

How effective is my college access program? Using the Every Student Succeeds Act (ESSA) to measure levels of evidence

Virginia College Access Network (VCAN) Conference, Richmond, Virginia
December 4, 2019



Meet your session facilitators



Jessica Mislevy, Ph.D.

REL Appalachia Partnership Lead
SRI International



Neil Seftor, Ph.D.

REL Appalachia Project Staff
SRI International



Jill Marcus, M.A.,

REL Appalachia Project Staff
Education Development Center

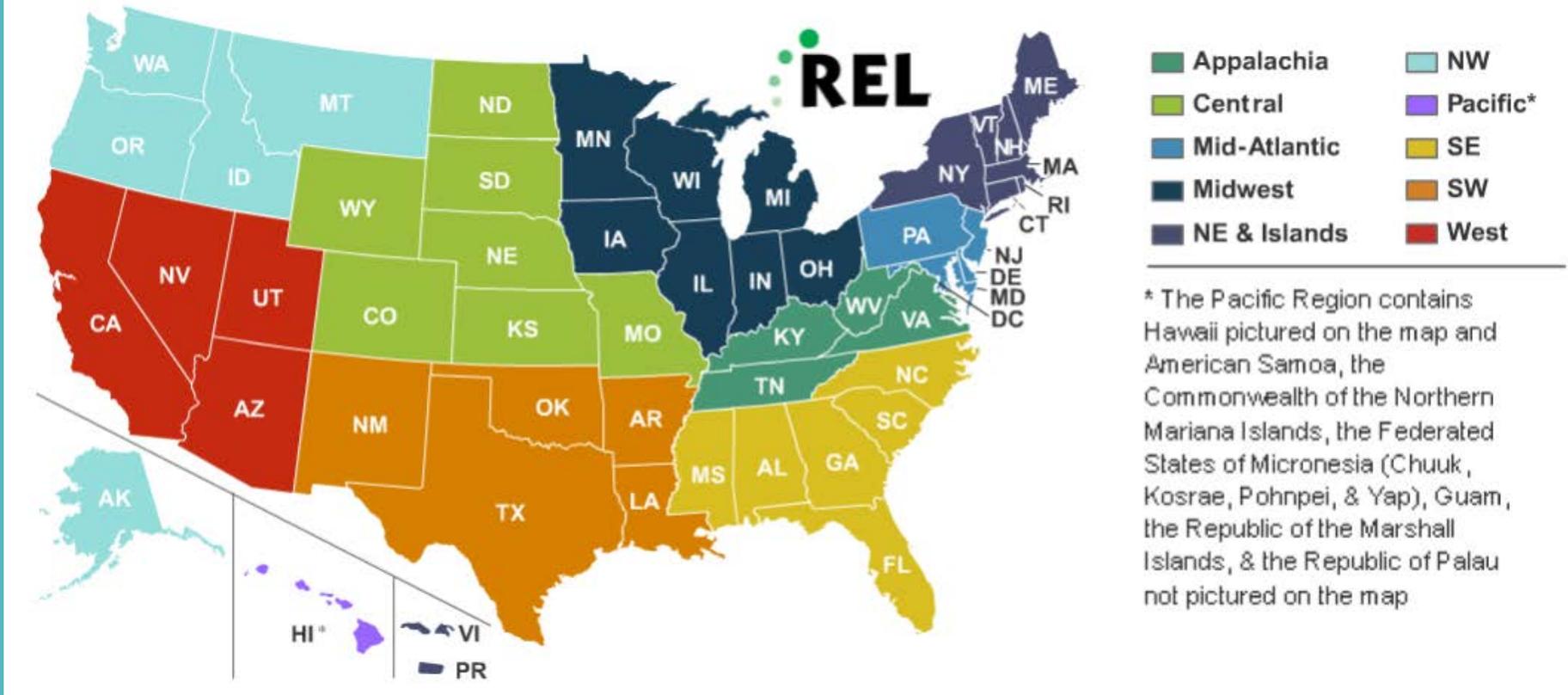
Session agenda



Topic
Introduction to REL Appalachia and our work on the transition to postsecondary
Overview and importance of ESSA levels of evidence
Orientation to evidence review project
Feedback on design plan: Part I
Feedback on design plan: Part II
Next steps

Introduction to REL Appalachia and our work on the transition to postsecondary

The Regional Educational Laboratories



The 10 Regional Educational Laboratories (RELs) work in partnership with stakeholders to conduct applied research and trainings.

The REL mission is to support a more evidence-based education system.

Virginia Improving Postsecondary Transitions partnership



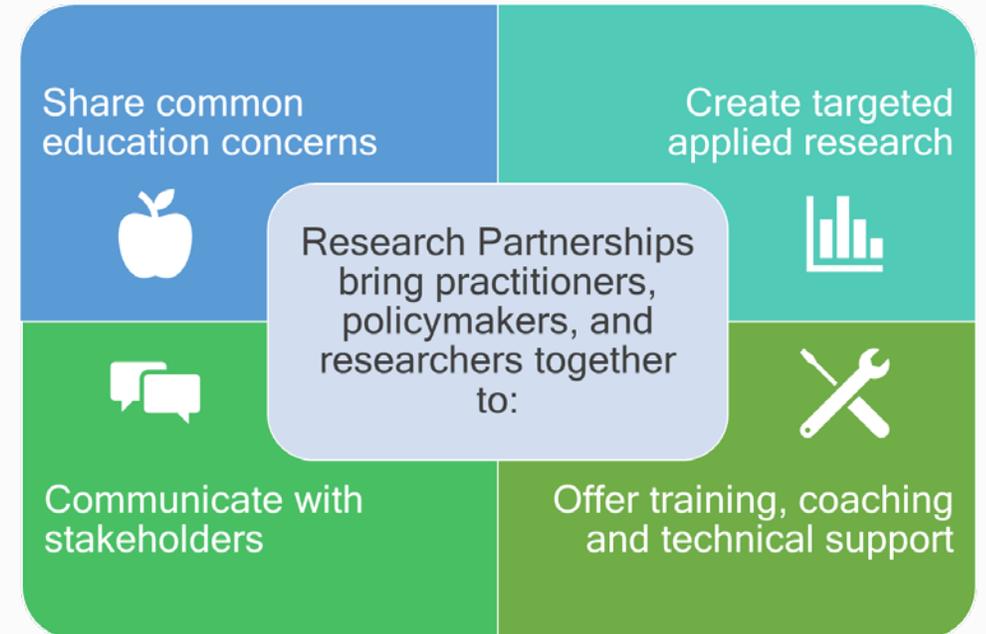
Leads: Jessica Mislevy & Deborah Jonas

Goal

- To identify, develop, and support the use of **evidence-based practices that strengthen high school graduates' transition** to college and careers.

Partners

- Virginia Department of Education (VDOE).
- Virginia Community College System (VCCS).
- State Council of Higher Education for Virginia (SCHEV) representatives.



