

An SEA Guide for Identifying Evidence-Based Interventions for School Improvement

Laurie Lee
John Hughes
Kevin Smith
Barbara Foorman

FLORIDA CENTER FOR READING RESEARCH, FLORIDA STATE UNIVERSITY



FLORIDA STATE UNIVERSITY
FLORIDA CENTER *for* READING RESEARCH

An SEA Guide for Identifying Evidence-Based Interventions for School Improvement

**Laurie Lee
John Hughes
Kevin Smith
Barbara Foorman**

FLORIDA CENTER FOR READING RESEARCH, FLORIDA STATE UNIVERSITY

UPDATED NOVEMBER 29, 2016

Acknowledgments

Funding Foundations:

William T. Grant Foundation

Overdeck Family Foundation

Annie E. Casey Foundation

Thanks to:

Florida Department of Education:

Sandra Dilger
Holly Edenfield
Jason Graham
Shannon Houston
Eileen McDaniel
Sonya Morris
Melissa Ramsey
Michael Stowell

South Carolina Department of Education:

Jennifer Morrison
Roy Stehle

Mississippi Department of Education:

Kim Benton
Robin Lemonis
Nathan Oakley
Sonja Robertson
Tennette Smith

Table of Contents

Introduction	1
Purpose of the Self-Study Guide.....	1
Flexibility with Responsibility – The Roles of the SEA and LEA.....	1
The Self-Study Process	3
Context for Use of the Self-Study Guide	6
SEA Self-Study Guide Tools	8
SEA Self-Study Guide Checklist.....	8
SEA Facilitator’s Checklist.....	8
SEA Team Member’s Checklist.....	8
SEA Scoring Template.....	8
SEA Scoring Guide.....	9
SEA Voting and Consensus Rating Form.....	10
SEA Planning Form.....	10
Appendix A. Annotated Bibliography.....	10
Appendix B. Theory of Action and Sample Logic Model.....	10
Preparing for Self-Study	11
Collecting Research.....	12
Evaluating Research.....	13
Follow up, Monitoring, and Evaluation	19
SEA Self-Study Guide Checklist	T-1
SEA Facilitator’s Checklist	T-10
SEA Team Member’s Checklist	T-15
SEA Scoring Template	T-19
SEA Scoring Guide	T-20
Area 1: Implementing Systemic Change.....	T-20
Area 2: Establishing Strong Leadership.....	T-26
Area 3: Improving Academic Instruction.....	T-32
Area 4: Developing and Retaining a High-Quality Staff.....	T-38
Area 5: Creating a Positive School Climate and Culture.....	T-43
SEA Voting and Consensus Rating Form	T-47
SEA Planning Form	T-50
Appendix A. Annotated Bibliography	A-1
Scoring Guide Area 1: Implementing Systemic Change.....	A-1
Scoring Guide Area 2: Establishing Strong Leadership.....	A-3
Scoring Guide Area 3: Improving Academic Instruction.....	A-5
Scoring Guide Area 4: Developing and Retaining a High-Quality Staff.....	A-7
Scoring Guide Area 5: Creating a Positive School Climate and Culture.....	A-10
Appendix B. Theory of Action and Sample Logic Model	B-1
References	Ref-1

Introduction

Purpose of the Self-Study Guide

The Every Student Succeeds Act (ESSA) challenges state education agencies (SEAs) to improve student outcomes by addressing the student-, teacher-, and school-level factors that drive achievement gains.

This Guide for Identifying Evidence-Based Interventions for School Improvement is intended to help State Education Agencies (SEAs) carefully consider the evidence supporting intervention options that they will require or recommend in their state ESSA plan and funding applications. SEAs could indicate in their state ESSA plan how they have used or will use the self-study process to identify interventions, in partnership with stakeholders. The purpose of the guide is to help SEAs:

1. evaluate the evidence base for interventions as they identify those to be included in the state plan for ESSA as options for schools in need of comprehensive or targeted support,
2. determine the interventions that have strong evidence, and are relevant and appropriate to meeting the needs of the Local Education Agencies (LEAs), and
3. plan to provide resources for LEAs to help them choose the best evidence-based option(s) for schools in need of comprehensive or targeted support to include in school improvement plans.

Flexibility with Responsibility – The Roles of the SEA and LEA

SEAs and LEAs are charged with implementing ESSA, with states being asked to ensure that LEAs are implementing evidence-based strategies, activities, and interventions in schools in need of significant improvement. Throughout this document, evidence-based strategies, activities, and interventions will be collectively referenced as “interventions.”

In the past, school improvement interventions that LEAs were allowed to use were very prescriptive, but ESSA provides states with the flexibility to delineate interventions, or help LEAs select interventions, provided they are evidence-based. This flexibility provides LEAs with an opportunity to help schools develop improvement plans that may prove to be more effective in increasing student achievement.

ESSA requires that SEAs identify schools in need of comprehensive support and targeted support as delineated in this chart provided by the U.S. Department of Education:

Category: Comprehensive Support and Improvement

Types of Schools	Description	Timeline for Identification	Initial year of identification
Lowest-Performing	Lowest-performing five percent of schools in the State participating in Title I.	At least once every three years	2018-2019
Low High School Graduation Rate	Any public high school in the State with a four-year adjusted cohort graduation rate at or below 67 percent, or below a higher percentage selected by the State, over no more than three years.	At least once every three years	2018-2019
Chronically Low-Performing Subgroup	Any school participating in Title I that was identified for targeted support and improvement because it had a subgroup of students performing at or below the performance of all students in the lowest-performing schools and did not improve after implementing a targeted support and improvement plan over a State-determined number of years.	At least once every three years	State-determined

