



**Regional Educational  
Laboratory Southeast**

# Understanding ESSA Levels of Evidence & their Application to Program Evaluation

Dr. Barbara Foorman, Professor, FSU & Director, REL Southeast

Dr. Phyllis Underwood, Sr. Research Associate, REL Southeast



## **Acknowledgement and disclaimer**

Information and materials for this presentation are supported by IES/NCEE's Regional Educational Laboratory Southeast at Florida State University (Contract ED-IES-17-C-0011) as resources and examples for the viewer's convenience. Their inclusion is not intended as an endorsement by the Regional Educational Laboratory Southeast or its funding source, the Institute of Education Sciences.

In addition, the instructional practices and assessments discussed or shown in this presentation are not intended to mandate, direct, or control a State's, local educational agency's, or school's specific instructional content, academic achievement system and assessments, curriculum, or program of instruction. State and local programs may use any instructional content, achievement system and assessments, curriculum, or program of instruction they wish.



# Session Outline

- Overview of the four ESSA Levels
- ESSA and Non-Regulatory Guidance
- Requirements for strong evidence
- Requirements for moderate evidence
- Promising evidence and demonstrates a rationale
- Applying ESSA levels of evidence to program evaluation



# Overview of the Four Levels of Evidence

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



# ESSA and Non-Regulatory Guidance

- ESSA identifies four levels of evidence
- The Department of Education's Non-Regulatory guidance provides recommendations, resources, and criteria for each of those levels
- The following slides build on that guidance to provide more detailed information about each level
- However, states are free to interpret and apply the four ESSA levels differently



# What Works Clearing House

- Non-regulatory guidance on ESSA draws from WWC standards
- WWC rates studies as:
  - Meets standards without reservations = strong evidence
  - Meets standards with reservations = moderate evidence
  - Does not meet standards = promising or demonstrates a rationale
- WWC is a useful resource for finding and evaluating studies
  - <https://ies.ed.gov/ncee/wwc/>



# Strong Evidence

- A well-designed and implemented experimental study
- Experiments require
  - An intervention or treatment
  - Subjects who receive the treatment and ones who do not
  - Subjects assigned randomly
- What is a “well-designed and implemented” experiment as defined by the WWC?
  - Appropriate randomization
  - Attrition
  - Confounds
- These types of studies can meet What Works Clearinghouse standards without reservations



# 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



# Attrition

- Attrition is the loss of subjects from the study
  - Attrition is common but
  - When attrition is high it compromises the outcome of random assignment
- Two types of attrition
  - Overall
  - Differential
- WWC offers guidance on attrition standards\* but at a minimum always look at how many subjects dropped out of a study



# Confounds

- Confounds are aspects of the experiment completely aligned to one group
  - Ex. One classroom is the intervention and one is the comparison
  - Ex. Intervention students are all ELs but comparison group has no ELs
  - Ex. Intervention is part of a larger package
- Confounds introduce an additional factor that compromises randomization



# Summary of Strong Evidence

- ESSA requires a well-designed and implemented experimental study
- WWC standards for this are:
  - Treatment and control groups
  - Random assignment to groups
  - Low attrition
  - No confounding factors
- As a result, would likely meet WWC standards without reservations



# Moderate Evidence

- A well-designed and implemented quasi-experimental design (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 what is well-designed or implemented
- However, generally a well-designed QED has the following
  - Strong break or forcing factor
  - Baseline equivalence
- These types of studies can meet WWC standards with reservations



# Forcing Variable

- The forcing variable or break point is the factor that creates the different groups
- The variable should be consistent and clear
  - Ex. A stable cut-score on a test allows comparing those just above and below the cut-off
    - But not if exemptions are permitted
  - Ex. A change in policy allows comparing those before and after
    - But not if the policy change is implemented at different times



# Baseline Equivalence

- Baseline equivalence is whether or not the intervention and comparison groups are similar on a key characteristic
- Without random assignment, the groups could differ
- Studies must take steps to demonstrate the groups were equivalent *prior to* 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 pre-test



## Baseline Equivalence, continued

- According to non-regulatory 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



# Summary of Moderate Evidence

- ESSA requires a well-designed and implemented quasi-experimental study
- ESSA does not define what is well-designed or implemented
- However, generally a well-designed QED has the following
  - At least two groups for comparison
  - Establishment of baseline equivalence



# 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
  - Typically has one group and examines predictors of an outcome
  - Controls are other key variables related to the outcome but not part of the research question
- These types of studies cannot meet WWC standards



## Promising, continued

- 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



# Summary of Promising Evidence

- Only one study group
- Uses terms like relationship, covariate, association, and predictor
- Uses statistical controls



# Demonstrates a Rationale

- 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 the outcome?
- Includes ongoing efforts to collect evidence
  - How will you evaluate the results?



