 Contact: cbcse@tc.columbia.edu

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

Assessing Relevance and Credibility of Available Evidence[1]

Existing studies of educational activities, strategies and interventions can be useful for assessing whether the evaluated programs and practices would work in an education decision-maker's own context. However, decision-makers may struggle to determine how relevant a study is to their own context and how seriously they should take the findings given varying levels of research quality and credibility. Decision-makers may want to consider results from a less rigorous study but adjust for the fact that the study may over- or under-estimate the effectiveness of the program being studied.

To account for these issues, DecisionMaker provides two indices developed by CBCSE that can be used to evaluate the relevance and credibility of studies: the Relevance Index and the Credibility Index. We recommend you review the items in each index below before reading each study you want to consider as evidence so you know what to look out for.

To find the information you will need for completing this worksheet, look for section headings in the study that use terms like "Methods" or "Study Design," "Measures," and "Sample" or "Setting." Search functions in a Word or PDF document can help with this (e.g., Ctrl F).

Key terms

A **target population** is the complete collection of students/teachers/schools or other units that we want to study or consider.[2] For example, the target population of a study on algebra skills of high school students in a particular school district may be all 150,000 students in every high school in the district.

An **outcome** is a change or impact caused by the program or strategy being evaluated, or it could be a characteristic of the respondents you want to measure.[3] An outcome of a reading program for third graders in an elementary school might be to raise reading skills among third grade students.

Measures are the items in a research study to which the participants respond[4], which are used to assess performance on the outcomes of interest.

[1] The design of these indices was influenced by a number of sources: Digital Promise's "Evaluating Studies of Ed Tech Products," NESTA's "Standards of Evidence," Alliance for Useful Evidence's "What Counts as Good Evidence," David Gough's "Weight of Evidence: A Framework for the Appraisal of the Quality and Relevance of Evidence," and REL Central's Draft "Applicability of Evidence-Based Interventions" Tool.
[2] Lohr. S. L. (2010). *Sampling: Design and Analysis*. 2nd ed. Boston, MA: Brooks/Cole, Cengage Learning.
[3] https://www.povertyactionlab.org/research-resources/measurement-and-data-collection
[4] http://www.uniteforsight.org/research-methodology/module4

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

A. Worksheet to Assess Relevance of Available Evidence to Your Purpose and Context:

Step 1. First, establish if prior evidence exists:

1. Are there prior studies conducted on this educational program or strategy (Solution Option) which assess its effectiveness in improving student or teacher outcomes?

Yes No

➔ If no, you may need to plan to collect your own data. (See suggestions for resources to help with this in *DecisionMaker*, Resources & Guidance, VII Evaluation Measures and Results).

Step 2. If you find studies that provide information on effectiveness, follow the steps below (Steps 2.a – 2.e) to use the Relevance Index for establishing whether this evidence is relevant to your context. Later, you will be guided to use the Credibility Index (Steps 3 – 13) to evaluate the credibility of studies you deem relevant to your context.

2. **To help you determine whether a study was conducted in a similar context to your school/ district/ state:**

   a. Think about which of the factors in the **Relevance Index** below are important to consider for your decision problem and fill in the corresponding radio buttons. You may decide to choose all or just a few.
   b. Read each study carefully and score it between 0 and 3 on each factor you chose. Your score should indicate how similar the study context is to your own context for this factor: *3= very similar, 2 = moderately similar, 1 = slightly similar, 0 = not at all similar.*
   c. Divide the total score earned by each study by the total maximum possible score on this index and then multiply by 100 to give you a Relevance Index between 0 and 100.