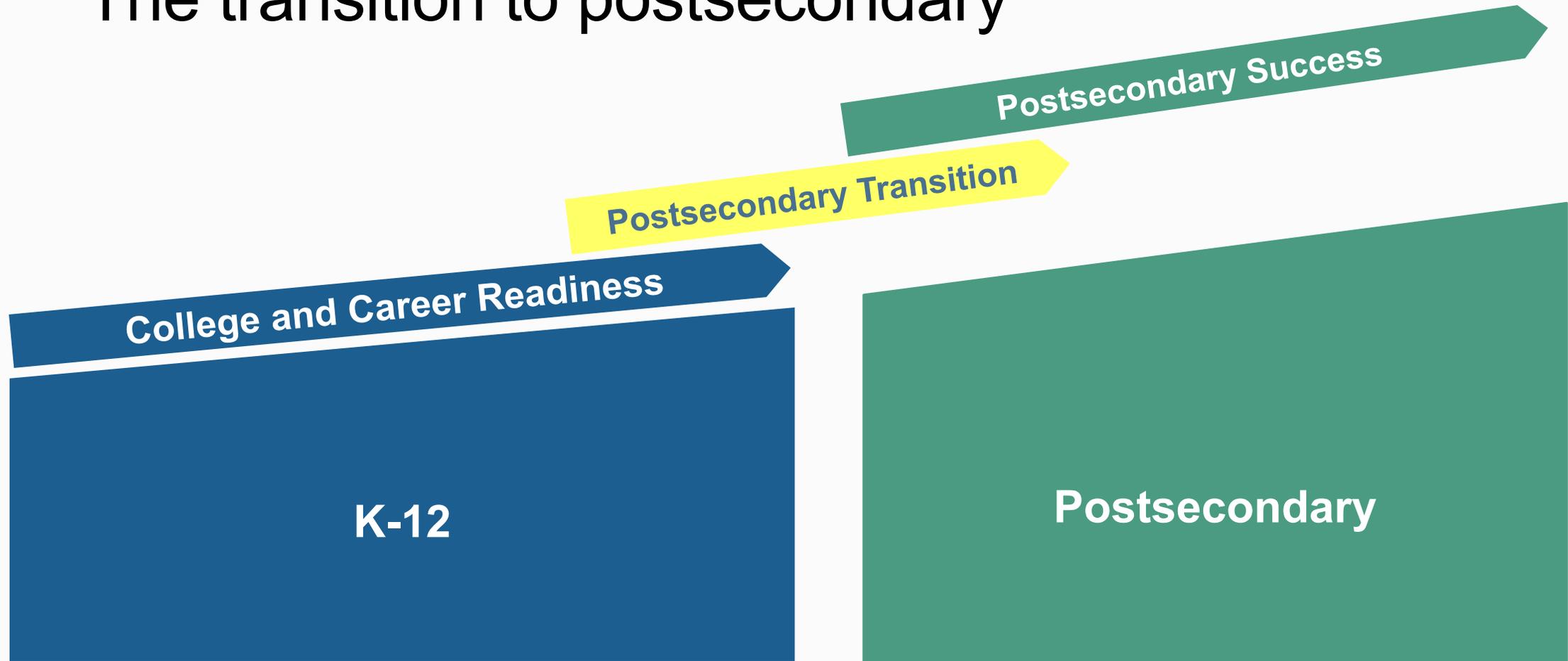
Why focus on postsecondary?

By 2020, 67 percent of jobs in Virginia will require postsecondary education or training.

- 30 percent of jobs in Virginia will require some college, an associate's degree, or a postsecondary vocational certificate.
- 23 percent of jobs in Virginia will require a bachelor's degree.
- 13 percent of jobs in Virginia will require a master's degree or more.



The transition to postsecondary



Common barriers for students to transition to postsecondary education

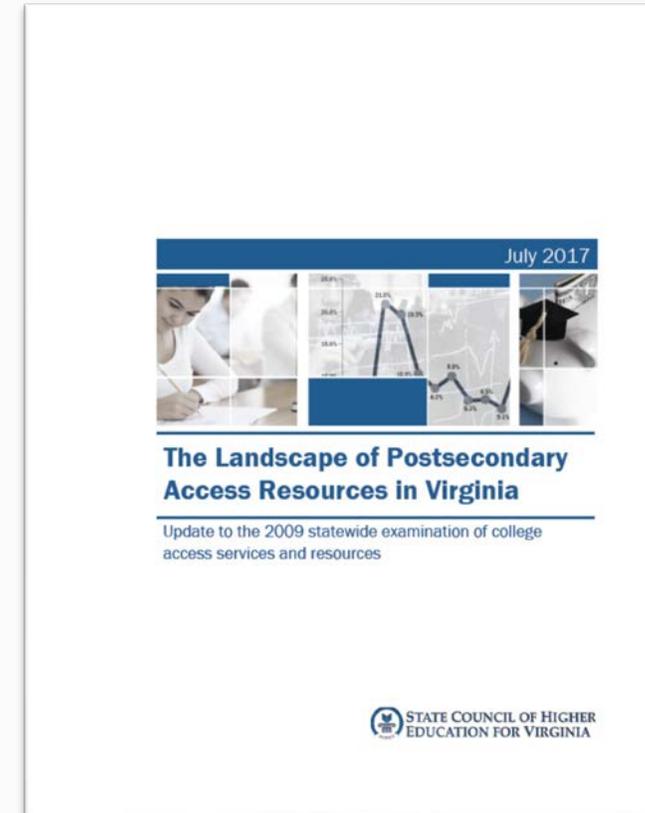
- Insufficient academic preparation
 - Curriculum and preparatory pathways lacking rigor.
- Limited financial resources
 - Missed aid deadlines.
- Lack of “college knowledge”
 - Unaware of requirements, expectations, norms, etc.
- Navigating social and emotional aspects of the transition
 - More responsibility.
 - New community.
 - Potentially being away from home.



Offering support: Virginia college access providers

In a 2017 SCHEV-commissioned study, the Metropolitan Education Research Consortium (MERC) analyzed data on the **services and resources available** to help students across the state enter postsecondary education.

All together, MERC identified **115 organizations** providing access services across 128 of Virginia's 131 school divisions.





Questions remain for state education agencies

REL AP's partners at VDOE, VCCS, and SCHEV seek to understand:

- To what extent are providers using **evidence-based practices** as a part of the programs they offer to increase access to and success in postsecondary education and training programs?
- To what extent are providers **evaluating** new and innovative programs to determine **effectiveness** and at what level of **rigor**?

College access providers also have questions and needs concerning their use of evidence-based practices

We need more systematic guidance on program implementation and evaluation practices across our multi-site program to improve quality and consistency. Everyone's doing their own thing, and we need to ensure each site offers at a minimum a core of evidence-based strategies.