Category: Targeted Support and Improvement

Types of Schools	Description	Timeline for Identification	Initial year of identification
Consistently Underperforming Subgroup	Any school with one or more consistently underperforming subgroups.	Annually	2019-2020
Low-Performing Subgroup	Any school in which one or more subgroups of students is performing at or below the performance of all students in the lowest-performing schools. These schools must receive additional targeted support under the law. If this type of school is a Title I school that does not improve after implementing a targeted support and improvement plan over a State-determined number of years, it becomes a school that has a chronically low-performing subgroup and is identified for comprehensive support and improvement.	At least once every three years	2018-2019

Along with the flexibility of ESSA comes the responsibility for LEAs, and ultimately SEAs, to ensure that evidence-based interventions are selected and implemented so that students attending schools in need of comprehensive or targeted support have the best opportunity to improve achievement. LEAs and schools in need of comprehensive or targeted support will develop school improvement plans which reflect these evidence-based interventions. LEAs will review and approve targeted support plans, and SEAs and LEAs will review and approve comprehensive support plans. **LEAs must conduct a needs assessment for schools identified in need of comprehensive support. Potential interventions should be evaluated on the basis of school needs and the evidence-based interventions selected for implementation should meet the needs of the school.**

The Self-Study Process

Self-study is a process that facilitates thoughtful investigation and discussion of an issue or topic so that decisions can be made through the collaboration of a variety of stakeholders. Although a time investment is required to prepare for discussions that focus on the topic or issue, engage in the discussions themselves, and subsequently plan for implementation of decisions made by the self-study team, the results of this collaboration can be invaluable.

In order to engage in the self-study process, a team must first be established. The self-study team may include school improvement specialists, content area specialists, exceptional student education (ESE) and English learner (EL) specialists, as well as those involved in professional development and leadership at the SEA knowledgeable in school improvement. The SEA may also choose to include representatives from LEAs such as district leaders, teachers and principals to help increase relevance and buy-in. A facilitator will then be selected to organize the work and may be chosen by SEA leadership or the team itself. Since the goal of this specific self-study is to decide upon interventions that might be recommended for schools needing improvement, the facilitator should have deep content knowledge of school improvement, be well-organized, a good listener, and be able to lead a discussion that encourages participation from all team members. SEAs may wish to consider using an external facilitator such as an individual from a university, Regional Education Laboratory, or Comprehensive Center.

The self-study process will help SEAs identify the strongest evidence-based interventions that the SEA will require or recommend for inclusion in LEA school improvement plans or LEA funding applications. SEAs could indicate in their state plan how they have used or will use this self-study to identify interventions, in partnership with stakeholders.

Prior to engaging in this process, the SEA should conduct a needs assessment to clearly identify the problems to be addressed (see Figure 2 below). Undoubtedly, schools in need of improvement throughout the state will exhibit a variety of issues and problems. It is important that the recommended interventions not only have a strong evidence-base, but that they address the issues that schools are facing. Figure 1 outlines the general steps in conducting the self-study.

Figure 1. The Self-Study Process: Conducting the Self-Study



Step 1: Preparation

Present Overview & Review Guide	Facilitator explains process to team	Team reviews guide and asks questions before proceeding to ratings
Collecting and Evaluating Research	Team members identify an evidence-based intervention and complete <i>SEA Scoring Template</i>	Facilitator distributes completed <i>SEA Scoring Templates</i> to team
Individual Rating	Team reviews relevant data and sources of evidence to help determine ratings	Team independently rates interventions submitted by team members and those provided in the <i>SEA Scoring Guide</i>



Step 2: Discussion

Consensus Rating	Facilitator guides the consensus rating process	Record recommendation of intervention as agreed upon by the team
-------------------------	---	--



Step 3: Planning

Documenting Next Steps	Team identifies 2-3 areas where support and resources for LEAs should be developed	Complete a detailed plan for next steps based on urgency, feasibility
-------------------------------	--	---

Step 1 is preparation. During this step the facilitator will describe the process to the team and ensure that everyone has the same understanding of the work. Each team member will review the sections of the self-study guide addressing the collection and evaluation of research and the ESSA Levels of Evidence, identify one or more potential interventions, and evaluate the level of evidence for them. These interventions may fall into the areas that have been identified in the *SEA Scoring Guide* (described in the upcoming Self-Study Guide Tools section), or they may fall into an entirely different category altogether. This is a critical activity since this guide is unable to address all of the potential interventions a state might consider, and more ideas for consideration will improve the results of the discussion step. In addition, the team members will complete the *SEA Scoring Guide*, considering the strategies and interventions provided, and reflecting upon whether or not they should be recommended for use in LEAs and schools.