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

Steps 2.a – c: Relevance Index

Contextual Factor	***Things to look for and think about***	***Important consideration for me as to whether this study is relevant to my context*** *(fill in radio button as needed)*	***Details of the Study***	***Study context is similar to mine*** *3= very 2 = moderately 1 = slightly 0 = not at all*
Recency of study	Extent to which the educational activity, strategy or intervention (Solution Option) is still likely to be applicable in my context today	○		
Student demographics	Age of students on which the program/ intervention was tested	○		
	Baseline performance of students before implementing the program	○		
	Percent economically disadvantaged, e.g., as indicated by FRPL	○		
	Percent classified as minority	○		
	Percent ELL	○		
	Percent Special Education	○		
School context	Charter vs. district school vs. private	○		
	Selective vs. open to all applicants	○		
	Qualifications of teachers (e.g., degree level, certification, tenure status, average experience)	○		
	Availability of necessary staff	○		
	Availability of necessary technology	○		
	Urban, rural or suburban	○		
Relevance of measure used	Whether the outcome measure reported in the study is relevant for your goals, e.g., if you are trying to reduce behavior incidents, a study that	○		

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

	reports suspension rates might be relevant, but one that reports attendance might be less relevant			
Scale	Is the scale at which the study was conducted similar to your context and needs (e.g., if you want to implement a reading program at large scale across several schools, did the study look at the effect of the intervention across many schools, or only at small scale in one school?)	○		
Total possible score (Y) = Number of factors you selected x 3 points = _____				**Total points you gave this study (X): _____**
Relevance Index = X/Y x 100 = _____%				

Note. FRPL = Free and Reduced Price Lunch; ELL = English Language Learner.

d. Compare the overall score for each study with the interpretation table below to come to a conclusion about which, if any, of the studies are relevant enough to your context.

STEP 2.d: Relevance Index Interpretation Table

Relevance Index (= Total points for this study/Total possible score x 100)	**Relevance Rating**
Less than 30%	Low Relevance
31 - 69%	Moderate Relevance
70% or higher	High Relevance

e. Record your assessment of each study in the **Relevance Summary Table** below.

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

STEP 2.e: Relevance Summary Table

Name of study	**Authors**	**Year of study**	(Column D) **Relevance Score**	(Column E) **Total possible score**	**Relevance Index** (Column D/ Column E)	**Relevance Rating** (High, Moderate, Low)	**Use this study as relevant evidence for this decision** (Yes/No)

Next Steps

- We recommend you move studies that earn a rating of High or Moderate relevance forward to determine whether they are credible.
- If you do decide to move forward with a study of low apparent relevance to your context, you may want to be sure that it scores high on credibility in the next step.
- If none of the studies you review score better then Low Relevance, you may want to consider designing your own study and collecting your own data.

Summary

The output of Steps 2a-2e above on Relevance is to provide a Relevance rating for each study you review, and a decision on whether to use any of these studies as evidence for the Solution Option(s) (educational activities, strategies or interventions) you are considering, or to collect your own data.

If one or more studies pass the Relevance threshold (greater than 30%), then you can proceed with the relevant study to the next step to assess credibility. Our Relevance threshold is merely a guideline, but you can choose your own threshold by which to accept studies that makes sense given the context and decision problem at hand.

If no study passes the threshold, then we recommend you collect your own data to evaluate the Solution Option(s) you are considering.

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

B. Worksheet to Assess the Credibility of Available Evidence

The Credibility Index can be used to assess how seriously you should take the findings of each study you reviewed which passed the 'Relevance threshold' in Part A above.

Credibility Index

Credibility Index Part I: Source of Study

3. For each section (1-2) in Part I of the Credibility Index, select the statement that most accurately reflects the study you are reviewing. Fill in the radio buttons to select ONE statement only per section and write in the points to award in the last column.
4. Add up the points you awarded in the last column for these two sections.

Credibility Part I:	***Details of the Study***	***What to look for and think about***	***Select ONE statement per section***	***Points to award***	***Write in points awarded***
Section 1: Who conducted the study?		It is not clear who conducted the study.	○	-1	
		The study was conducted by the program vendor.	○	0	
		The study was conducted by an external evaluator hired by the program vendor.	○	1	
		The study was conducted by an external evaluator acting as an independent third party, i.e., not paid by the vendor. This may include collaborations between the evaluator and implementing partners such as school districts or other educational or research institutions, but excluding the program vendor. The key distinction is that those conducting and/or commissioning the evaluation should not stand to make a profit	○	2	

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

		from the sale of the product/ intervention being tested.			
Section 2: Who published the study?		It is not clear where the study was published.	○	-1	
		The study was published by the vendor.	○	0	
		The study was published by a third party (i.e., other than the vendor) but not a peer-reviewed journal, e.g., a university, research organization, school district or other government agency. Keep in mind some technical reports are also later published in a peer-reviewed journal, so you may wish to check Google Scholar to see if there is a later published version.	○	1	
		The study was published in a peer-reviewed journal.	○	2	
Total points for Part I: Sections 1-2				**Max possible: 4**	Total points awarded: __

Credibility Index Part II: Implementation and Goals of the Intervention

5. In Part II of the Credibility Index below, assess how well the study meets its stated goals, how clearly it explains the components of the treatment and how it was implemented. Fill in the radio buttons to select ONE statement only per section and write in the points to award in the last column.