# Debrief Questions

- Do you have any questions about the 4 ESSA levels of evidence?
- What obstacles do you face in finding strong levels of evidence for programs/interventions?
- If an adequate evidence level for a program/intervention does not exist for your population of students, how could your district consider evaluating the program/intervention yourselves?



# Applying ESSA Levels of Evidence to Program Evaluation

- What is the intended/stated goal of the program evaluation?
- What are the effective components of the program?  
[Derived from research-based constructs?]
- What are the relevant outcomes?
- What are the direct (and indirect) effects between the program and intended outcomes?
- Was the program delivered with fidelity?



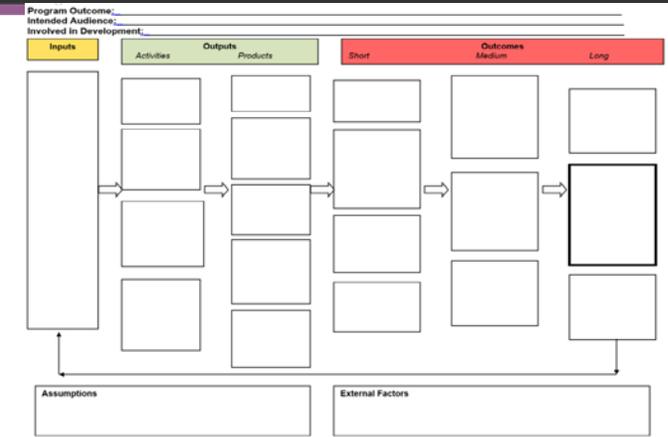
Research  
Article  
Review  
Rubric

Review Criteria	Comments
<b>All Studies</b>	
<b>Literature Review</b> – Review the authors’ theoretical framework, definition of concepts, and the conceptual basis for the study.	
<b>Research Questions</b> – Identify the research questions and their appropriateness to the theoretical framework	
<b>Methodology</b> – What is the population?	
<b>Methodology</b> – What is the sample?	
<b>Methodology</b> – What is the intervention?	
<b>Methodology</b> – Is there a control or comparison group? If there are two or more groups, are they randomly assigned?	
<b>Methodology</b> – Based on the answers above, what is the study design (RCT, QED, Correlational, other)?	
<b>RCTs (Potentially Strong Evidence)</b>	
<b>Methodology</b> – Is there evidence of a confound?	
<b>Methodology</b> – Is there evidence of attrition? Is the overall or differential attrition high?	
<b>QEDs (Potentially Moderate Evidence)</b>	
<b>Methodology</b> – How were the groups established? Was the criterion clear and consistent?	
<b>Methodology</b> – Do the authors examine baseline equivalence?	
<b>Correlational Studies (Potentially Promising Evidence)</b>	
<b>Methodology</b> – What is the predictor or independent variable? Is there more than one?	
<b>Methodology</b> – What are the statistical controls? Are they appropriate?	
<b>Other Study Designs (Demonstrates a Rationale)</b>	
<b>Methodology</b> – Is there a logic model? If not, is the theoretical framework clear?	
<b>All Studies</b>	
<b>Results</b> – Are the results conclusive, and appropriate given the study design?	
<b>Discussion</b> – Are the conclusions appropriate to the results and are limitations appropriately discussed by the author?	