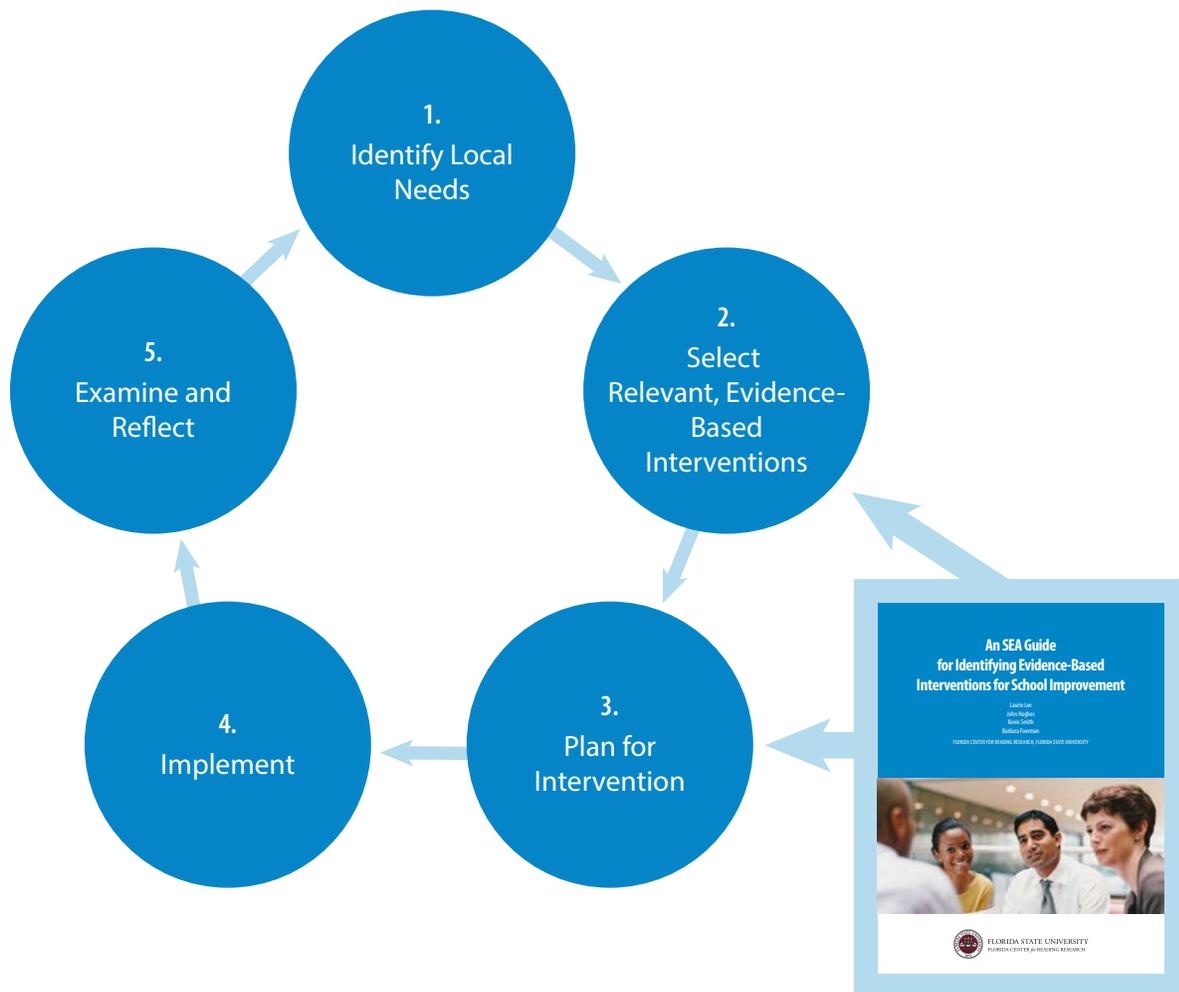
During Step 2 team members discuss all of the various ideas for interventions that the SEA might permit or recommend, and the individual ratings that team members assigned on the *SEA Scoring Template* (described in the upcoming Self-Study Guide Tools section) and the *SEA Scoring Guide*. It is during this step that the SEA will settle on the options that LEAs will be authorized to use if the state is providing a list of interventions from which LEAs must choose. Having a broad range of strategies and interventions is important, but it is equally important that they be based on the best available evidence. In addition, it is critical that strategies and interventions meet the needs that have been identified in the state.

During the final step, the SEA team members discuss priorities, potential resource development, and anticipated challenges in implementation of the strategies. Next steps may be determined with a timeline established and team members assigned to tasks. The facilitator leads the discussion and information is recorded on the *SEA Planning Form* (described in the upcoming Self-Study Guide Tools section).

Context for Use of the Self-Study Guide

Guidance released by the U.S. Department of Education on September 16, 2016 and available at <http://www2.ed.gov/policy/elsec/leg/essa/guidanceusesinvestment.pdf> provides a series of steps that can promote continuous improvement and support better outcomes for students. These steps include:

Figure 2. Steps to Promote Continuous Improvement



Source: Adapted from U.S. Department of Education, 2016.

The use of this self-study guide will be most helpful in addressing steps two and three above.

SEAs should select, or help LEAs select, evidence-based interventions (step two) that best meet the needs identified in the school-level needs assessment and that address the root causes of underperformance. While the level of evidence should be as strong as possible, it is just as important that the interventions meet the needs identified in step one. In addition, the guidance encourages SEAs and LEAs to look at the overall body of relevant evidence rather than just one study when selecting interventions. Moreover, the evidence base should reflect a preponderance of statistically significant, positive effects rather than statistically significant, negative effects. Finally, in cases of minimal evidence, the role of strong theory and logic is paramount.

The guiding questions included in the self-study guide may help team participants consider whether an intervention may meet the needs of schools in the state and begin planning for implementation (step three). The questions may also provoke thinking about resources available as well as technical assistance and support that SEAs may need to offer to LEAs for successful implementation.

SEA Self-Study Guide Tools

The SEA Guide for Identifying Evidence-Based Interventions for School Improvement consists of the following nine tools: *SEA Self-Study Guide Checklist*, *SEA Facilitator's Checklist*, *SEA Team Member's Checklist*, *SEA Scoring Template*, *SEA Scoring Guide*, *SEA School Voting and Consensus Rating Form*, *SEA Planning Form*, *Appendix A*, and *Appendix B*. These are described below.

SEA Self-Study Guide Checklist

This checklist delineates in chronological order the steps of the self-study process for facilitators and team members. The tool assists those involved in the self-study in ensuring that all tasks are completed.