6. Add up the points you awarded in the last column for these two sections.

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

Credibility Part II:	***Details of the Study***	***Things to look for and think about***	***Select ONE statement per section***	***Points to award***	***Write in points awarded***
Section 3: Length of time participants are exposed to the educational program/strategy being studied		Length of exposure is not clear from the study.	○	-1	
		Length of exposure is too short to make a difference.	○	0	
		Length of exposure is too long to reflect likely effect in regular practice.	○	1	
		Length of exposure is about right.	○	2	
Section 4: Meaningful outcomes		It is not clear which outcomes are being measured, e.g., it is not clear whether the study is evaluating geometry skills or algebra skills.	○	-1	
		The outcomes measured are not at all aligned with the ultimate goal for implementing the intervention, e.g., the study investigates whether an after-school supplemental math program improves geometry skills, despite the fact that the program aims to improve algebra skills.	○	0	
		The outcomes measured only capture short-term behavioral changes but not longer-term educational outcomes, e.g., the study only documents whether students are attending an after-school math program, but does not measure whether their math skills are improving.	○	1	
		The outcomes measured are aligned with some but not all of the stated goals for implementing the intervention, e.g., the study is measuring algebra skills when the primary goal of the program is to improve both algebra and geometry skills.	○	2	

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

		The outcomes measured are aligned with all of the stated goals for implementing the intervention, e.g., the study is measuring algebra skills when improving algebra skills is the primary goal of the program.	○	3	
Section 5: Definition of treatment: Is it clear what program participants were supposed to do compared with business as usual conditions?		The components of the program/treatment are not clear.	○	-1	
		The components of the program/treatment are partially clear.	○	1	
		The components of the program/treatment are completely clear.	○	2	
Section 6: Implementation of the treatment How clear are the details of implementation (dosage, frequency, whether it was implemented in/out of the classroom, before/ after school hours etc.)?		The implementation details of the intervention are not clear.	○	-1	
		The implementation details of the intervention are partially clear.	○	1	
		The implementation details are completely clear.	○	2	
Total points for Part II: Sections 3-6				**Max possible: 9**	Total points awarded: __

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

Credibility Index Part III: Study Sample

What is a sample? *Imagine that you are the superintendent of a large, diverse school district, and you want to investigate the social and emotional competencies of all high school students in your district. However, the school district contains 80 high schools with a total student population of 150,000. You have a limited budget, and decide to ask a research team to collect the data for you. It might not be feasible to go into all 80 high schools in the district and get all 150,000 students to take an assessment of social and emotional competencies, so the research team will likely select a subset of students, i.e., a sample, to make the data collection process more feasible.*

Key terms

A **sample** is a subset of a population, for example, a subset of the students/teachers/schools that make up a target population for an evaluation.[5]

A sample is **representative** if the sample is similar to the target population on all important characteristics.

Sample Size: The number of units (e.g., students/teachers/schools) in a sample.

Statistical power: The probability that the estimate of the program effect will be found statistically significant if an effect of that size is determined to have occurred.[6]

There are two concerns with drawing a sample in order to get trustworthy results in an effectiveness study:

- Is the sample representative?
  - For example, if the district is 50% FRPL, 75% minority and 13% ELL, the researchers should aim to draw a sample that has a similar distribution of these characteristics.

- Is the sample large enough to detect an effect when indeed there is one?
  - For example, if the sample only has 10 participants and the study aims to measure outcomes for social emotional competencies, this would probably be too small a sample. However, if the study aims to measure ease of implementation across 10 different classrooms, this sample size would be more reasonable.