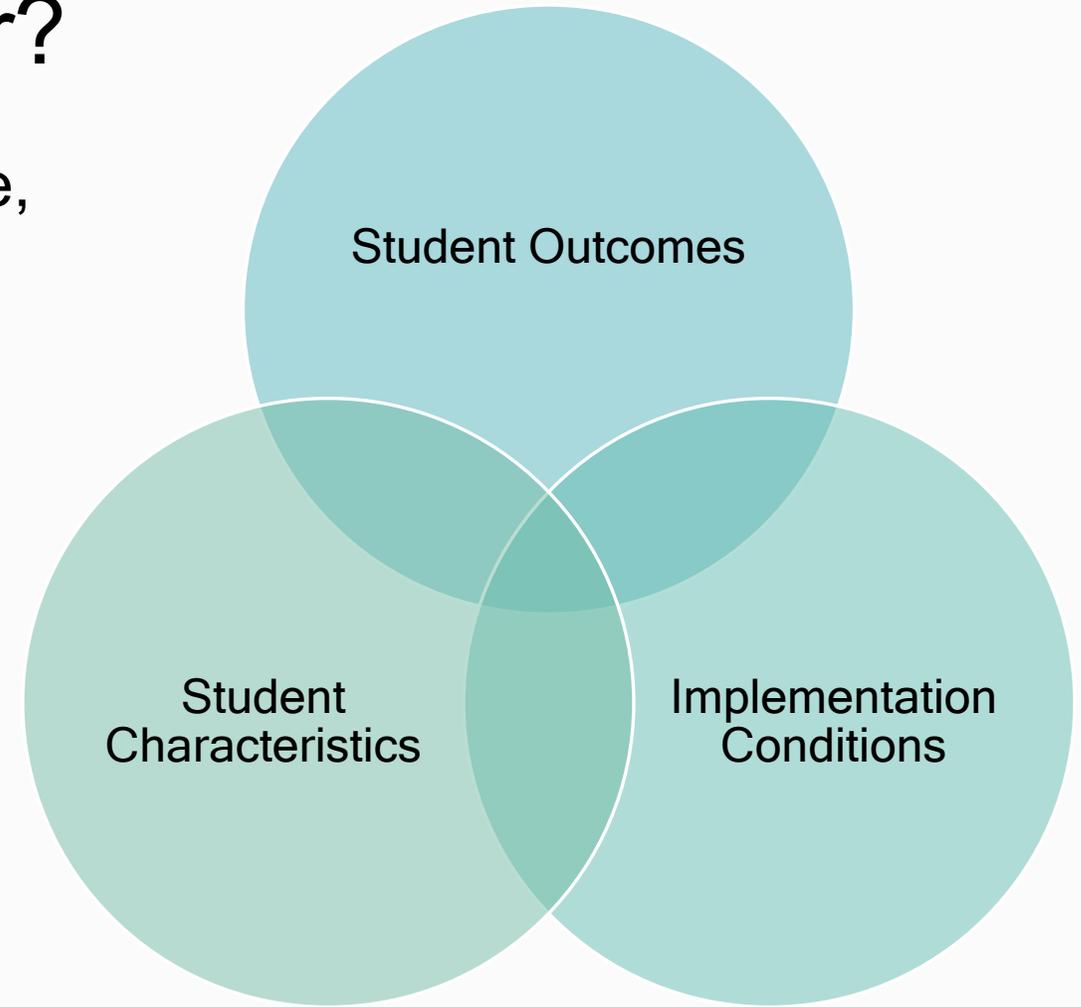
We need to do a better job “telling our story” by providing the kinds of information and data that state, federal, and foundation funders want—and expect—to make funding decisions.

We need a way to talk with people working in schools about what is and isn't working based on data and evidence.

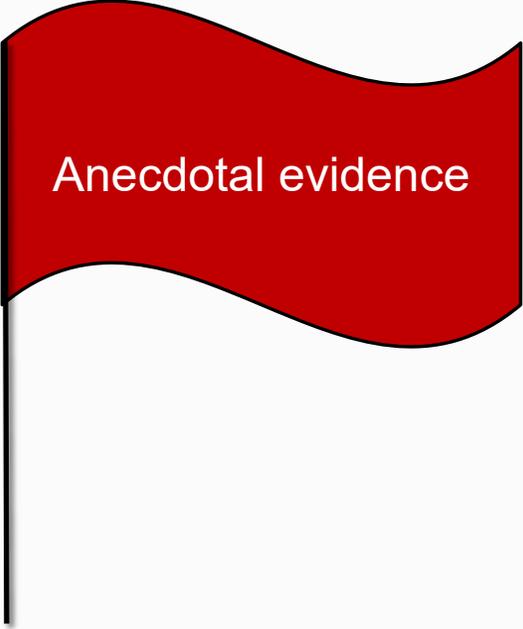
Overview of the Every Student Succeeds Act (ESSA) levels of evidence

Why does evidence matter?

Before you adopt a program or practice, you want to know whether it works, for whom, and under what conditions.



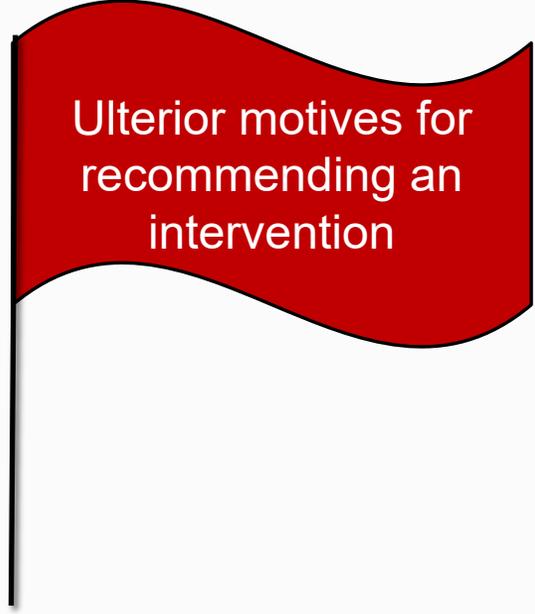
When to be cautious about “evidence” claims: Some potential red flags



Anecdotal evidence



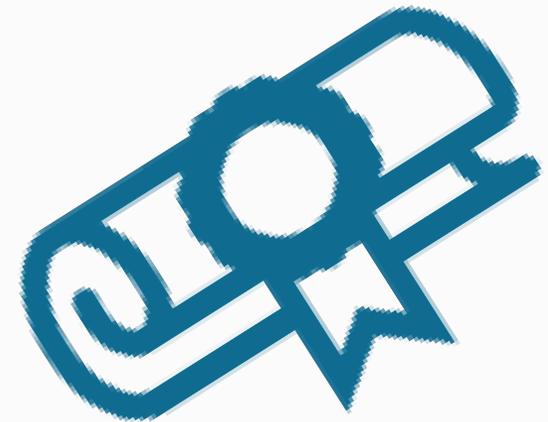
Source does not verify
quality of evidence



Ulterior motives for
recommending an
intervention

The Every Student Succeeds Act (ESSA)

- 2015 reauthorization of the Elementary and Secondary Education Act (ESEA), the nation's educational law governing K-12 public education.
- Requires rigorous and comprehensive state-developed plans designed to close achievement gaps, increase equity, improve the quality of instruction, and increase outcomes for all students.
- Directs educators to implement interventions grounded in research.
- Requires education leaders to include evidence-based practices in improvement plans for low-performing schools.