SEA Facilitator's Checklist

While the *SEA Self-Study Guide Checklist* delineates tasks of everyone involved in the self-study process, this checklist reflects only the responsibilities of the **facilitator** throughout preparation, discussion, and planning for next steps. This tool assists facilitators in ensuring that all tasks are completed.

SEA Team Member's Checklist

While the *SEA Self-Study Guide Checklist* delineates the tasks of everyone involved in the self-study process, this checklist reflects only the responsibilities of **each team member** throughout preparation, discussion, and planning for next steps. This tool assists team members in ensuring that all tasks are completed.

SEA Scoring Template

This blank template provides an opportunity for each team member to identify one or more interventions that are appropriate and relevant to the needs of the schools in the state, determine the strength of the associated evidence base and the fit and feasibility of the intervention, and record this information prior to the start of the self-study process. The form includes fields to enter the following information:

- a broad overall area to which the intervention pertains that could be an area identified in the *SEA Scoring Guide*, or another area altogether,
- the specific intervention identified by research to be considered,
- the evidence level based on a body of collected research,
- a summary of the collection of research reviewed which may include the results and significance of the studies, and
- additional information identified locally that pertains to the needs that schools will want to consider such as school improvement plans or student achievement data.
- Guiding questions will facilitate a discussion among team members. Guiding questions may include any number of factors. Some common ones to consider include:
 - the level of satisfaction among the group with the evidence-level of the intervention,
 - the extent to which the intervention was conducted on a student population that is relevant to the state or district context,

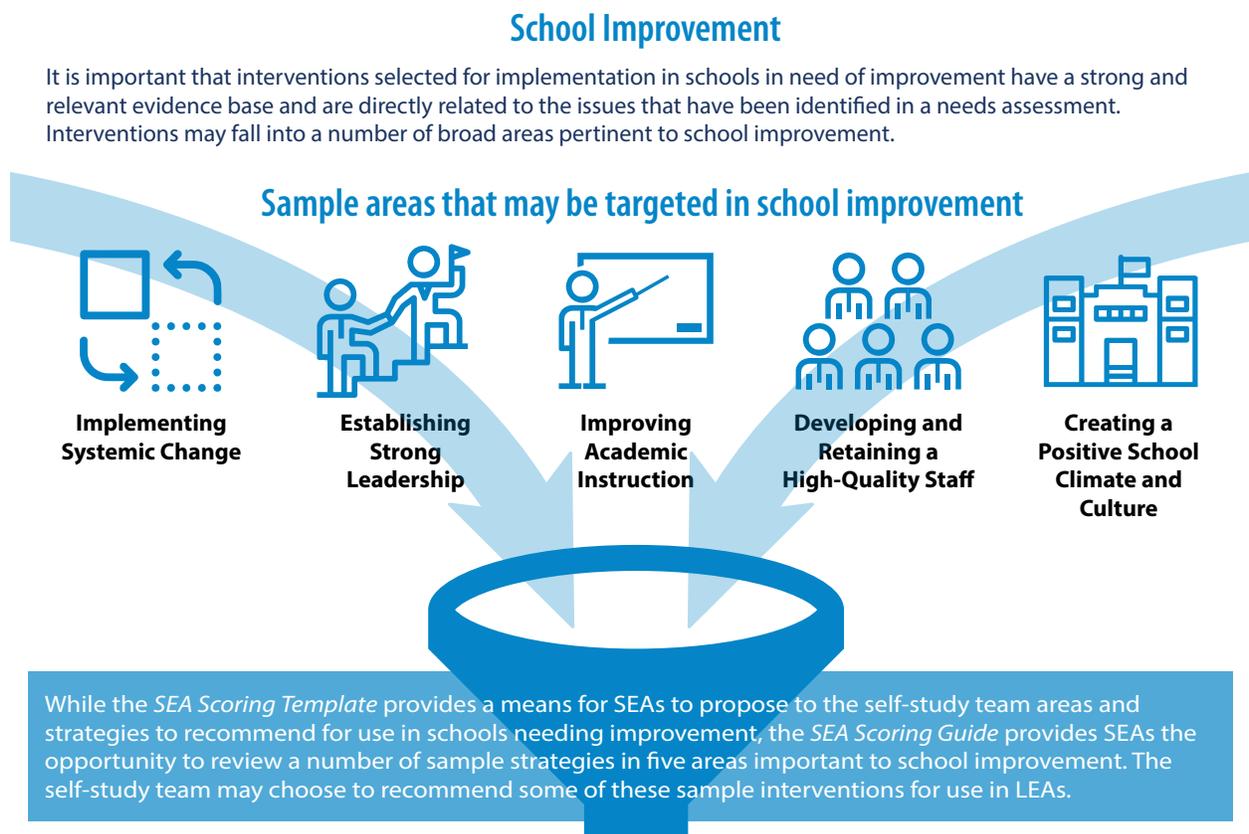
- the types of schools where the intervention might work best, and
- the possible cost/benefit of implementation.

A rating scale is also included in the template so that, after careful consideration, self-study team members can determine whether they (1) do not recommend, (2) recommend, or (3) strongly recommend an intervention. The *SEA Scoring Guide* (described below) may be used as an example for completing the *SEA Scoring Template*.

