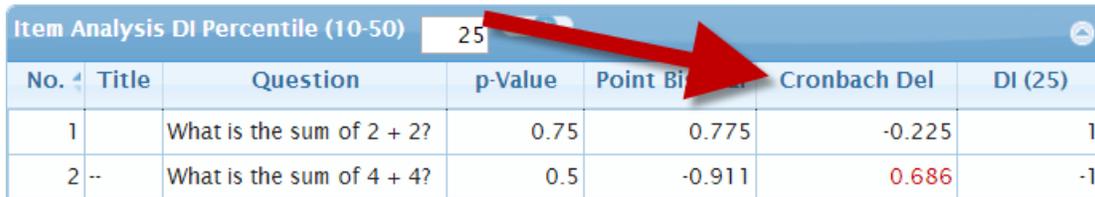


Cronbach Alpha with Deletion



No.	Title	Question	p-Value	Point Bi	Cronbach Del	DI (25)
1		What is the sum of 2 + 2?	0.75	0.775	-0.225	1
2	--	What is the sum of 4 + 4?	0.5	-0.911	0.686	-1

Cronbach Alpha with Deletion helps assess test question reliability.

How? It asks whether the exam as a whole is more reliable if you simply **delete** the question under review. Basically, the Cronbach Alpha with Deletion re-runs the exam's KR-20 without the question under review. If the exam as a whole is more reliable without it, there's probably something wrong with that question.

The Cronbach Alpha with Deletion generally ranges between **0.0** and **+1.0**, but it can fall below 0.0 with smaller sample sizes.

More important than its range is how the Cronbach Alpha with Deletion compares to the exam's KR-20. If a question's Cronbach Alpha with Deletion is **greater than** the exam's KR-20, it means the exam as a whole is more reliable without it. For example, take a look at Question No. 2 in the picture above. If the exam's KR-20 is 0.545, then we know Question No. 2 is "suspect" because its Cronbach Alpha with Deletion of **0.686** is greater than 0.545.

EAC suggestion: Look out for questions with a Cronbach Alpha with Deletion greater than the exam's KR-20. These questions, which EAC Visual Data **highlights in red**, decrease overall test reliability and should be considered suspect.

Learn more at [The EACs of Test Statistics](https://eacVisualData.com/eacs)

eacVisualData.com/eacs

Christopher Heisen, Managing Director
phone: 800.954.0380 ext 101
email: Christopher@eacVisualData.com

Educational Assessments Corporation



eacVisualData.com

distractor point biserial

The **distractor point biserial correlation** digs deeper than the item statistics and measures the reliability of **each answer choice** presented to students. How? It correlates student scores on each answer choice with their scores on the test as a whole.

The driving assumption is simple: Students who score well on the test as a whole should on average select the correct answer choice for each question. Students who struggle on the test as a whole should on average select an incorrect answer choice for each question. If an answer choice deviates from this assumption, the distractor point biserial correlation lets us know.

The distractor point biserial correlation ranges from a low of **-1.0** to a high of **+1.0**. The closer a **correct** answer choice's distractor point biserial correlation is to **+1.0** the more reliable the answer choice is considered because it discriminates well among students who mastered the test material and those who did not. This answer choice "works well."

By the same token, the closer an **incorrect answer** choice's distractor point biserial correlation is to **-1.0** the more reliable the answer choice is considered because it discriminates well among students who did not master the test material and those who did.

Since making sense of distractor point biserial correlations can be challenging, we provide a simple Distractors Key in EAC Visual Data.

EAC suggestion: Refer to the Distractors Key to assess whether each answer choice is "working" as expected (i.e., whether each answer choice discriminates among students who have mastered the test material and those who have not). Consider changing answer choices that aren't "working" as expected and also those that students don't select at all. If no student selected an answer choice, that answer choice isn't really a "distractor" after all.