Defining evidence-based interventions

The Every Student Succeeds Act (ESSA) defines evidence-based interventions as *"[P]ractices or programs that have evidence to show that they are effective at producing results and improving outcomes when implemented."*

Tier I – strong evidence: supported by one or more well-designed and well-implemented **randomized control experimental studies**

Tier II – moderate evidence: supported by one or more well-designed and well-implemented **quasi-experimental studies**

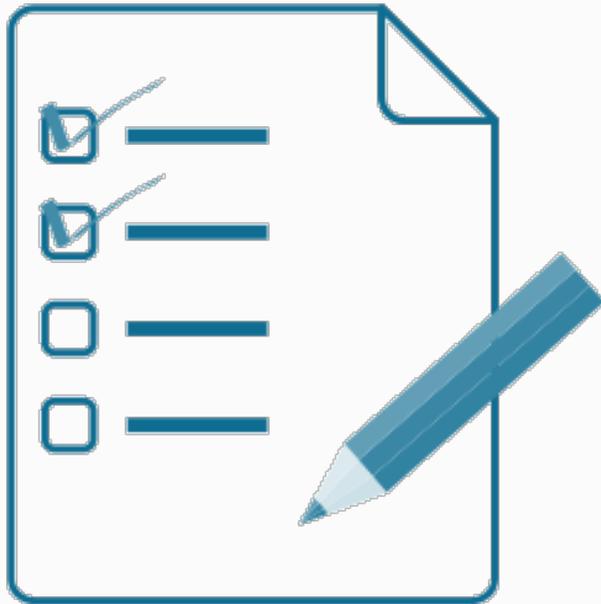
Tier III – promising evidence: supported by one or more well-designed and well-implemented **correlational studies**

Tier IV – demonstrates a rationale: has a **well-defined logic model**, is **supported by research**, and **efforts to evaluate** are under way

For full description of ESSA evidence standards, see <https://ed.gov/policy/elsec/leg/essa/guidanceusesinvestment.pdf>

Evidence review project plan

Planned project



REL AP will help our partners at SCHEV, VCCS, and VDOE codevelop and pilot a **systematic evidence review protocol for practitioner use** to examine:

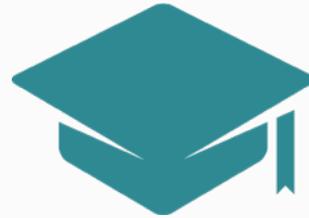
- College access providers' adoption of evidence-based postsecondary transition strategies.
- The rigor with which they are evaluating new and innovative strategies.
- Indicators of program effectiveness.

How would you use the evidence review protocol?



State agencies could use the information to target support for such activities as:

- Selecting interventions.
- Designing and carrying out evaluations.
- Sponsoring/identifying funding for evaluations.



Program providers could use the protocol as a self-assessment to:

- Strengthen their practices and the rigor of their evaluation efforts.
- Identify new and innovative strategies that merit further study.



School/division leaders could use the information to:

- Understand what qualifies as strong evidence.
- Inform their decisions about whether to adopt a program in their context.

A use case scenario

Review protocol contents outline

- Introduction to the protocol and instructions for use.
- Part I: Evidence-based programs and practices for supporting the transition to postsecondary.
- Part II: Assessing the level of evidence for postsecondary transition programs and practices.
- Other considerations and additional resources.

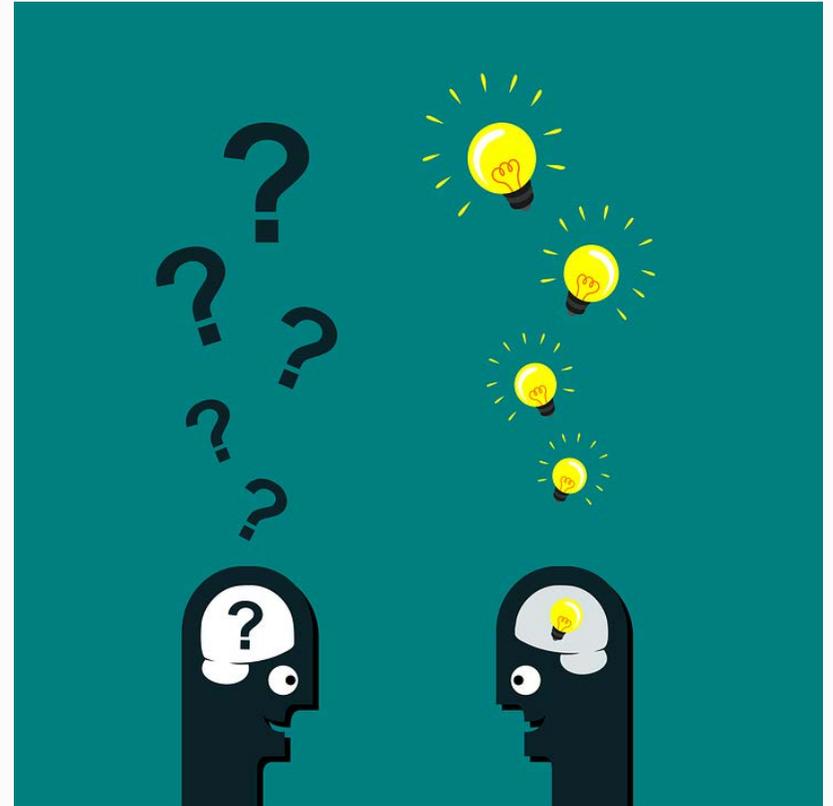
I'm a college access provider and I want to know the evidence base for the program I'm offering. I decide to use this protocol to conduct a self-assessment of my program.

First, I look through the list of evidence-based programs and practices provided in Part I of the protocol. Is what I'm doing already on the list? If so, what level of evidence has it demonstrated in prior rigorous evaluations?

If it's not already on the list, I proceed to Part II of the protocol to assess the level of evidence it may meet. Here I follow a set of guided questions. For example, "Does the program have a theory of change for why it is likely to work? If the program has not already been evaluated, are evaluation efforts under way? If yes, then the program meets 'Demonstrates a Rationale' (ESSA Tier IV)."

Now that I've learned my program meets ESSA Tier IV evidence, I'll look at some of the additional resources to determine my next steps, such as using an evaluation toolkit to design an evaluation of my program that could provide "Promising" ESSA Tier III evidence.