SEA Scoring Guide

The *SEA Scoring Guide* includes already identified examples of evidence-based strategies and interventions, along with a summary of the research base, the ESSA evidence-base level, state-level information that may be helpful to consider, and guiding questions for discussion. The content of the *SEA Scoring Guide* is organized into five areas: implementing systemic change; establishing strong leadership; improving academic instruction; developing and retaining a high-quality staff; and creating a positive school climate and culture. The areas chosen for the guide were based on those identified in the Institute of Education Sciences (IES) Practice Guide *Turning Around Chronically Low-Performing Schools*. A literature review was conducted identifying interventions associated with the areas. In addition, literature was also reviewed pertaining to the systemic interventions previously required for use in schools needing improvement. **The *SEA Scoring Guide* is not meant to be an all-inclusive or recommended list of school improvement interventions, but rather contains examples of interventions that might meet the needs of schools needing comprehensive or targeted support.**

Figure 3. Areas Associated with School Improvement



As the facilitator and self-study team members review the information in the scoring guide, work through the rating system individually, and then engage in discussion, they thoughtfully consider whether or not to recommend an intervention for their state. The interventions recommended may become a menu from which LEAs may choose based on the needs of the school. It may be that an evidence-level is strong for an intervention, but the state has not experienced much success using that specific approach. Also, team members should strongly consider what has already been done in the state, and the effectiveness of current strategies and interventions. It may be that an evidence-level may be strong for an intervention but the state has not experienced much success in using that specific approach. Perhaps some interventions should replace others based on that experience. An annotated bibliography of the research supporting each scoring guide area is provided in *Appendix A*.

SEA Voting and Consensus Rating Form

After the *SEA Scoring Guide* is completed, the facilitator guides the self-study team through a consensus rating process. The team uses the *SEA Voting and Consensus Rating Form* to reach agreement on whether the proposed intervention should be recommended as an option for schools requiring comprehensive or targeted support in the ESSA state plan. The most important part of this process is the discussion that goes into consensus rating. The scores on the *SEA Voting and Consensus Rating Form* should reflect this facilitated discussion.

SEA Planning Form

This form is used to establish priorities, ideas regarding resource development for LEAs, and any anticipated challenges. The facilitator leads the discussion centered on these topics and uses the form to record ideas.

Appendix A. Annotated Bibliography

This appendix describes key references that provide additional support for each of the scoring guide areas. Research from each study referenced in the LEA and School Scoring Guide is summarized.

Appendix B. Theory of Action and Sample Logic Model

This appendix provides information pertaining to theory of action and also includes a sample logic model to help familiarize participants with these concepts.

Preparing for Self-Study

In preparation for the self-study process, leadership at SEAs recruit team members to participate. Leadership at SEAs recruit members for the self-study team. The team should be comprised of a wide range of individuals so as to include as much knowledge and as many skills as possible. Members typically include researchers, content area specialists, exceptional student education (ESE) and English learner (EL) specialists, those involved with professional development, and senior leadership at the SEA. The SEA may also wish to include representatives from LEAs such as district leaders, teachers, and principals. The names of team members and facilitator may be recorded on the *SEA Voting and Consensus Rating Form*.

Leadership at the SEA or the team members select a dedicated and knowledgeable facilitator such as the school improvement director or ESSA state plan project manager. The facilitator should have deep content knowledge of school improvement, be well-organized, a good listener, and be able to lead a discussion that encourages participation from all team members.

Once the team is established, the following steps should be followed:

1. The facilitator studies the materials provided to conduct the self-study process so that he/she can effectively guide team members through the process. The facilitator gathers all pertinent data and evidence pertaining to the interventions.
2. The facilitator distributes a blank *SEA Scoring Template*, the *SEA Scoring Guide, Appendix A, Appendix B*, as well as any other relevant data or evidence to each team member, and provides a timeline for team members to review materials.
3. The facilitator schedules a short meeting after team members have reviewed the documents to discuss any questions.
4. The facilitator asks each member to re-read the sections of the self-study guide addressing the collecting and evaluating of research and the ESSA Levels of Evidence. The facilitator then requests that team members research an area pertinent to school improvement in order to identify a specific evidence-based intervention for consideration by the team during the self-study process, and to complete the *SEA Scoring Template*. Research areas could include those addressed in this guide: implementing systemic change, establishing strong leadership, improving academic instruction, developing and retaining high-quality staff, and creating a positive school climate and culture. Alternatively, research could include other areas selected by the team member or SEA. The team can work individually or with a partner or small group to identify a broad area and then a more specific intervention to investigate. The team may collect research on as many interventions as they choose. Team members may wish to share their selected intervention(s) with one another so there is no duplication of effort.
5. The facilitator establishes a deadline for completion and submission of the *SEA Scoring Templates* and communicates that to the team.
6. Each team member re-reads the sections of the self-study guide addressing the collecting and evaluating of research and the ESSA Levels of Evidence, reviews research, completes the *SEA Scoring Template* using the *SEA Scoring Guide* as an example, and returns the completed template to the facilitator by the established deadline.
7. The facilitator distributes the completed templates to all team members and instructs members to rate these strategies and interventions according to the scale on the template and to complete the *SEA Scoring Guide*.

8. The facilitator informs team members of the timeline for their review and schedules a consensus rating process meeting.
9. Team members review the completed *SEA Scoring Templates* they received from the facilitator. They may also rate the interventions in the *SEA Scoring Guide* to individually reflect their thoughts regarding the recommendation of any interventions after reviewing the summary of research and any data or evidence provided by the facilitator. A team member who does not know how to rate a specific area may abstain from rating it.

Collecting Research

To collect the research necessary to identify a range of evidence-based interventions, team members should search professional educational journals and websites of reputable organizations. Some databases and websites to consider include:

What Works Clearinghouse: <http://ies.ed.gov/ncee/wwc/>

ERIC: <http://www.eric.ed.gov/>

JSTOR: <http://www.jstor.org/action/showAdvancedSearch>

Google Scholar: www.google.com/scholar

Institute of Education Sciences (IES) Resources: <http://ies.ed.gov>

Blueprints for Healthy Youth Development Database: <http://www.blueprintsprograms.com/>

Results First Clearinghouse: <http://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2014/09/results-first-clearinghouse-database>

The search process begins by identifying relevant keywords. The search should not focus on just a few search terms, such as “school turnaround” but should be broad so as to capture as many relevant studies as possible. Examples of keywords include:

School turnaround	Focus school	Reading intervention
School improvement	Effective schools	Professional development
Low-performing schools	Randomized control trial	Mathematics intervention

Keywords can be combined to look for specific ideas, such as ‘best practices’ and ‘professional development’ and ‘principals’ to find ways to better train school leaders.

