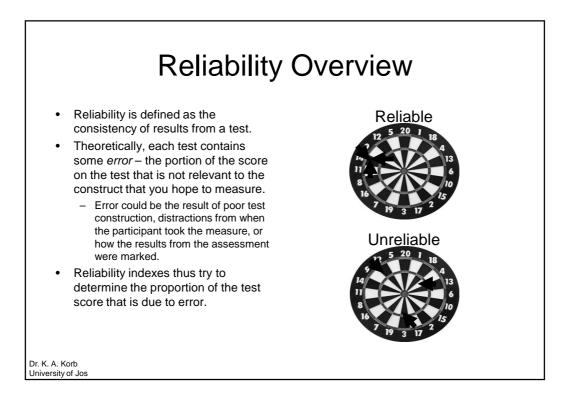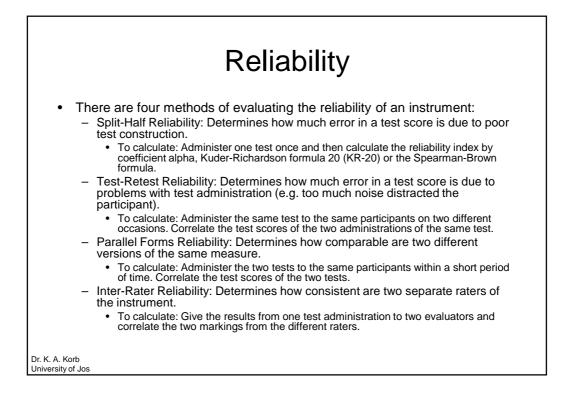# Calculating Reliability of Quantitative Measures

Dr. K. A. Korb
University of Jos

Reliability Overview

- Reliability is defined as the consistency of results from a test.
- Theoretically, each test contains some *error* – the portion of the score on the test that is not relevant to the construct that you hope to measure.
  - Error could be the result of poor test construction, distractions from when the participant took the measure, or how the results from the assessment were marked.
- Reliability indexes thus try to determine the proportion of the test score that is due to error.

Reliable

Unreliable

Reliability

- There are four methods of evaluating the reliability of an instrument:
  - Split-Half Reliability: Determines how much error in a test score is due to poor test construction.
    - To calculate: Administer one test once and then calculate the reliability index by coefficient alpha, Kuder-Richardson formula 20 (KR-20) or the Spearman-Brown formula.
  - Test-Retest Reliability: Determines how much error in a test score is due to problems with test administration (e.g. too much noise distracted the participant).
    - To calculate: Administer the same test to the same participants on two different occasions. Correlate the test scores of the two administrations of the same test.
  - Parallel Forms Reliability: Determines how comparable are two different versions of the same measure.
    - To calculate: Administer the two tests to the same participants within a short period of time. Correlate the test scores of the two tests.
  - Inter-Rater Reliability: Determines how consistent are two separate raters of the instrument.
    - To calculate: Give the results from one test administration to two evaluators and correlate the two markings from the different raters.

Dr. K. A. Korb
University of Jos

Split-Half Reliability

- When you are validating a measure, you will most likely be interested in evaluating the split-half reliability of your instrument.
  - This method will tell you how consistently your measure assesses the construct of interest.
    - If your measure assesses multiple constructs, split-half reliability will be considerably lower. Therefore, separate the constructs that you are measuring into different parts of the questionnaire and calculate the reliability separately for each construct.
    - Likewise, if you get a low reliability coefficient, then your measure is probably measuring more constructs than it is designed to measure. Revise your measure to focus more directly on the construct of interest.
  - If you have dichotomous items (e.g., right-wrong answers) as you would with multiple choice exams, the KR-20 formula is the best accepted statistic.
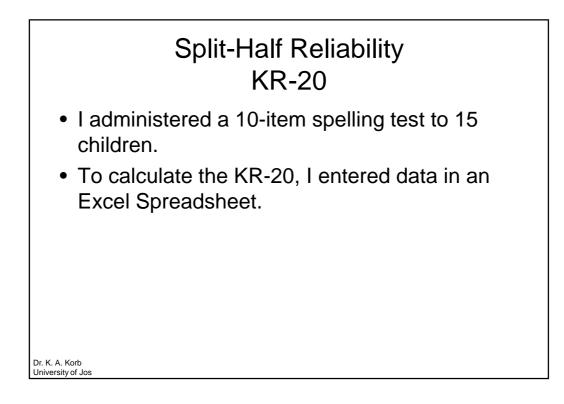 - If you have a Likert scale or other types of items, use the Spearman-Brown formula.