Distractors Key		
Incorrect	PB	Correct
Warning	1	Useful
	.5	
	0	
	-.5	
Useful		Warning
	-1	

Learn more at [The EACs of Test Statistics](https://www.eacvisualdata.com/eacs)

[eacVisualData.com/eacs](https://www.eacvisualdata.com/eacs)

Christopher Heisen, Managing Director

phone: 800.954.0380 ext 101

email: Christopher@eacVisualData.com

Educational Assessments Corporation



[eacVisualData.com](https://www.eacvisualdata.com)

KR-20

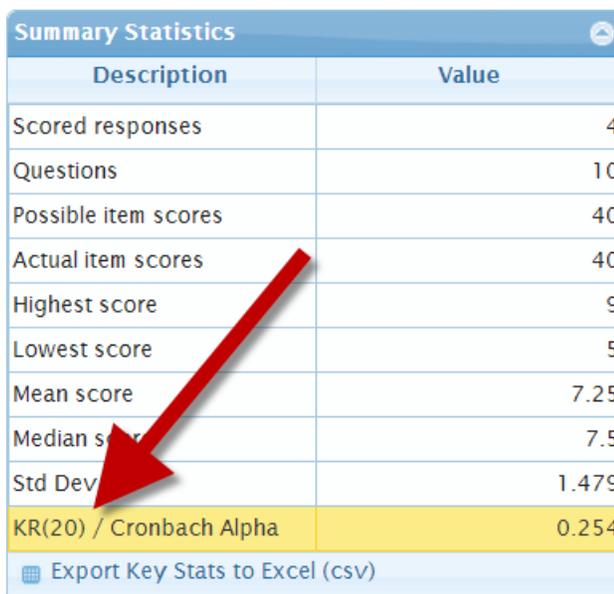
The **KR-20**, or Kuder-Richardson Formula, measures overall test reliability.

Essentially it lets you know whether the exam as a whole discriminated among students who mastered the subject matter and those who did not.

The KR-20 generally ranges between **0.0** and **+1.0**, but it can fall below 0.0 with smaller sample sizes.

The closer the KR-20 is to +1.0 the more reliable an exam is considered because its questions do a good job consistently discriminating among higher and lower performing students.

A KR-20 of 0.0 means the exam questions didn't discriminate at all. Imagine a test where all 20 students answered all 40 questions correctly. The test didn't discriminate among any of them, and its KR-20 of 0.0 makes perfect sense.



Summary Statistics	
Description	Value
Scored responses	4
Questions	10
Possible item scores	40
Actual item scores	40
Highest score	9
Lowest score	5
Mean score	7.25
Median score	7.5
Std Dev	1.479
KR(20) / Cronbach Alpha	0.254

Export Key Stats to Excel (csv)

EAC suggestion: The interpretation of the KR-20 depends on the purpose of the test. Most high stakes exams are intended to distinguish those students who have mastered the material from those who have not. For these, shoot for a KR-20 of **+0.50 or higher**. A KR-20 of less than +0.30 is considered poor no matter the sample size. If the purpose of the exam is to ensure that ALL students have mastered essential skills or concepts or the test is a "confidence builder" with intentionally easy questions, look for a KR-20 close to 0.00.

Learn more at **The EACs of Test Statistics**

eacVisualData.com/eacs

Christopher Heisen, Managing Director

phone: 800.954.0380 ext 101

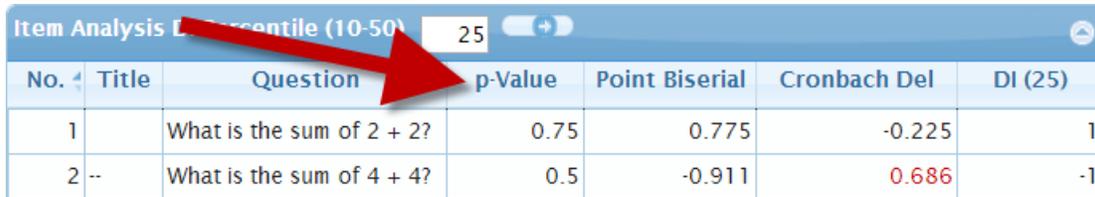
email: Christopher@eacVisualData.com

Educational Assessments Corporation

EAC

eacVisualData.com

p-Value



No.	Title	Question	p-Value	Point Biserial	Cronbach Del	DI (25)
1		What is the sum of 2 + 2?	0.75	0.775	-0.225	1
2	--	What is the sum of 4 + 4?	0.5	-0.911	0.686	-1

In the branch of statistics dealing with test reliability, and unlike other branches of statistics, **p-Value** is a simple measure of question difficulty.

