How is the Relevance Index (RI) Calculated?

The Relevance Index (RI) Excel tool assesses how similar a study's context is to your own context using the ratings you select for each important consideration, or contextual factor. Conceptually, the RI is calculated as follows:

For the contextual factors you choose:

$$\frac{\text{Sum of (your rating of how similar the study's context is to yours } \mathbf{x} \text{ importance value)}}{\text{Sum of importance values you assign to each contextual factor } \mathbf{x}\ 3}$$
(Numerator) / *(Denominator)*

The mathematical basis for the RI is described below.

STEP 1: Generate the denominator of the RI by selecting the importance of each consideration in Column C, *"Important consideration for me:"* By selecting how important each contextual consideration, or factor, is for you, you are assigning the following values to each one:

Importance rating		**Importance value assigned (IV)**
Not at all/unavailable	=	0
Slightly	=	1
Moderately	=	2
Very	=	3

The denominator of the RI is generated by multiplying the sum of importance values assigned to each contextual factor by 3. This represents the maximum number of points a study can score in your analysis. The more factors you indicate are important to you, the higher the maximum possible score will be. If you selected all 16 factors listed and indicated they were all *very* important, the maximum possible score would be 16 x 3 x 3 = 144. If you chose only four factors and indicated they were each moderately important, the maximum possible score would be 4 x 2 x 3 = 24. *(Note that if you write in additional factors, the maximum possible score increases accordingly).*

The reason for multiplying each importance value by 3 is because this is the highest score that a study can earn for similarity of each contextual factor in the study to your own context. You will assess similarity in the next step.

STEP 2: Generate the numerator of the RI using ratings in Column H (or L, P, T, X etc.), "Study context is similar to mine:" You assign points to each important consideration by rating how similar the details of the study are to your own context. A maximum of 3 points can be awarded as follows:

Similarity of study context to mine	**Similarity points assigned (SP)**	
Not at all	=	0
Slightly	=	1
Moderately	=	2
Very	=	3

The numerator of the RI is the sum of the product of each factor's importance value and similarity points.

Example application of RI: You evaluated a study about use of digital math tools and chose five contextual factors on which to base your judgement about the study's relevance to your context. Your assessment was as shown in Table 1.

Table 1. Example scoring in an application of the RI to a study about digital math tools

Contextual factor	Importance rating	Importance value assigned (IV)	Similarity rating	Similarity points assigned (SP)	Factor' s IV x SP
	Denominator of RI		**Numerator of RI**		
Recency	Very	3	Slightly	1	3
Age of students	Moderately	2	Very	2	4
Availability of necessary technology	Very	3	Slightly	1	3
Relevance of outcome measure used	Very	3	Moderately	2	6
Scale	Slightly	1	Not at all	0	0
TOTAL		**12**			**16**
	Multiply by 3 = 12 x 3 = **36**				

First, the denominator of the RI is generated by summing the importance values you assigned to each contextual factor selected, and multiplying the sum by three:

$$Denominator = (3 + 2 + 3 + 3 + 1) \times 3 = 12 \times 3 = 36$$

Second, the numerator of the RI is generated by multiplying the importance value you assigned to each factor you selected by the similarity points you awarded, and summing the product:

$$Numerator = (3 \times 1) + (2 \times 2) + (3 \times 1) + (3 \times 2) + (1 \times 0) = 16$$

STEP 3: Calculating the RI: The RI is calculated by dividing the numerator (i.e., the sum of each factor's importance value and similarity points from Step 2) by the denominator (i.e., the total possible points obtained from Step 1):

$$RI = \frac{[IV1 * SP1] + [IV2 * SP2] + [IV3 * SP3] \ldots}{3\ x\ (IV1 + IV2 + IV3 \ldots)}$$

The RI for this example using the scores in Table 1 is:

$$RI = \frac{16}{36} = 44\%$$

STEP 4: Interpreting the RI: The interpretation for the RI is presented in Table 2.

Table 2. Interpretation of RI

Relevance Index	Relevance Rating
Less than 30%	Low Relevance
31 - 69%	Moderate Relevance
70% or higher	High Relevance

STEP 5. We recommend you move forward with studies that earn a rating of High or Moderate

relevance to determine whether they are credible. However, these are only our recommendations, and you may choose to move forward with studies at lower or higher cutoff points based on your own standards.

How is the Credibility Index (CI) Calculated?

STEP 1. By selecting the relevant answer/statement for each question listed, you are assigning the corresponding scores in Table 3 to the study for each question.

STEP 2. The CI is calculated by summing all the points you obtained for each question. The maximum possible score a study can earn on the CI is 27 and the minimum is -9.

