's a real expensive program, maybe you want to find something else that has a greater effect size or is cheaper. I really can't do it justice, but it's all just to say that 0.35 probably isn't too shabby, but there's no hard and fast rule with effect size. And we'll make sure to get that resource out to all of you when it's released very soon.

I just want to say thank you to everyone who attended. It was great to see the turnout, it was great to see the level of engagement. We really appreciate your comments and questions. It's awesome that people want to know more, demystify the research. So thank you. I want to thank my colleagues at REL West and all our facilitators who monitored our breakout rooms from WestEd and REL West, and I especially want to thank our partners at the Arizona Department of Education, Sarah Bondy and Amy Boza. They've been great partners all along, and we really appreciate them being part of this webinar today and sharing their experience with this tool we've created. Thank you. And I hope everyone has a great day and a great week.

[End at 0:55:44]