In addition to searching for individual articles and studies, SEAs may find resources that combine multiple studies in a specific area helpful. The What Works Clearinghouse (WWC) practice guides, for example, synthesize a large number of studies and identify those with the most supporting evidence. Similarly, organizations like RAND have pulled together multiple studies to provide summaries of what the research has found.¹

¹ http://www.rand.org/pubs/research_reports/RR1550.html

Evaluating Research

One of the most challenging steps for many SEAs will be evaluating the research that they collect to match it to the appropriate levels of evidence. This section provides some general guidance on how to determine the level of evidence for a study; however, a number of resources exist that can help SEAs with this task. One is the What Works Clearinghouse², sponsored by the Institute for Education Sciences. The WWC rates research studies according to a set of standards³ and provides information about the rigor of those studies. Because the guidance around ESSA levels of evidence refer to and utilize WWC standards, those standards are referenced throughout this section and readers should familiarize themselves with them. Another resource is the Best Evidence Encyclopedia housed at Johns Hopkins University.⁴

Additionally, there are a number of organizations that SEAs can reach out to for support in evaluating research. Federally funded organizations such as the Regional Comprehensive Centers and content centers⁵ and the Regional Educational Laboratories⁶ are well-suited to provide states with this kind of support. SEAs can partner with universities that have centers and individual faculty with expertise in these topics. The National Network of Education Research–Practice Partnerships can provide support to SEAs that want to explore these kinds of research–practice partnerships.⁷

What are the ESSA levels of evidence?

ESSA recognizes four levels of evidence. This section is designed to help SEA and LEA staff understand these different levels and apply them to research they are considering for school turn-around and related purposes. A summary of the four levels of evidence is shown in Figure 4:

Figure 4. ESSA Levels of Evidence

Category One		
Demonstrates a statistically significant effect on improving student outcomes or other relevant outcomes	1 Strong Evidence	based on at least 1 well -designed and well-implemented experimental study
	2 Moderate Evidence	based on at least 1 well-designed and well-implemented quasi-experimental study
	3 Promising Evidence	based on at least 1 well-designed and well-implemented correlational study with statistical controls for selection bias
Category Two		
Demonstrates a rationale based on high-quality research findings or positive evaluation that such activity, strategy, or intervention is likely to improve student outcomes or other relevant outcomes	4 Demonstrates a Rationale	includes ongoing efforts to examine the effects of such activity, strategy, or intervention

Source: Source: Adapted from *Chiefs for Change*, 2016.

2 <http://ies.ed.gov/ncee/wwc/default.aspx>

3 <http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19>

4 <http://www.bestevidence.org/>

5 <http://www2.ed.gov/about/contacts/gen/othersites/compcenters.html>

6 <http://ies.ed.gov/ncee/edlabs/>

7 <http://nnerpp.rice.edu/>

For each of the first three levels, the research studies must demonstrate a “statistically significant effect on improving student outcomes or other relevant outcomes.” Statistically significant means that the difference observed in the study is not likely due to chance. Implied by this requirement is that the results are positive and not overridden by statistically significant negative results from other studies with moderate or strong levels of evidence. In many cases, multiple studies of the same intervention will yield different results and it is possible that some could be positive and others negative while all still being statistically significant.

A result can be statistically significant but not substantively important. That is, a positive effect can be statistically significant but the effect may be so small as to be unimportant in practical terms. The impact is often described as an *effect size*, which is the magnitude of the difference between intervention groups measured as the proportion of a standard deviation. For example, an effect size of 0.25 means that an average student in one intervention group would be expected to have scored 0.25 standard deviation more had they participated in the other intervention group. The WWC considers an effect size of greater than or equal to 0.25 to be a substantively important difference. While not specifically required under ESSA, it is strongly recommended that when reviewing research the effect size should be considered along with the statistical significance.⁸

In addition, the first three evidence levels each expect that the studies have large and multi-site samples and that the samples reflect populations or settings similar to those proposed to receive the intervention. These are critical considerations. A well-designed study with strong evidence for an intervention for early grade students may not be suitable for adolescents. Similarly, an intervention from a study conducted in an urban school may not be appropriate for a rural school. Ensuring that the sample was large, from multiple sites, and similar to the target population will increase the chances of success.

Finally, the fourth level, demonstrates a rationale, can be thought of as an evidence-building opportunity. That is, evaluation of an intervention with minimal evidence but strong supporting logic for its potential to improve outcomes is an opportunity to begin developing evidence of its effectiveness.

What is strong evidence?

Strong evidence is defined as “a well-designed and well-implemented experimental study.” The Department of Education considers a study to be well-designed and implemented if it meets WWC standards without reservations. One of the first steps in reviewing any research is to check the WWC to see if a study has been rated.

But if a study has not been reviewed by the WWC, it is still possible to determine the appropriate level of evidence. For strong evidence this will require some form of an experiment or a regression discontinuity design.

⁸ Throughout this report a number of terms are used, such as statistically significant, substantively important and intervention. A good resource that defines many of these terms can be found at the What Works Clearinghouse which provides an online glossary at: <http://ies.ed.gov/ncee/wwc/Glossary.aspx>.