Table 3. Scores assigned to each study characteristic when applying the CI rubric

Q1. Who conducted the study?	Corresponding points
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 from the sale of the product/ intervention being tested.	2
Q2. Who published the study?	Corresponding points
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
Q3. Length of time participants are exposed to the educational program/strategy being studied?	Corresponding points
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
Q4. Components of the program: Is it clear what program participants were supposed to do compared with business as usual conditions?	Corresponding points
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
Q5. Implementation of the program: How clear are the details of implementation (dosage, frequency, whether it was implemented in/out of the classroom, before/ after school hours etc.)?	Corresponding points
The implementation details of the program are not clear.	-1

The implementation details of the program are partially clear.	1
The implementation details of the program are completely clear.	2
Q6. Meaningful outcomes?	Corresponding points
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
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
Q7. Sample characteristics: Are the characteristics of the sample clear?	Corresponding points
The characteristics of the study sample are not clear.	-1
The characteristics of the study sample are moderately clear.	1
The characteristics of the study sample are very clear.	2
Q8. 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?)	Corresponding points
The size of the study sample is unclear.	-1
The study does not have an adequate number of participants.	0
The study had a fairly adequate number of participants.	1
The study has a very adequate number of participants.	2
Q9. First determine whether there is a comparison group of any kind.	Corresponding points
It is not clear whether there is a comparison group.	-1
There is no comparison group.	-1
Study includes comparison group which does not participate in the program being studied.	1
Q10. Then, identify how that comparison group was selected. (Only answer 10a and 10b if the study includes a comparison group)	
10a. Clarity of selection process for program v. comparison groups.	Corresponding points
Unclear how study participants were selected to be in the program or the comparison group.	-1

The selection process for program v. comparison group participants is clearly described.	0
10b. Study design	Corresponding points
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 and gender, and then comparing outcomes for the matched pairs (propensity score matching techniques).	1
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 eligibility cutoff, and therefore receive the intervention, with students/teachers/schools who are just below the eligibility cutoff, and therefore do not receive the intervention. This design should ensure the two groups are highly comparable.	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.	5
The study uses a randomized controlled trial (RCT) plus utilizes propensity score matching or statistical matching techniques.	6
10c. Addressing other possible explanations for the results: If the study does not assign students/teachers/schools at random to participate in the program or in the comparison group, do the authors 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?	Corresponding points
The study does not address other possible explanations for the results.	0
The study discusses other possible explanations for the results.	1
Q11. Measuring outcomes over time	Corresponding points
The study includes only post-test measures.	0
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.	1
The study assesses outcomes multiple times before/ during and after the intervention, which may include a second post-test several months after the intervention has ended.	2

STEP 3. Use Table 4, the Credibility Index Interpretation Table, to find the credibility band your score falls into. Again, these are CBCSE's suggested cut offs and you may prefer to establish your own. We suggest that you do not use studies that earn low credibility for making high-stakes decisions.

Table 4. Credibility Index Interpretation Table

Credibility Index	Credibility rating	Credibility parameter
Less than 10	Low Credibility	0.2
10-19	Medium Credibility	0.6
20-27	High Credibility	1.0

STEP 4 (OPTIONAL). Adjust effect size: For low or moderate credibility studies, you can consider making a downward adjustment to reflect the fact that your confidence in the findings is not high. The credibility parameter is a suggested adjustment that serves this purpose. 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 "Program A" reported an effect size of 0.3 but received a Credibility rating of "Moderately credible," this is how you would proceed:

Credibility Parameter for a "Moderately credible" study = 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 Program A. If you are using *DecisionMaker*, you would enter 0.18 in the Evaluation Data Table.

Example Application of Credibility Index

Table E1 provides an example of the CI scoring for the study of eSpark described in "Evaluating Digital Math Tools in the Field" (Hollands & Pan, 2018). As indicated at the bottom of the table, the study earns a CI of 20 (maximum possible is 27).

Table E1. Example CI scoring of the eSpark study in Evaluating Digital Math Tools in the Field

	Question	Selected statement	Corresponding points
Q1	Who conducted the study?	The study was conducted by an external evaluator acting as an independent third party, i.e., not paid by the vendor.	2
Q2	Who published the study?	The study was published in a peer-reviewed journal.	2
Q3	Length of time participants are exposed to the educational program/strategy being studied	Length of exposure is about right.	2
Q4	Components of the program: Is it clear what program participants were supposed to do compared with business as usual conditions?	The components of the program/treatment are completely clear.	2
Q5	Implementation of the program: 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 program are completely clear.	2
Q6	Meaningful outcomes	The outcomes measured are aligned with all of the stated goals for implementing the intervention: the study measures math skills and improving math skills is the primary goal of the programs.	3
Q7	Sample characteristics: Are the characteristics of the sample clear?	The characteristics of the study sample are very clear.	2
Q8	Sample size: Does the size of the sample, i.e., the number of participants in the study, seem adequate?	The study had a fairly adequate number of participants.	1
Q9	First determine whether there is a comparison group of any kind.	Study includes comparison group which does not participate in the program being studied.	1
Q10a	Clarity of selection process for program v. comparison groups.	The selection process for program v. comparison group participants is clearly described.	0
Q10b	Study design	The study compares outcomes for	1

		students 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 and gender, and then comparing outcomes for the matched pairs (propensity score matching techniques).	
Q10c	Addressing other possible explanations for the results	The study discusses other possible explanations for the results.	1
Q11	Measuring outcomes over time	The study includes before and after measures: a pre-test before the intervention and a post-test after the intervention.	1
		Total points awarded:	**20**
		Total possible points	**27**