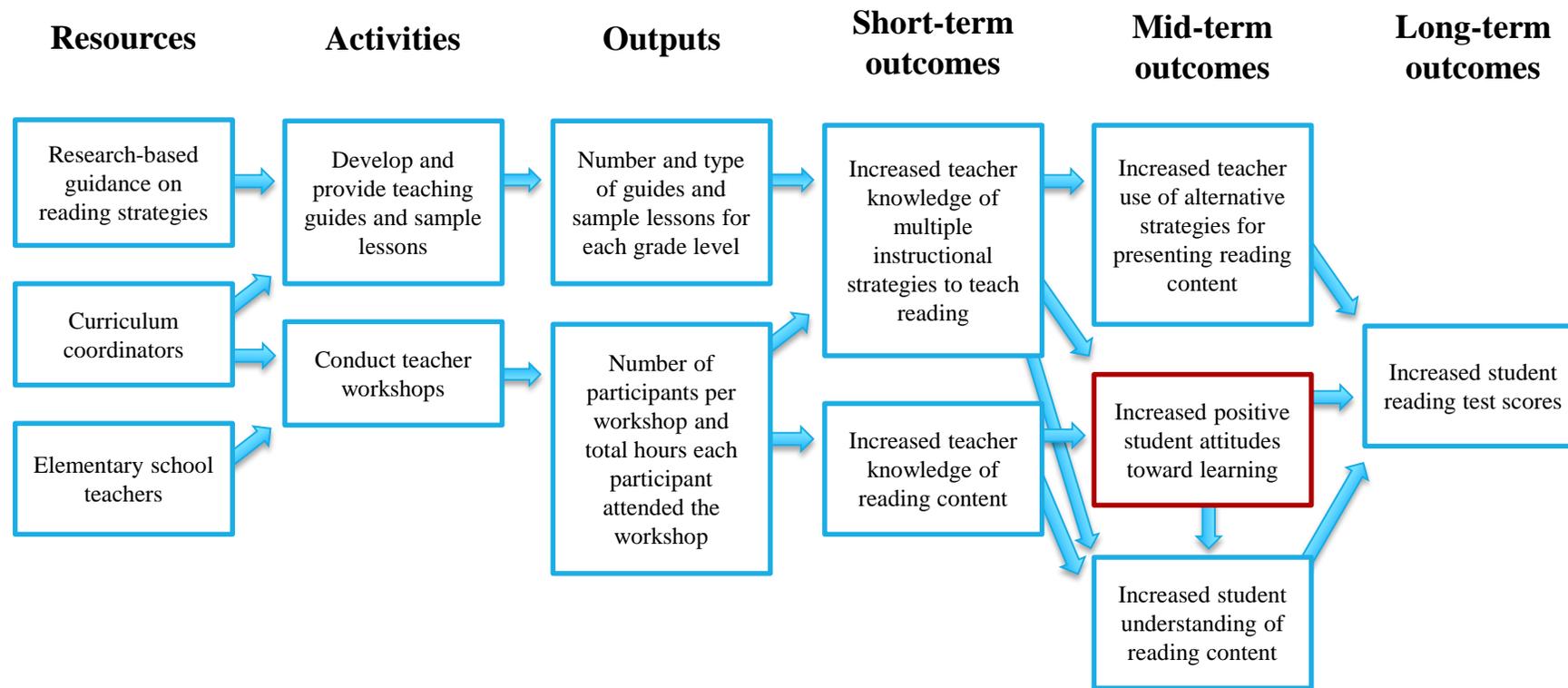
# Using a Logic Model for Planning Program Evaluation

Logic models are visual representations of the theory of action underlying educational programs and interventions



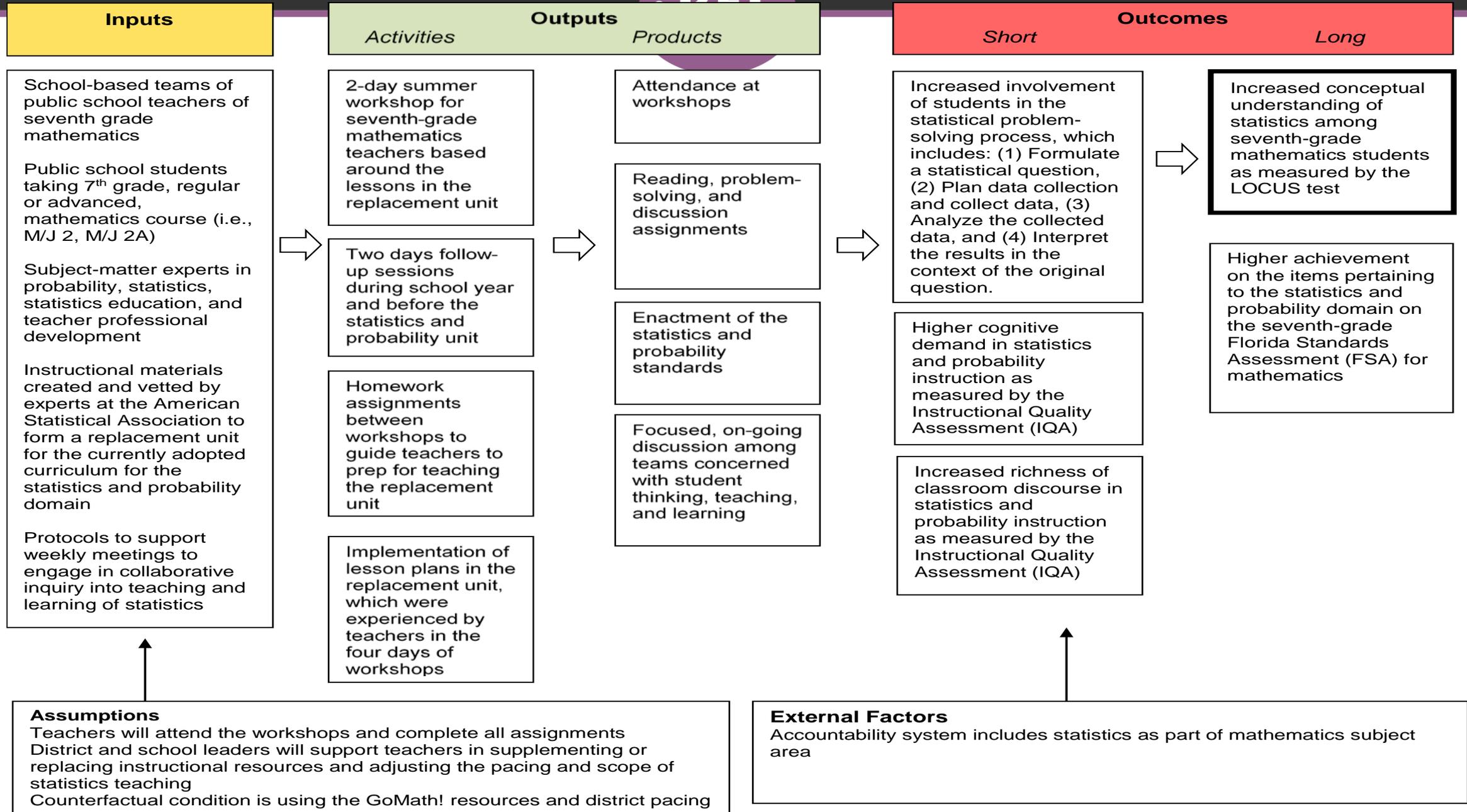
A logic model exercise completed by the planning team can:

- Lead to consensus on final program outcome(s)
- Determine inputs (resources) to consider in the process
- Determine outputs (activities to complete expected products)
- Determine short, medium, and long-term outcomes
  - In many evaluation projects, these are typically teacher & student outcomes
- Consider assumptions & external factors that may impact planning & delivery of the program



# Sample Logic Model

Source: <http://files.eric.ed.gov/fulltext/ED544779.pdf>





## Ex #2: Impact of Word Knowledge Instruction on Literacy Outcomes in Grade 5 (Foorman et al.)

- Addresses Hillsborough County Public School's (HCPS) and the nation's growing achievement gaps due to **English language proficiency** and socio-economic status through a rigorous RCT
- Measures the impact of Word Knowledge Instruction on:
  - awareness of the meanings of prefixes & suffixes;
  - vocabulary and syntactic knowledge, and
  - reading comprehension

# The Impact of Word Knowledge Instruction on Literacy Outcomes in Grade 5

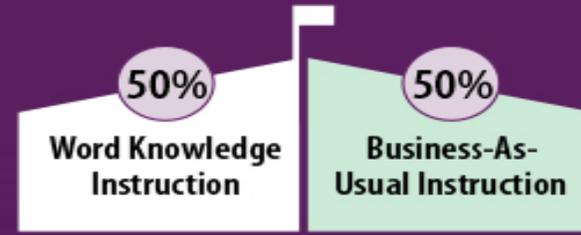
This study will measure the impact of Word Knowledge Instruction on:

- awareness of the meanings of prefixes and suffixes,
- vocabulary and syntactic knowledge, and
- reading comprehension.

## Requirements to Participate

- Approximately 44 schools
- At least 60% of students eligible for free or reduced price lunch
- At least 2 English Language Arts teachers of record in grade 5
- Schools with English learner students will be prioritized

## Random Assignment Within Schools



## Student Benefits

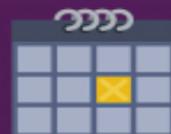
Standards-based instruction that:

- Increases awareness of the meanings of prefixes and suffixes
- Enhances academic vocabulary knowledge
- Enables inferencing of new word meanings
- Facilitates comprehension of text

## Teacher/School Benefits

- \$500 to each school to support grade 5 instruction
- Free PD for teachers randomly assigned to the Word Knowledge condition and an honorarium for summer training
- If Word Knowledge Instruction proves effective, instructional lessons and PD materials will be provided free to HCPS