The p-Value ranges from a low of **0.0** to a high of **+1.0**.

The closer the p-Value is to 0.0 the more difficult the question. For example, a p-Value of 0.0 means that no student answered the question correctly and therefore it's a really hard question. If an item's p-Value is unexpectedly close to 0.0, be sure to check the answer key.

The closer the p-Value is to +1.0 the easier the question. For example, a p-Value of +1.0 means that every student answered the question correctly and therefore it's a really easy question.

EAC suggestion: On high stakes exams, shoot for p-Values between **+0.50 and +0.85** for most test questions. A p-Value less than +0.50 means the question may be too difficult or you should double-check the answer key. A p-Value greater than +0.85 means the question may be too easy or most students have mastered that concept.

Learn more at **The EACs of Test Statistics**

eacVisualData.com/eacs

Christopher Heisen, Managing Director

phone: 800.954.0380 ext 101

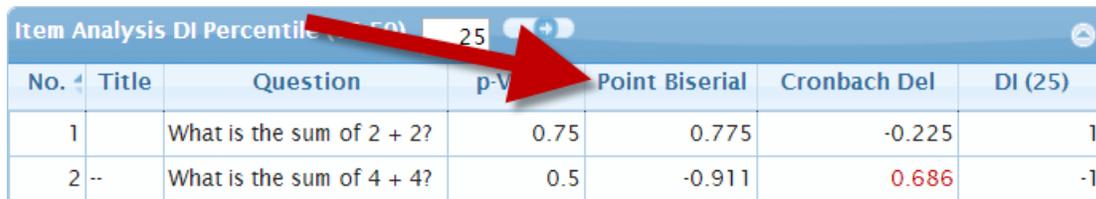
email: Christopher@eacVisualData.com

Educational Assessments Corporation



eacVisualData.com

point biserial correlation



No.	Title	Question	p-V	Point Biserial	Cronbach Del	DI (25)
1		What is the sum of 2 + 2?	0.75	0.775	-0.225	1
2	--	What is the sum of 4 + 4?	0.5	-0.911	0.686	-1

The **point biserial correlation** measures item reliability. How? It correlates student scores on one particular question with their scores on the test as a whole.

The driving assumption is simple: Students who score well on the test as a whole should on average score well on the question under review. Students who struggle on the test as a whole should on average struggle on the question under review. If a question deviates from this assumption (aka, a "suspect" question), the point biserial correlation lets us know.

The point biserial correlation ranges from a low of **-1.0** to a high of **+1.0**. The closer the point biserial correlation is to +1.0 the more reliable the question is considered because it discriminates well among students who mastered the test material and those who did not.

A negative point biserial correlation means that students who performed well on the test as a whole tended to miss the question under review and students who didn't perform as well on the test as a whole got it right. It's a red flag, and there are a number of possible things to check. Is the answer key correct? Is the question clearly worded? If it's multiple choice, are the choices too similar?

EAC suggestion: For high stakes exams intended to distinguish among students who mastered the material from those who did not, shoot for questions with point biserial correlations **greater than +0.30**. They're very good items. Questions with point biserial correlations less than +0.09 are considered poor. Questions with point biserial correlations between +0.09 and +0.30 are considered reasonably good.

Learn more at [The EACs of Test Statistics](#)

eacVisualData.com/eacs

Christopher Heisen, Managing Director

phone: 800.954.0380 ext 101

email: Christopher@eacVisualData.com

Educational Assessments Corporation



eacVisualData.com