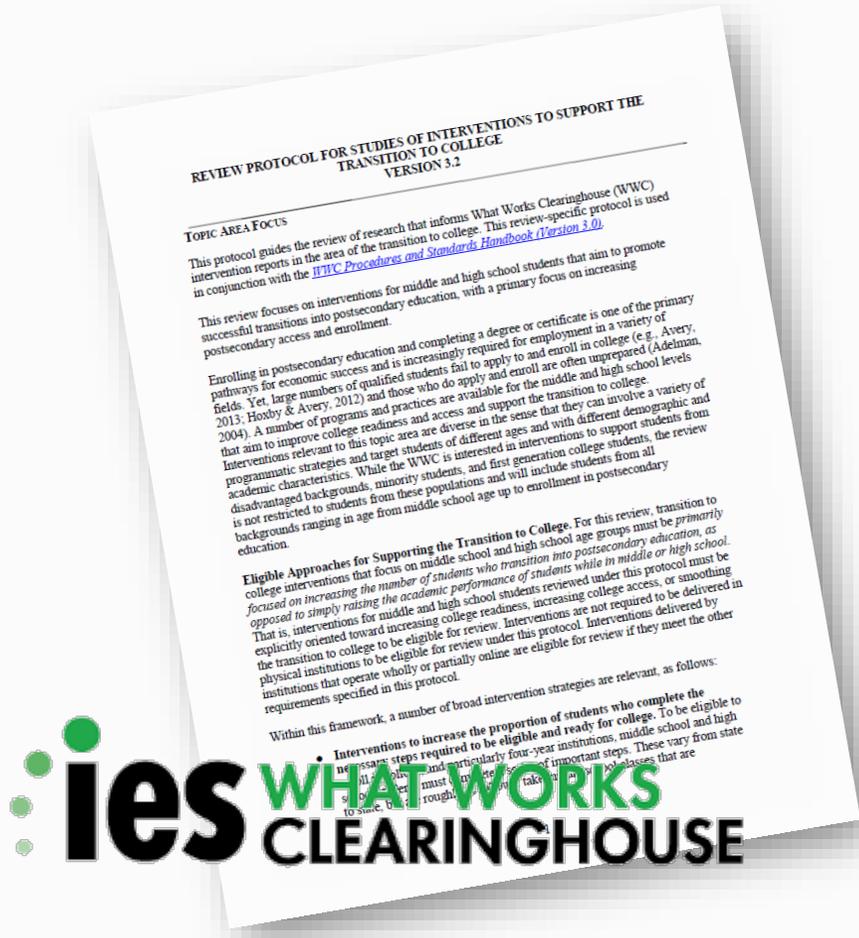
Questions?



Part I: Evidence-based programs and practices for supporting the transition to postsecondary

“IS WHAT I’M DOING ALREADY ON THE LIST? IF SO, WHAT LEVEL OF EVIDENCE HAS IT DEMONSTRATED IN PRIOR RIGOROUS EVALUATIONS?”

What Works Clearinghouse

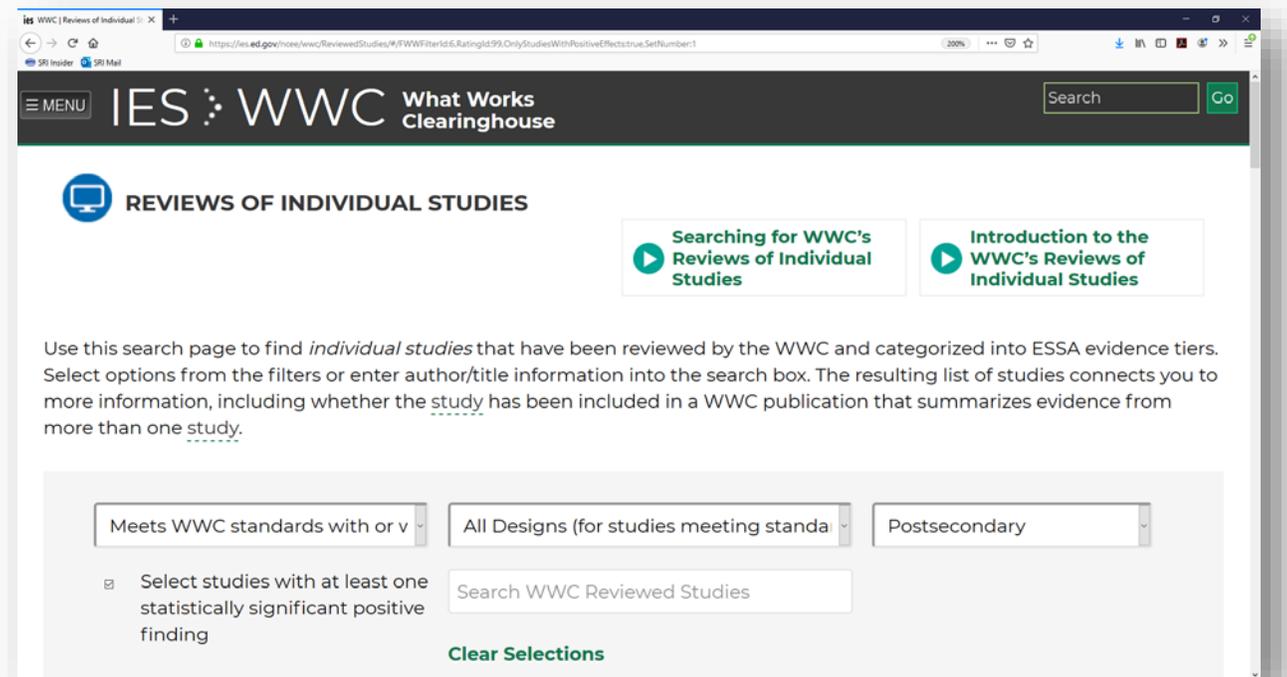


- The ESSA levels of evidence were informed by the What Works Clearinghouse (WWC).
- The WWC provides study ratings based on the strength of evidence using a consistent and transparent set of standards.
- The website includes a search page to find studies that have been reviewed by the WWC and categorized into ESSA evidence tiers.

Search and summary in progress

With input from our partners at SCHEV, VCCS, and VDOE, REL AP is:

- Defining criteria to **systematically search the WWC** for evidence-based postsecondary transition programs and strategies.
- Determining an approach to **summarize the findings** in a way that is understandable, useful, and sufficient for target users.



Evidence map (handout 2)

- Rows are programs/practices.
- Columns are outcomes affected.
- Cells include the ESSA evidence level (**strong** or **moderate**) and number of studies.