Dr. K. A. Korb
University of Jos

Split-Half Reliability
KR-20

- NOTE: Only use the KR-20 if each item has a right answer. Do NOT use with a Likert scale.
- Formula:

$$r_{KR20} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\Sigma pq}{\sigma^2}\right)$$

  - r_{KR20} is the Kuder-Richardson formula 20
  - k is the total number of test items
  - Σ indicates to sum
  - p is the proportion of the test takers who pass an item
  - q is the proportion of test takers who fail an item
  - σ^2 is the variation of the entire test

Split-Half Reliability
KR-20

- I administered a 10-item spelling test to 15 children.
- To calculate the KR-20, I entered data in an Excel Spreadsheet.

Student Name	Math Problem									
	1. 5+3	**2.** 7+2	**3.** 6+3	**4.** 9+1	**5.** 8+6	**6.** 7+5	**7.** 4+7	**8.** 9+2	**9.** 8+4	**10.** 5+6
Sunday	1	1	1	1	1	1	1	1	1	1
Monday	1	0	0	1	0	0	1	1	0	1
Linda	1	0	1	0	0	1	1	1	1	0
Lois	1	0	1	1	1	0	0	1	0	0
Ayuba	0	0	0	0	0	1	1	0	1	1
Andrea	0	1	1	1	1	1	1	1	1	1
Thomas	0	1	1	1	1	1	1	1	1	1
Anna	0	0	1	1	0	1	1	0	1	0
Amos	0	1	1	1	1	1	1	1	1	1
Martha	0	0	1	1	0	1	0	1	1	1
Sabina	0	0	1	1	0	0	0	0	0	1
Augustine	1	1	0	0	0	1	0	0	1	1
Priscilla	1	1	1	1	1	1	1	1	1	1
Tunde	0	1	1	1	0	0	0	0	1	0
Daniel	0	1	1	1	1	1	1	1	1	1

$$r_{KR20} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\Sigma pq}{\sigma^2}\right)$$

$$k = 10$$

- The first value is k, the number of items. My test had 10 items, so k = 10.
- Next we need to calculate p for each item, the proportion of the sample who answered each item correctly.

$$r_{KR20} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\Sigma pq}{\sigma^2}\right)$$

Student	Math Problem									
Name	**1.** 5+3	**2.** 7+2	**3.** 6+3	**4.** 9+1	**5.** 8+6	**6.** 7+5	**7.** 4+7	**8.** 9+2	**9.** 8+4	**10.** 5+6
Sunday	1	1	1	1	1	1	1	1	1	1
Monday	1	0	0	1	0	0	1	1	0	1
Linda	1	0	1	0	0	1	1	1	1	0
Lois	1	0	1	1	1	0	0	1	0	0
Ayuba	0	0	0	0	0	1	1	0	1	1
Andrea	0	1	1	1	1	1	1	1	1	1
Thomas	0	1	1	1	1	1	1	1	1	1
Anna	0	0	1	1	0	1	1	0	1	0
Amos	0	1	1	1	1	1	1	1	1	1
Martha	0	0	1	1	0	1	0	1	1	1
Sabina	0	0	1	1	0	0	0	0	0	1
Augustine	1	1	0	0	0	1	0	0	1	1
Priscilla	1	1	1	1	1	1	1	1	1	1
Tunde	0	1	1	1	0	0	0	0	1	0
Daniel	0	1	1	1	1	1	1	1	1	1
Number of 1's	6	8	12	12	7	11	10	10	12	11
Proportion Passed (p)	0.40	0.53	0.80	0.80	0.47	0.73	0.67	0.67	0.80	0.73

To calculate the proportion of the sample who answered the item correctly, I first counted the number of 1's for each item. This gives the total number of students who answered the item correctly.