[5] Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2019). *Evaluation: A systematic approach*. 8th Edition. Sage Publications.
[6] Ibid.

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

7. Assess the study in question on its representativeness and sample size in Part III of the Credibility Index below. Fill in the radio buttons to select ONE statement only per section and write in the points to award in the last column.
8. Add up the points you awarded in the last column for the two sections.

Credibility Part III:	***Details of the Study***	***Things to look for and think about***	***Select ONE statement per section***	***Points to award***	***Write in points awarded***
Section 7 Representativeness: How much does the study sample mirror the target population of the study? Think about the key factors that matter for what the study is trying to measure – does the sample have a similar distribution of these factors compared with the target population? (Note: This is different from assessing whether the sample mirrors *your own* population, which pertains to Relevance).		The characteristics of the study sample are not clear.	○	-1	
		The study sample is not representative of the target population.	○	0	
		The study sample is moderately representative of the target population.	○	1	
		The study sample is highly representative of the target population.	○	2	
Section 8 Sample size: Does the size of the sample, i.e., the number of participants in the study, seem adequate? (For researchers: do you think there is enough power to detect an effect if indeed there is one?)		The size of the study sample is unclear.	○	-1	
		The study did not have an adequate number of participants.	○	0	
		The study had a fairly adequate number of participants.	○	1	
		The study had a very adequate number of participants.	○	2	
Total points for Part III: Section 7 and 8				**Max possible: 4**	Total points awarded: __

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

Credibility Index Part IV: Rigor of Methodology

Key terms

A comparison group is a group that did not participate in a program and whose results on the outcome of interest can be compared with those of the group that *did* participate in the program. The key challenge in a good efficacy study is to find a comparison group that closely resembles program participants, meaning both groups should be similar on average across all main observable characteristics, such as student demographics or school characteristics.

To isolate the effects of a social program, researchers conducting effectiveness studies need to measure the outcomes for the individuals exposed to the program (the "treatment group") and find a credible way to estimate the outcomes that would have occurred in the absence of the program.[7] To do so, researchers must identify a comparison or "control" group that is similar to the group exposed to the program except for participation in the program.

There are two considerations:

- The most important consideration is **identification of a credible comparison group**. A credible comparison group is one that is similar to the group that received the program on characteristics that are relevant for the goals of program. To assess how credible a comparison group is, ask **how** program participants were selected to participate in the program: were program participants selected because of certain student, teacher or school characteristics?
    - For example, were students performing below grade level in literacy chosen to participate in a reading program? If so, does the comparison group perform at the same baseline reading levels as program participants?
    - Or were schools with motivated principals and strong infrastructure selected to participate in the program? If so, does the comparison group have equally motivated principals and similar infrastructure to the program schools?

- The second consideration is **whether data on outcomes are collected multiple times** for the treatment and the comparison group, e.g., before and after the program, and on subsequent occasions. This can be useful if you want to account for baseline differences between the two groups, or if you want to measure longer-term outcomes.

[7] Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2019). *Evaluation: A systematic approach*. 8th Edition. Sage Publications.

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

9. Fill in the radio buttons to select ALL statements in Part IV of the Credibility Index below that apply to the study you are reviewing and write in the points to award in the last column.
10. Add up the points you awarded in the last column for Section 9.

Credibility Part IV:	*Details of the Study*	*Things to look for and think about*	*Select ALL that apply*	*Points to award*	*Write in points awarded*
Section 9: Rigor of methodology for evaluating outcomes	1. Determining whether there is a credible comparison group				
	1a. First determine whether there is a comparison group of any kind.				
		It is not clear whether there is a comparison group.	○	-1	
		There is no comparison group.	○	-1	
		The study includes a comparison group which does not participate in the program being studied.	○	1	
	1b. Then, identify how that comparison group was selected.				
		It is unclear how program participants were selected.	○	-1	
		Program participants were selected based on certain observable characteristics (e.g., gender, academic performance), and the comparison group is not similar on those characteristics. For example, the lowest-performing students were selected to participate in a reading program, and the comparison group includes high-performing students.	○	0	
		The study compares outcomes for students/teachers/schools who are receiving the program with outcomes for counterparts who have similar characteristics but are not participating in the program. It may do so either by identifying a comparison group that shares several known characteristics with the program participants, e.g., same grade, gender, SES (statistical matching), or by first matching program participants with non-participants who could have been just as likely to participate in the program, as predicted by known characteristics such as age	○	1	