Every Student Succeeds Act (ESSA) Evidence Summary

Postsecondary Transitions Interventions with Strong or Moderate Evidence	Middle School Attendance	Middle School General Academic Achievement	Staying in High School	High School Attendance	High School General Academic Achievement	Progressing in High School	Completing High School	College Readiness	College Access and Enrollment	College Attendance	College Credit Accumulation and Persistence	College General Academic Achievement	College Degree Attainment	Employment and Labor Market
WWC Intervention Report: Dual Enrollment Programs (2017) Dual enrollment programs allow high school students to take college courses and earn college credits while still attending high school.			Strong 1	Strong 1	Strong 2		Strong 2	Strong 1	Strong 4		Moderate 2	Strong 1	Strong 5	
WWC Intervention Report: Summer Counseling (2018) Counseling services between high school graduation and college enrollment by college counselors or peer mentors via text messaging campaigns, e-mail, phone, in-person meetings, instant messaging, or social media.									Strong 5		Strong 1			
WWC Intervention Report: First Year Experience Courses (2016) Courses for first-year college students to support the academic performance, social development, persistence, and degree completion.											Moderate 3	Moderate 2	Moderate 1	
WWC Study Review of Intrusive Intervention Abelman & Molina (2001). Style over substance revisited: A longitudinal analysis of intrusive intervention.											Strong			
WWC Study Review of Early College High Schools AIR & SRI (2013). Early college, early success: Early College High School Initiative impact study.							Strong		Strong				Strong	
WWC Study Review of Dual Enrollment Programs An (2013). The impact of dual enrollment on college degree attainment: Do low SES students benefit?													Moderate	

- Each row represents evidence for an intervention designed to support successful transitions from secondary to postsecondary education and training programs.
- Outcomes for which there is **Strong** or **Moderate** ESSA evidence are indicated by a box with the number of supporting studies.
- An empty cell only indicates a lack of ESSA Strong or Moderate evidence; there could be lesser or negative evidence, or no research.
- The list includes only what has been reviewed by the What Works Clearinghouse (WWC) and is not comprehensive of all programs or studies.
- Links go to WWC intervention or study pages that may include population and setting information useful for determining overlap.
- For a full description of ESSA evidence standards, see <https://www2.ed.gov/policy/elsec/leg/essa/guidanceuseinvestment.pdf>.

These materials were produced for the Virginia College Access Network (VCAN) and were presented on December 4, 2019 at the VCAN conference.

Zooming in...

Postsecondary Transitions Interventions with Strong or Moderate Evidence	<i>Middle School Attendance</i>	<i>Middle School General Academic Achievement</i>	<i>Staying in High School</i>	<i>High School Attendance</i>	<i>High School General Academic Achievement</i>
<p>WWC Intervention Report: Dual Enrollment Programs (2017)</p> <p>Dual enrollment programs allow high school students to take college courses and earn college credits while still attending high school.</p>			<p>Strong 1</p>	<p>Strong 1</p>	<p>Strong 2</p>
<p>WWC Intervention Report: Summer Counseling (2018)</p> <p>Counseling services between high school graduation and college enrollment by college counselors or peer mentors via text messaging campaigns, e-mail, phone, in-person meetings, instant messaging, or social media.</p>					



Discussion activity

1. How could you envision using this information about evidence-based programs and practices?
2. Is the summary information understandable?
3. Is anything missing that you think users would need or want?
4. What do you think it means if your program/practice is not on the list?
5. Other strengths or areas in need of improvement?

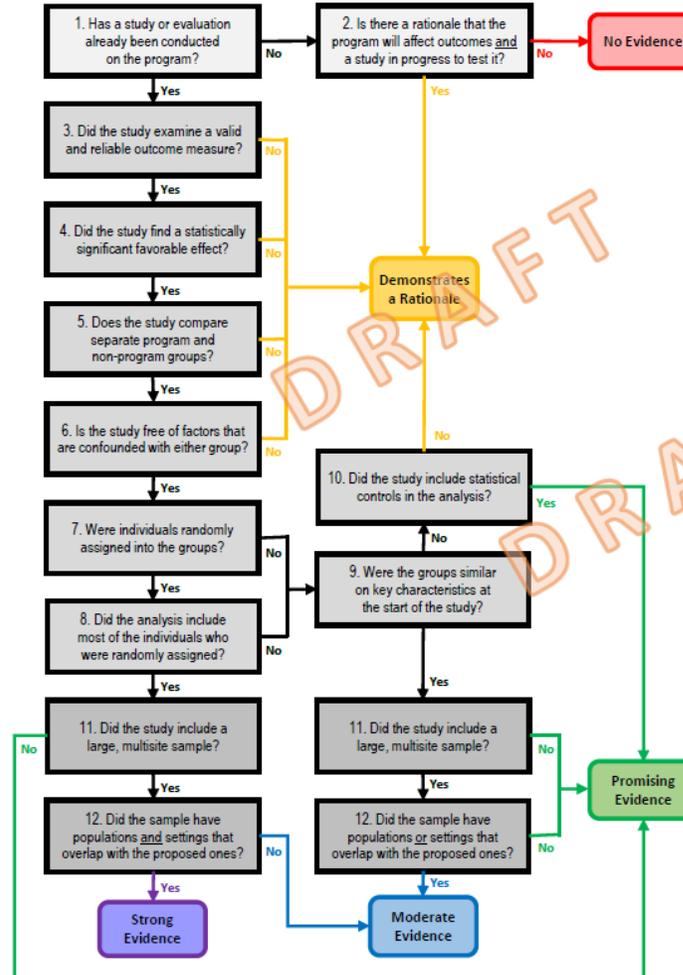
Part II: Assessing the ESSA level of evidence for programs and practices

“IF THE PROGRAM OR PRACTICE I OFFER IS NOT ALREADY LISTED IN PART I, HOW DO I ASSESS THE LEVEL OF EVIDENCE IT MAY MEET?”

Evidence review flow chart (handout 3)

- On the left: Guided set of questions to assess ESSA evidence levels.
- On the right: Supporting information and links to additional resources.

Assessing Every Student Succeeds Act (ESSA) Evidence Levels



Does the program have evidence?

2. Is there a rationale that the program will affect outcomes?
Is there reason to believe that there is a relationship between key activities of the program and relevant outcomes? Studies will often describe this relationship through a logic model or theory of change. [\(For more information\)](#)

How strong is the evidence?

3. Does the study examine a valid and reliable outcome measure?
A measure is **valid** if it accurately measures what it claims to measure, such as whether a student enrolled in college. It is **reliable** if results are stable and consistent across people, settings, and time, such as with a standardized test.

4. Does the study find a statistically significant favorable effect?
An effect is **favorable** if an outcome changes in a way that reflects improvement, such as raising a score or lowering the dropout rate. A finding is said to be **statistically significant** if the probability that an observed difference could have occurred by chance alone is less than five percent. This is often reported using a p -value, with statistical significance indicated by $p < 0.05$.

6. Is the study free of factors that are confounded with either group?
A component of a study is a **confounding factor** if it is completely aligned with one group and completely excluded from the other, such as when the program group is from one school and the non-program group is from a different school. This makes it impossible to tell if any difference in outcomes is due to the program, the confounding factor, or both. [\(For more information\)](#)

7. Were individuals randomly assigned into the groups?
Individuals are considered to have been randomly assigned if they are assigned to study groups entirely by chance and everyone has some possibility of being assigned to each study group.