The essential components of an experiment in educational research include:

- some kind of intervention or treatment designed to change outcomes,
- subjects who receive the intervention (typically called an experimental or treatment group),
- subjects who do not receive the intervention (typically called the control group), and
- random assignment of experimental and control groups.

To qualify as an experiment, there must be some factor that is manipulated. This is called the *treatment* and could be a curriculum, a teaching strategy, a school policy, or anything similar. For example, a district might implement a new math intervention. This would be provided to some students at some schools but not to others. Thus, an educational aspect is changed for some individuals and held constant for others.

The students (or teachers or schools) that receive the intervention or are part of the factor that is manipulated are the *experimental* or *treatment* group (and possibly a comparison group). Those for whom instruction is unchanged are part of the *control* group, often called the “business-as-usual group.”

Note, however, that random assignment is particularly critical. Whenever two different groups receive different treatments, changes in outcomes could be a result of the different treatment but also because of differences in the groups. For example, if a school wanted to test a new reading program it might decide to give some classrooms the new program but other classrooms use the original reading program. This creates two groups to compare but if the students in the classes are different (maybe one group is more advanced than the other), differences in outcomes might be due to differences in the students and not the new program. The best way to overcome this risk is to randomly assign students (or teachers or schools) to either the treatment or control group. True random assignment helps ensure that the two groups are likely to be similar to each other and that any differences in outcomes are due to the treatment and not to differences between the subjects in the two groups.

Whether or not an experiment is well-designed and well-executed is not simple to determine. There are numerous factors that could weaken confidence in an experiment’s results, more than can be described here. Readers should look at resources such as the What Works Clearinghouse, which has developed standards to help judge the level of rigor for many educational studies.

For this guide there are two critical limitations to focus on that can help identify studies that were not well designed or well executed. The first limitation is *attrition*. Attrition is the loss of subjects from the experiment. Even if the subjects are randomly assigned at the beginning, if enough members of either group leave the experiment, it can effectively undo the randomization process. The individuals who leave are likely to differ from those who stay, and, thus, if enough leave the results could be biased. The WWC provides guidance on appropriate levels of attrition.⁹

The second limitation is any kind of *confound*. A confound occurs when some aspect of the experiment is completely aligned with one aspect of the study conditions, even if all subjects were randomly assigned. A confound can be thought of as an “extra” factor that was not taken into account that could explain the observed differences between the two groups. The most common confound occurs when there is only one unit (that is, teacher, classroom, school, or district) assigned to each group. For example, consider two classrooms taught by different teachers. One classroom comprises the intervention group and the other comprises the control group. The teachers could be randomly assigned to the treatment or control conditions but there would still be a confound because there was only one teacher in each condition. If the study found that the intervention classroom performed better than the control classroom, an alternative explanation for the observed difference could be related to differ-

⁹ http://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_brief_attrition_080715.pdf

ences between the classroom teachers and not the intervention. Another example of a confound in an RCT is overalignment of the outcome measure and the intervention. If the outcome measure is a direct measure of the intervention, then the results are confounded. An intervention that teaches specific spelling words and then measures the results with a test of those same words would be overaligned. Inclusion of a norm-referenced spelling test would be necessary to prove the intervention's effectiveness beyond a taught spelling list.

Like an experimental design, a regression discontinuity design (RDD) can meet WWC standards without reservations and can be considered strong evidence. An RDD determines causal impacts by examining interventions that occur just above and below a cut-off of some kind. In these cases, the cut-off, such as a cut-score on a test, splits the population of interest into two groups that can be compared. The logic is that subjects just above and just below the cut-off are likely very similar and so can be compared. An RDD study must meet several requirements to qualify as strong evidence, including establishing the equivalence between the two groups and avoiding confounds. For more information on how an RDD can meet WWC standards without reservations, please see the WWC reference resources.¹⁰

Summary of key things to look for:

- meeting WWC standards without reservations,
- experimental or treatment group (and the possible addition of a comparison group),
- control group that does not receive the treatment,
- groups formed by random assignment or a discontinuity such as a cut-score,
- low attrition, and
- the absence of a confound.

What is moderate evidence?

Moderate evidence is based on at least one study using a quasi-experimental design (QED).¹¹ What is the difference between an experiment and a quasi-experiment? The major difference is that a QED lacks random assignment of subjects to groups and instead, a QED leverages some natural change, such as implementation of a new program, to create treatment and control groups. QED studies are common because many educational policies and practices are implemented across the board or with a small pilot group that was not randomly assigned. For example, a few school principals might volunteer their schools to participate in a new initiative. Results from those schools might then be compared to schools that did not volunteer. This creates a *treatment* and a *control* group but lacks random assignment. As noted above, when subjects are not randomly assigned it increases the risk that any observed differences in outcomes are due to other factors. In this example one might wonder if the principals who volunteered were especially excited or interested in the intervention, or perhaps more creative leaders, and that it was their leadership and interest that drove changes in outcomes.

A common QED is to compare changes in the pre-test and post-test scores for students in two different groups. This looks like an experiment except that the two groups were not randomly assigned. The researchers would try to select groups that are similar on key criteria, such as English learner status or economic status, so that the groups can be compared. A related approach is to statistically match students. One way this is done is by taking each student who received an intervention and finding a

¹⁰ <http://ies.ed.gov/ncee/wwc/Document/258>

¹¹ Note that an RDD is a type of quasi-experimental design but it can still meet WWC standards without reservations and thus potentially can qualify as strong evidence.

statistical “twin” who did not receive the intervention and then comparing results.