Second, I divided the number of students who answered the item correctly by the number of students who took the test, 15 in this case.

$$r_{KR20} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\Sigma pq}{\sigma^2}\right)$$

- Next we need to calculate q for each item, the proportion of the sample who answered each item incorrectly.
- Since students either passed or failed each item, the sum p + q = 1.
  - The proportion of a whole sample is always 1.
  - Since the whole sample either passed or failed an item, p + q will always equal 1.

$$r_{KR20} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\Sigma pq}{\sigma^2}\right)$$

Student	Math Problem									
Name	1. 5+3	2. 7+2	3. 6+3	4. 9+1	5. 8+6	6. 7+5	7. 4+7	8. 9+2	9. 8+4	10. 5+6
Number of 1's	6	8	12	12	7	11	10	10	12	11
Proportion Passed (p)	0.40	0.53	0.80	0.80	0.47	0.73	0.67	0.67	0.80	0.73
Proportion Failed (q)	0.60	0.47	0.20	0.20	0.53	0.27	0.33	0.33	0.20	0.27

I calculated the percentage who failed by the formula 1 – p, or 1 minus the proportion who passed the item.

You will get the same answer if you count up the number of 0's for each item and then divide by 15.

Dr. K. A. Korb
University of Jos

$$r_{KR20} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\Sigma pq}{\sigma^2}\right)$$

- Now that we have p and q for each item, the formula says that we need to multiply p by q for each item.
- Once we multiply p by q, we need to add up these values for all of the items (the Σ symbol means to add up across all values).

Dr. K. A. Korb
University of Jos

$$r_{KR20} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\Sigma pq}{\sigma^2}\right)$$

$$\Sigma pq = 2.05$$

Student	Math Problem									
Name	1. 5+3	2. 7+2	3. 6+3	4. 9+1	5. 8+6	6. 7+5	7. 4+7	8. 9+2	9. 8+4	10. 5+6
Number of 1's	6	8	12	12	7	11	10	10	12	11
Proportion Passed (p)	0.40	0.53	0.80	0.80	0.47	0.73	0.67	0.67	0.80	0.73
Proportion Failed (q)	0.60	0.47	0.20	0.20	0.53	0.27	0.33	0.33	0.20	0.27
p x q	0.24	0.25	0.16	0.16	0.25	0.20	0.22	0.22	0.16	0.20

In this column, I took p times q.
For example, (0.40 * 0.60) = 0.24

- Once we have p x q for every item, we sum up these values.

.24 + .25 + .16 + … + .20 = 2.05

Dr. K. A. Korb
University of Jos

$$r_{KR20} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\Sigma pq}{\sigma^2}\right)$$

- Finally, we have to calculate σ^2, or the variance of the total test scores.

Dr. K. A. Korb
University of Jos

$$r_{KR20} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\Sigma pq}{\sigma^2}\right)$$

$$\sigma^2 = 5.57$$

For each student, I calculated their total exam score by counting the number of 1's they had.