8. Did the analysis include most of the individuals who were randomly assigned?
Random assignment creates groups at the start of a study that are assumed to be similar. Attrition occurs when an analysis does not include all the randomly assigned individuals, such as due to moving to another district or being absent on the day of an assessment. If there is high attrition for the whole sample, or differences in the rates of attrition for the study groups, the groups used in the analysis may no longer be similar. [\(For more information\)](#)

9. Were the groups similar on key characteristics at the start of the study?
When groups are not randomly assigned or there is high attrition, the study needs to demonstrate that the groups were similar. The groups are said to have **baseline equivalence** if the study shows that the groups of individuals used in the analysis were similar on characteristics that could be observed before the study, such as prior achievement and socio-economic status. [\(For more information\)](#)

10. Did the evaluation include statistical controls in the analysis?
If the groups are not similar, studies may try to account for these differences by including measures from the start of the study in the analysis. Examples of common statistical controls include analysis of covariance (ANCOVA), hierarchical linear modeling (HLM), and regression adjustment.

How does the evidence apply to me?

11. Did the evaluation include a large, multi-site sample?
An analytic sample is **large** if it includes at least 350 individuals or at least 50 groups that contain 10 or more individuals. A **multi-site** sample consists of more than one site, where site can be defined as a school, local education agency, locality, or state.

12. Did the sample have populations and/or settings that overlap with the proposed ones?
When a study is conducted with a population or setting like those where a program is proposed to be used, the findings may be more relevant. A study can demonstrate **overlap** with the proposed context if it shows that the program has a statistically significant favorable effect on the specific population or subgroup of interest, or in the same type of setting.

This guide provides information that can help determine the highest possible evidence level for a program, but additional technical analysis may be needed to make a firm determination. For a full description of ESSA evidence standards, see <https://www2.ed.gov/policy/elsec/leg/essa/guidanceseinvestment.pdf>.

These materials were produced for the Virginia College Access Network (VCAN) and were presented on December 4, 2019 at the VCAN conference.

A delicate balance



For the protocol, we'll need your help to balance factors such as:

- Accuracy of technical concepts with accessibility of language, and
- Level of detail provided with length and complexity of protocol.

The goal is to produce a data collection protocol that is **useful** and **usable** for our partners and their stakeholders.



Application activity

Instructions

1. Read through the study description (handouts 4 & 5).
2. Use the flowchart (handout 3) to determine highest ESSA evidence-level rating this study is potentially eligible for based on the information you have.
3. Jot down notes with any questions, thoughts, reactions, etc., to the process.

Now let's go over the WWC's answers

Question	Summer Bridge	School Counseling
1. Has a study or evaluation already been conducted on the program?	Yes	Yes
3. Did the study examine a valid and reliable outcome measure?	Yes	Yes
4. Did the study find a statistically significant favorable effect?	Yes	Yes
5. Does the study compare separate program and non-program groups?	Yes	Yes
6. Is the study free of factors that are confounded with either group?	Yes	Yes
7. Were individuals randomly assigned into groups?	No	Yes
8. Did the analysis include most of the individuals who were randomly assigned?		Yes
9. Were the groups similar on key characteristics at the start of the study?	Yes	
11. Did the study include a large, multisite sample?	No	Yes
12. Did the sample have populations and/or settings that overlap with the proposed ones?		???
Highest ESSA evidence level based on available information	Promising	Strong



Discussion activity

1. Thinking about your first impression: How user-friendly was the draft review protocol?
2. If we make the changes that you recommend, how user-friendly do you think the review protocol will be?
3. What did you like most about the protocol? What features should we be certain to keep?
4. What other questions or comments do you have?

Questions?

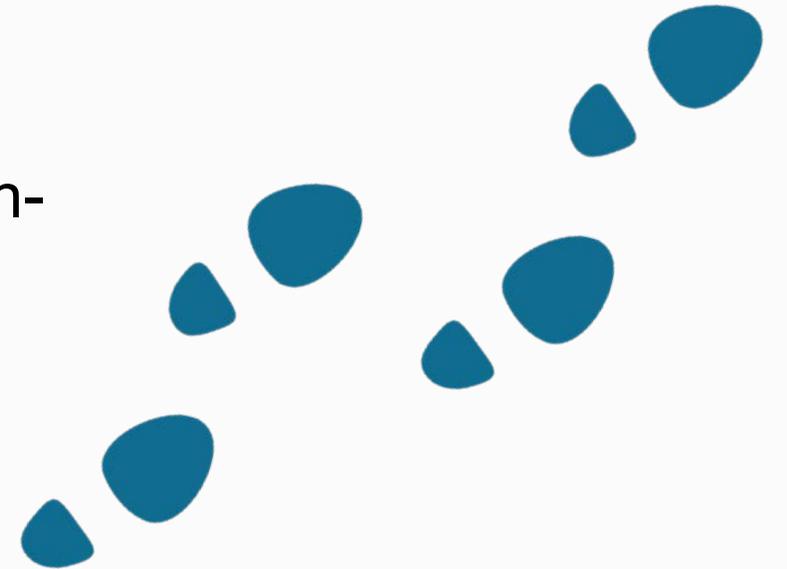


Next steps

REL AP and our partners at SCHEV, VCCCS, and VDOE are seeking volunteers to help pilot the protocol in early 2020.

- Do you want to know what level of ESSA evidence your programs or practices meet?
- Has your program been evaluated?
- Did the study compare separate program and non-program groups? (*preferred*)

If you are interested to participate, contact:
Jessica Mislevy (jessica.mislevy@sri.com)





Thank you!

Contact REL Appalachia

General inquiries:

RELAppalachia@sri.com

**Questions related to improving
postsecondary transitions in Virginia:**

Jessica.Mislevy@sri.com

Supporting slides

Four levels of evidence in ESSA

Strong

- At least one well-designed and implemented experimental study.

Moderate

- At least one well-designed and implemented quasi-experimental study.

Promising

- At least one well-designed and implemented correlational study.
- Includes controls for statistical bias.

Demonstrates a Rationale

- Well-specified logic model or theory of action
- Includes ongoing efforts to collect evidence.

Strong evidence

A well-designed and well-implemented experimental study

- Experimental studies require:
 - An intervention or treatment.
 - Subjects who receive the treatment and ones who do not.
 - Random assignment to the intervention and comparison groups.
- Well-designed and well-implemented experiments require:
 - Distinct intervention and comparison groups.
 - Appropriate randomization.
 - Valid and reliable measures.
 - Low attrition.
 - No confounds.
 - Large sample and multisite sample that overlaps with the populations and settings proposed to receive the intervention.

Strong evidence: Requires randomization

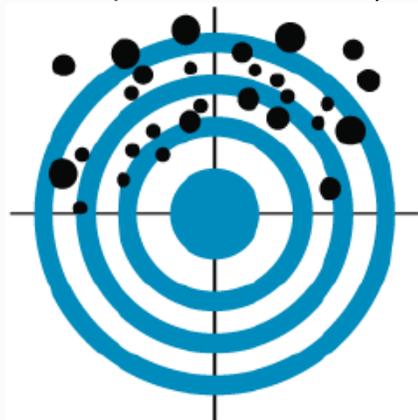
- Randomization is critical.
 - Random assignment ensures the treatment and control groups are as similar as possible.
 - Without randomization, unobserved characteristics may interfere.
- Random is defined as entirely by chance, and every subject has a chance to be in either group.
- Assignment occurs before the intervention.