## Time Table



**20 weeks** of instruction

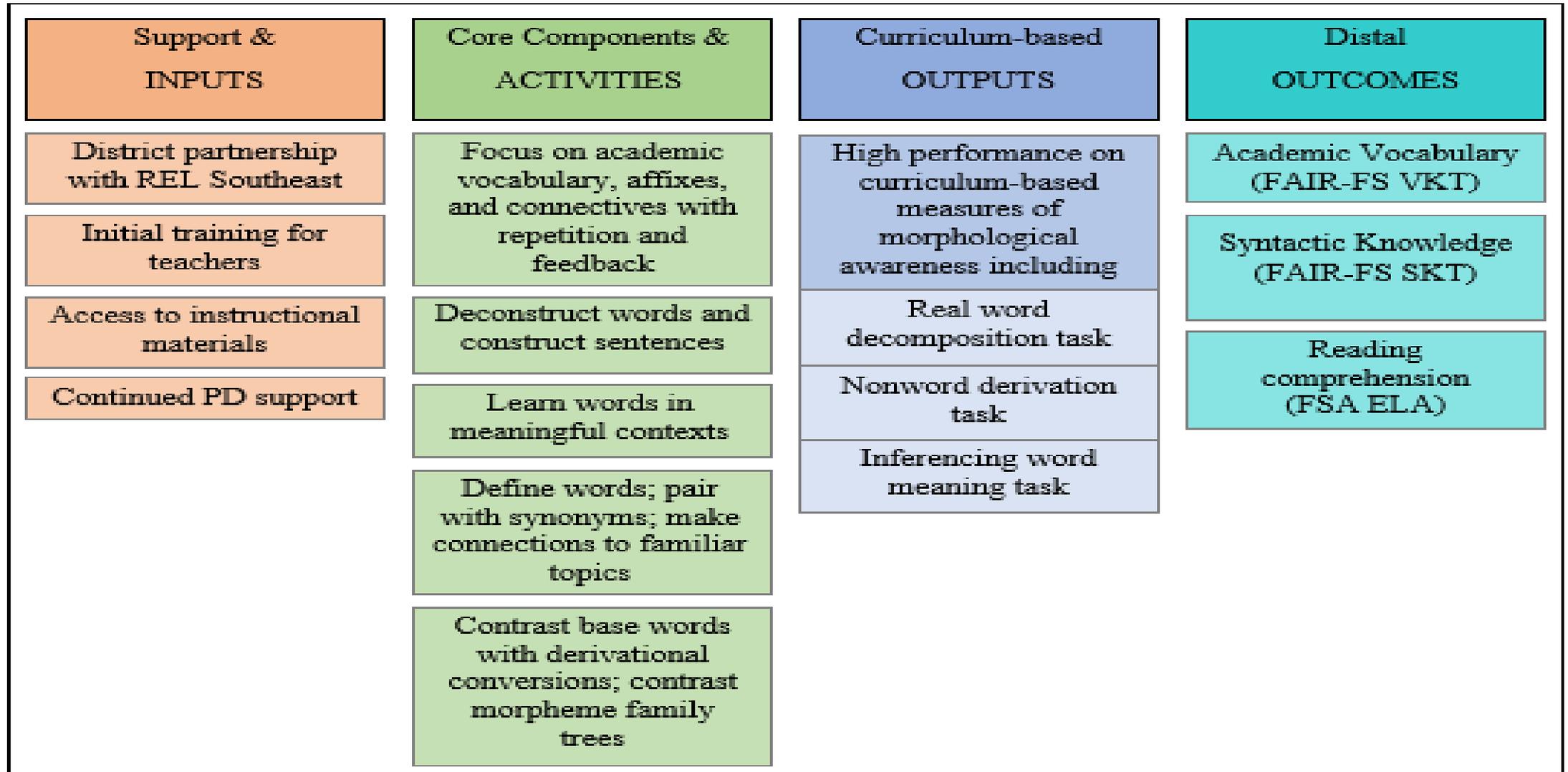


**15-minute** daily Word Knowledge Instruction in place of current district word study

**2018/19 school year:** September 10, 2018 through mid-March, 2019



# Word Knowledge Instruction Logic Model





# PICO Framework: Useful in Evaluating Evidence

<b>P</b>	<b>Population</b>
<b>I</b>	<b>Intervention</b>
<b>C</b>	<b>Comparison</b>
<b>O</b>	<b>Outcome</b>

<https://www.ncbi.nlm.nih.gov/pubmedhealth/PMHT0029906/>



# Debrief Questions

- What is a program/intervention/set of strategies you would like to evaluate?
- What are your research questions?
- What is your theory of action (logic model)?
- How will you determine fidelity of implementation?
- What are the obstacles to your evaluation? Solutions?



# Questions?

Barbara Foorman

[bfoorman@fcrr.org](mailto:bfoorman@fcrr.org)

Phyllis Underwood

[punderwood@fcrr.org](mailto:punderwood@fcrr.org)

For more information about the REL Southeast, please visit our website at [rel-se.fsu.edu](http://rel-se.fsu.edu)