As with experiments, deciding whether or not a QED is well-designed and well-executed is not simple to determine. Again, readers should look at resources such as the What Works Clearinghouse, which provides information about the level of rigor for many educational studies. A study that meets WWC standards with reservations qualifies as moderate evidence.

Perhaps the single most critical factor to consider in a QED is whether or not the study was able to establish *baseline equivalence* between the two groups. As noted above, experiments use random assignment to try to ensure that the two groups studied are as equal as possible and often include pretest scores as covariates so as to improve analytic precision. Without random assignment, researchers use other ways to ensure that groups are similar, such as comparing them on key variables like race, economic status, and test scores. Verifying that two groups are comparable on pre-test scores is an excellent way to establish baseline equivalence.

Without randomized assignment there will remain a concern about unobservable group differences that weaken confidence in the results. For example, two students with the same pre-test scores could have very different levels of motivation, which could in turn result in one improving more than another. Concerns about unobserved differences are why even a well-executed QED is rated as only having moderate evidence.

Summary of key things to look for:

- experimental or treatment group (and the possible addition of a comparison group),
- control group,
- establishing or failing to establish baseline equivalence, and
- no random assignment.

What is promising evidence?

Promising evidence comes from correlational studies. In a correlational study there is no assignment to treatment and control groups. Instead, a correlational study uses *predictors* or independent variables to look for a *relationship* between some factor and the outcome of interest within a group or groups of subjects. For example, suppose a school enacted a program to encourage students to read more books during the school year by offering prizes. At the end of the year a researcher might see if the number of books read is a good predictor of changes in student test scores. All students would be in the analysis so there is just one study group. The number of books serves as the independent variable or predictor of interest while other factors such as prior test scores might be used as *control variables* or *covariates*. Nonetheless, a positive association between number of books read and increase in student test scores would be difficult to interpret because of the lack of a control group and potential confounds.

The phrase “statistical controls for selection bias” refers to some of these control variables or covariates. Selection bias refers to the possibility that the process of selecting or identifying the study subjects introduces some kind of systematic error that could invalidate the results. A common problem is selective participation in a treatment. For example, as part of a new policy the district assigns reading coaches to specific schools. Because the schools were not randomly assigned, or assigned based on a cut-score, there would not be a good comparison group. Instead, a researcher wanting to understand if the new policy was effective would have to use statistical controls to try to adjust for differences between the schools with coaches and those without. Thus, conclusions from the available data would be limited.

Researchers often try to overcome selection bias by checking that key factors, such as test scores and demographics, are similar between those receiving the treatment and those that did not. Putting these variables into a model allows researchers to statistically control for those factors. To meet the standard of promising evidence, a correlational study must have those kinds of statistical controls. Note that statistical controls may also be used in a QED or even an RCT to add analytic precision and to guard against possible confounds such as variation in district policy implementation.

Correlational studies are considered promising evidence because there is no way to assign causality to the results. Mathematically, all a correlation can demonstrate is that two variables are related to each other. Logic might indicate a causal path, such as reading coaches lead to higher tests scores. But without random assignment there are other competing explanations for the correlation. In this example, reading coaches might lead to improved scores. But it is also plausible that the schools with coaches adopted other changes that led to higher test scores. A correlational analysis can only show an association, it cannot explain a causal relationship. That is why such studies are only rated promising.

Key things to look for to identify a correlational study:

- only one study group (no separate treatment and control groups),
- terms such as “relationship,” “covariate,” and “predictor,” and
- presence of statistical controls.

What qualifies as demonstrates a rationale?

The final level of evidence provides flexibility to work with interventions that have not been studied much or at all. Part of the goal for this flexibility is that allowing schools and districts to test new interventions may add to our knowledge of what works. Note that ESSA limits the use of funds for practices in this category. For example, the 7% of Title I, Part A funds set aside for school improvement efforts must use interventions supported by research in the top three tiers.

For the purposes of this guide, two aspects are notable. First, there should be a theory of change providing a basis for expecting an intervention to result in an improvement. The theory of change should be well-constructed and well-established, such as by using a logic model. Readers are encouraged to develop logic models for these kinds of interventions.¹² An example of a logic model for evaluating the effects of professional development of student reading outcomes is provided In Appendix B.

Second, it is expected that SEAs and LEAs will carefully monitor progress of the selected strategies. Ideally the interventions should be evaluated through well-designed experiments but an LEA or SEA should at least set up an evaluation before applying the intervention. This would require, minimally, identifying the expected outcomes, tracking implementation, collecting follow up data, and conducting the analyses. Implementing an intervention with no way to measure or understand its consequences deprives the larger educational field an opportunity to learn more about the intervention.

Keys to consider:

- What is the logic model explaining the theory of change?
- How will the practice be evaluated?
- How will you know if it worked, or didn't work?

¹² A good resource to help with creating logic models can be found at <http://ies.ed.gov/pubsearch/pubsinfo.asp?pubid=REL2015057>.

Follow up, Monitoring, and Evaluation

Inherent to implementing evidence-based interventions to improve school performance is frequent monitoring of progress. Leaders at all levels need to address the crucially important questions of: “Are we on track?” and, if not, “Why not, and what are we going to do about it?”

Most SEAs will use their existing accountability systems for monitoring and evaluation; however, there are ways to enhance these systems. SEAs may want to consider the following questions:

How do SEAs and LEAs identify indicators appropriate to the interventions selected?

How large of an improvement and how quickly should SEAs and LEAs see a difference in the indicators?

What steps might SEAs and LEAs take to increase the intensity of the new interventions to accelerate student growth?

At what point do the indicators suggest that interventions be changed?

Answers to these questions entail that leaders build an infrastructure at the school level that addresses improvement in instruction, leadership, teaching, and professional development, with the necessary resources and accountability to be successful.