Student	Math Problem										Total Exam Score
Name	1. 5+3	2. 7+2	3. 6+3	4. 9+1	5. 8+6	6. 7+5	7. 4+7	8. 9+2	9. 8+4	10. 5+6	
Sunday	1	1	1	1	1	1	1	1	1	1	10
Monday	1	0	0	1	0	0	1	1	0	1	5
Linda	1	0	1	0	0	1	1	1	1	0	6
Lois	1	0	1	1	1	0	0	1	0	0	5
Ayuba	0	0	0	0	0	1	1	0	1	1	4
Andrea	0	1	1	1	1	1	1	1	1	1	9
Thomas	0	1	1	1	1	1	1	1	1	1	9
Anna	0	0	1	1	0	1	1	0	1	0	5
Amos	0	1	1	1	1	1	1	1	1	1	9
Martha	0	0	1	1	0	1	0	1	1	1	6
Sabina	0	0	1	1	0	0	0	0	0	1	3
Augustine	1	1	0	0	0	1	0	0	1	1	5
Priscilla	1	1	1	1	1	1	1	1	1	1	10
Tunde	0	1	1	1	0	0	0	0	1	0	4
Daniel	0	1	1	1	1	1	1	1	1	1	9

The variation of the Total Exam Score is the squared standard deviation. I discussed calculating the standard deviation in the example of a Descriptive Research Study in side 34. The standard deviation of the Total Exam Score is 2.36. By taking 2.36 * 2.36, we get the variance of 5.57.

$$r_{KR20} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\Sigma pq}{\sigma^2}\right)$$

$$k = 10$$
$$\Sigma pq = 2.05$$
$$\sigma^2 = 5.57$$

- Now that we know all of the values in the equation, we can calculate r_{KR20}.

$$r_{KR20} = \left(\frac{10}{10-1}\right)\left(1 - \frac{2.05}{5.57}\right)$$

$$r_{KR20} = 1.11 * 0.63$$

$$r_{KR20} = 0.70$$

Split-Half Reliability
Likert Tests

- If you administer a Likert Scale or have another measure that does not have just one correct answer, the preferable statistic to calculate the split-half reliability is coefficient alpha (otherwise called Cronbach's alpha).
    - However, coefficient alpha is difficult to calculate by hand. If you have access to SPSS, use coefficient alpha to calculate the reliability.
    - However, if you must calculate the reliability by hand, use the Spearman Brown formula. Spearman Brown is not as accurate, but is much easier to calculate.

Coefficient Alpha	**Spearman-Brown Formula**
$r_\alpha = \left(\dfrac{k}{k-1}\right)\left(1 - \dfrac{\Sigma\sigma_i^2}{\sigma^2}\right)$	$r_{SB} = \dfrac{2r_{hh}}{1 + r_{hh}}$
where σ_i^2 = variance of one test item. Other variables are identical to the KR-20 formula.	where r_{hh} = Pearson correlation of scores in the two half tests.

Split-Half Reliability
Spearman Brown Formula

- To demonstrate calculating the Spearman Brown formula, I used the PANAS Questionnaire that was administered in the Descriptive Research Study.
    - See the PowerPoint for the Descriptive Research Study for more information on the measure.
- The PANAS measures two constructs via Likert Scale: Positive Affect and Negative Affect.
    - When we calculate reliability, we have to calculate it for each separate construct that we measure.
        - The purpose of reliability is to determine how much error is present in the test score. If we included questions for multiple constructs together, the reliability formula would assume that the difference in constructs is error, which would give us a very low reliability estimate.
    - Therefore, I first had to separate the items on the questionnaire into essentially two separate tests: one for positive affect and one for negative affect.
        - The following calculations will only focus on the reliability estimate for positive affect. We would have to do the same process separately for negative affect.

$$r_{SB} = \frac{2r_{hh}}{1 + r_{hh}}$$

These are the 10 items that measured positive affect on the PANAS.