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

		and gender, and then comparing outcomes for the matched pairs (propensity score matching techniques).[8]			
		The intervention is provided to students/teachers/schools who are above a cutoff point for eligibility. The study compares participants who are just above the cutoff, and therefore receive the intervention, with students/teachers/schools who are just below the cutoff, and therefore do not receive the intervention. This design should ensure the two groups are highly comparable.	O	3	
		The study uses a randomized controlled trial (RCT) in which students/teachers/schools are chosen at random to either participate in the program or to serve in a comparison group.	O	5	
		If the study does not assign students/teachers/schools at random to participate in the program or in the comparison group, does it address issues of bias/confounding variables in some way? E.g., does the study attempt to account for, or at least discuss the possibility of, other factors besides the treatment explaining the difference in outcomes between treatment and control groups?	O	1	
	2. Measuring outcomes over time				
		The study includes before and after measures, e.g., a pre-test/survey/observation before the intervention and a post-test/survey/observation after the intervention.	O	1	
		The study includes a second post-test several months after the intervention ends.	O	1	
		The study assesses outcomes multiple times before/during and after the intervention.	O	1	
Total for Part IV: Section 9				**Max Possible: 10**	Total points awarded: __

[8] https://www.povertyactionlab.org/sites/default/files/resources/2016.08.31-Impact-Evaluation-Methods.pdf

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

11. Now add up the points earned by this study for Parts I, II, III and IV of the Credibility Index

Summary Score for Credibility Index

Totals Points for Parts I-IV:	Max Possible	Points earned
Part I	4	
Part II	9	
Part III	4	
Part IV	10	
GRAND TOTAL	**27**	Total points awarded: __

12. Use the Credibility Index Interpretation Table below to find the credibility band your score falls into.

Credibility Index Interpretation Table

Credibility Index (Total Points for Parts I, II, III, IV)	**Credibility Rating**	**Credibility Parameter**
Less than 10	Low credibility	0.2
10-19	Moderate credibility	0.6
20-27	High credibility	1

13. For low or moderate credibility studies, you can multiply the effect size found in the study you reviewed by the relevant Credibility Parameter to adjust it downwards. Note that this adjustment should be viewed as a reflection of your professional judgement rather than as scientific evidence!

For example, if you are trying to assess a computer-assisted learning program for impact on standardized test scores and a study you reviewed of "Option 1" reported an effect size of 0.3 but received a Credibility rating of "Moderately credible," this is how you would proceed:

RELEVANCE AND CREDIBILITY INDICES FOR STUDIES OF EDUCATIONAL INTERVENTIONS

Credibility Parameter: Moderately credible = 0.6
Impact on standardized test scores (as taken from the evaluation study) = 0.3

Multiply the effect size reported in the study by the Credibility Parameter: [0.3] x [0.6] = [0.18]

Use the new effect size, 0.18, as the expected effectiveness for Option 1. In *DecisionMaker* you would enter 0.18 in the evaluation measures table.

14. Use the Relevance and Credibility Summary Table below to document your assessments of each study you reviewed.

Relevance and Credibility Summary Table

Name of study	**Authors**	**Yr of study**	**Relevance Index**	**Relevance Rating** (High, Moderate, Low)	**Use study as relevant evidence for decision** (Yes/No)	**Credibility Index** (Total Score adding Parts I, II, III, IV)	**Credibility Rating** (High, Moderate, Low)	**Effect size reported in study**	**Adjusted effect size**

Summary

- **The output of Step B on Credibility is to assign a low, medium or high Credibility Parameter to a study that evaluates the effectiveness of an educational program or strategy. This parameter will function as a weight between 0 and 1.**
- **You can multiply the effect size found in a study by this weight to adjust it for credibility. Effect sizes from high credibility studies would obviously remain unchanged while the effect sizes for low and moderate credibility studies would become smaller.**
- **You may also simply consider not relying at all on studies that have very low credibility.**