Measures must be valid and reliable

- Researchers need to demonstrate that their outcome measures work.
- Two key criteria for evaluating measures are:
 - **Reliability**: degree to which a measure produces stable and consistent results.
 - **Validity**: extent to which scores from a measure represent what it intends to.
- WWC standards assume standardized (state) tests have face validity and are reliable.

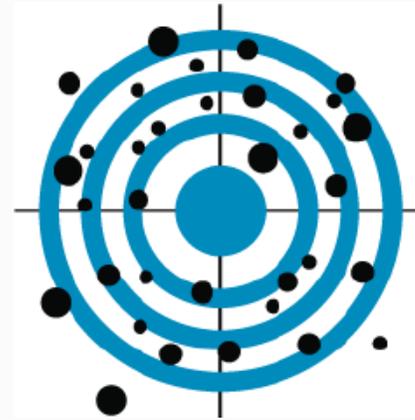
Figure adapted from Barford (2014)



Not reliable or valid



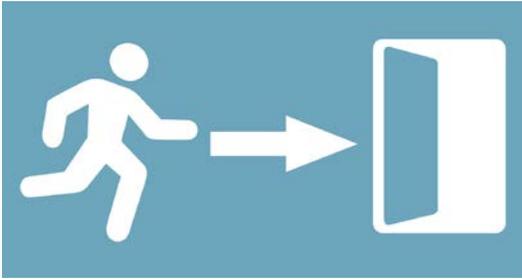
Reliable but not valid



Valid but not reliable



Valid and reliable



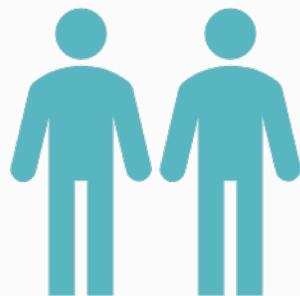
Strong evidence: Attrition matters

- Attrition is the loss of subjects from the study.
- Attrition is common, but when it is high, it compromises the outcome of random assignment.
- Two types of attrition
 - Overall: attrition for all study participants.
 - Differential: difference in attrition between intervention and comparison groups.
- WWC offers guidance on attrition standards,* but at a minimum always look at how many subjects dropped out of a study.

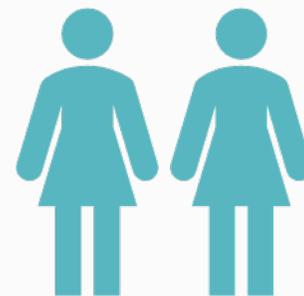
* https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_brief_attrition_080715.pdf

Strong evidence: Confounds matter, too

- Confounds are aspects of the experiment completely aligned to one group.
 - Ex. One classroom delivers the intervention, and one delivers the treatment.
 - Ex. Intervention studies all English learners (ELs), but treatment group has no ELs.
 - Ex. Intervention is part of a larger package.
- Confounds introduce an additional factor that compromises randomization.



Intervention



Comparison

Moderate evidence

A well-designed and implemented quasi-experimental (QED) study

- QEDs lack randomization.
- Instead, they leverage some natural change to create groups.
 - Ex. Comparing before and after a policy change.
- ESSA does not define well-designed or implemented.
- However, generally a well-designed QED has the following:
 - Strong break or forcing factor.
 - Valid and reliable measures .
 - Baseline equivalence.
- These types of studies can meet WWC standards *with* reservations

Moderate evidence: Baseline equivalence



- Baseline equivalence means that the intervention and comparison groups are similar on key characteristics.
- Without random assignment, the groups could differ.
- Researchers must take steps to demonstrate that the groups were equivalent **before** the intervention (i.e., at baseline).
- Baseline should be established on a characteristic similar to the outcome or correlated with it.
 - Ex. Prior year test score or a pretest.

Moderate evidence: Baseline equivalence (cont'd)

According to nonregulatory guidance,*

- If equivalence can be established, the study can be considered moderate evidence.
- If the baseline differences are small, statistical controls can be used.
- If the baseline differences are large, the study is not well designed and implemented.

* <https://www2.ed.gov/policy/elsec/leg/essa/guidanceusesinvestment.pdf>



Promising evidence

At least one well-designed and implemented correlational study that includes controls for statistical bias

- Correlational means the study looks at associations, not impacts.
- Such a study typically has one group and examines predictors of an outcome.
- Controls are other key variables related to the outcome but are not part of the research question.
- These types of studies cannot meet WWC standards.





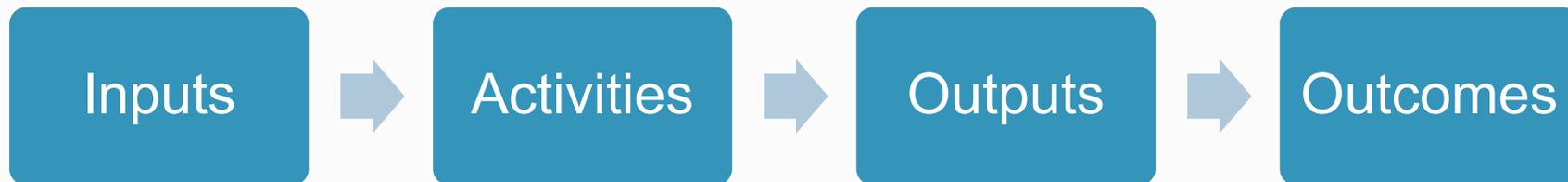
Promising evidence (cont'd)

- Correlational studies cannot measure impacts.
 - No random assignment.
 - No comparison groups.
 - No ability to establish baseline equivalence.
- Ex. Study shows students who report reading more books score higher on end-of-year test.
 - Controls for prior test scores, race, gender, and economic status.
 - But measures only the association between reading and scores.
 - Cannot conclude that assigning more books to read would increase scores.

Demonstrates a rationale

Well-specified logic model or theory of action

- Well-specified logic model or theory of action
 - What features of the intervention seem likely to result in improved outcomes?
 - What is the connection between the intervention and outcome?
- Includes ongoing efforts to collect evidence
 - How will you evaluate the results?



Does it work? Interpreting study findings

- Look for
 - **Positive direction:** favors the intervention group.
 - **Statistical significance:** the likelihood that the difference between groups is due to chance is less than 5 percent ($p < .05$).
 - **Substantive importance:** has an effect size—a standardized measure of the magnitude of an effect—of 0.25 or greater, regardless of statistical significance.





Study findings: What to look for

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	p-value
Castleman et al., 2014^a								
<i>Continuous first-year enrollment (%)</i>	Full sample	1,397 students	82.4 (na)	78.5 (na)	3.9	0.15	+6	< .05

Direction

Magnitude

Statistical Significance

Source: https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_summer_counseling_032718.pdf