S/No	Questionnaire Item Number									
	1	3	5	9	10	12	14	16	17	19
1	5	4	4	5	1	4	4	5	3	4
2	5	3	3	4	5	2	3	4	5	4
3	5	3	3	4	5	2	3	4	5	4
4	5	3	3	4	3	4	3	5	4	3
5	5	3	3	4	3	4	3	5	4	3
6	5	3	5	3	1	3	2	3	3	1
7	5	4	4	4	3	3	3	4	4	3
8	5	3	5	4	3	4	4	4	5	3
9	5	5	3	4	2	5	3	5	5	4
10	5	4	3	4	3	4	4	4	4	4
11	4	4	4	3	3	4	4	4	5	4
12	4	3	3	3	1	3	3	5	4	3
13	5	5	3	3	2	3	4	4	3	3
14	3	2	2	3	1	4	3	4	3	4

The data for each participant is the code for what they selected for each item: 1 is slightly or not at all, 2 is a little, 3 is moderately, 4 is quite a bit, and 5 is extremely.

Fourteen participants took the test.

- The first step is to split the questions into half. The recommended procedure is to assign every other item to one half of the test.
  - If you simply take the first half of the items, the participants may have become tired at the end of the questionnaire and the reliability estimate will be artificially lower.

Dr. K. A. Korb
University of Jos

$$r_{SB} = \frac{2r_{hh}}{1 + r_{hh}}$$

The first half total was calculated by adding up the scores for items 1, 5, 10, 14, and 17.

S/No	Questionnaire Item Number										1 Half Total	2 Half Total
	1	3	5	9	10	12	14	16	17	19		
1	5	4	4	5	1	4	4	5	3	4	17	22
2	5	3	3	4	5	2	3	4	5	4	21	17
3	5	3	3	4	5	2	3	4	5	4	21	17
4	5	3	3	4	3	4	3	5	4	3	18	19
5	5	3	3	4	3	4	3	5	4	3	18	19
6	5	3	5	3	1	3	2	3	3	1	16	13
7	5	4	4	4	3	3	3	4	4	3	19	18
8	5	3	5	4	3	4	4	4	5	3	22	18
9	5	5	3	4	2	5	3	5	5	4	18	23
10	5	4	3	4	3	4	4	4	4	4	19	20
11	4	4	4	3	3	4	4	4	5	4	20	19
12	4	3	3	3	1	3	3	5	4	3	15	17
13	5	5	3	3	2	3	4	4	3	3	17	18
14	3	2	2	3	1	4	3	4	3	4	12	17

The second half total was calculated by adding up the scores for items 3, 9, 12, 16, and 19

Dr. K. A. Korb
University of Jos

Split-Half Reliability
Spearman Brown Formula

- Now that we have our two halves of the test, we have to calculate the Pearson Product-Moment Correlation between them.

$$r_{xy} = \frac{\Sigma(X - \overline{X})\,(Y - \overline{Y})}{\sqrt{[\Sigma\,(X - \overline{X})^2]\,[\Sigma Y - \overline{Y})^2]}}$$

In our case, X = one person's score on the first half of items, $\overline{X}$ = the mean score on the first half of items, Y = one person's score on the second half of items, $\overline{Y}$ = the mean score on the second half of items.

$$r_{xy} = \frac{\Sigma(X - \overline{X})\,(Y - \overline{Y})}{\sqrt{[\Sigma(X - \overline{X})^2]\,[\Sigma(Y - \overline{Y})^2]}}$$

S/No	1 Half Total	2 Half Total
1	17	22
2	21	17
3	21	17
4	18	19
5	18	19
6	16	13
7	19	18
8	22	18
9	18	23
10	19	20
11	20	19
12	15	17
13	17	18
14	12	17
Mean	**18.1**	**18.4**

The first half total is X.

The second half total is Y.

We first have to calculate the mean for both halves.

This is $\overline{X}$.

This is $\overline{Y}$.

$$r_{xy} = \frac{\Sigma(X - \overline{X})\,(Y - \overline{Y})}{\sqrt{[\Sigma(X - \overline{X})^2]\,[\Sigma(Y - \overline{Y})^2]}}$$

S/No	1 Half Total	2 Half Total	X - $\overline{X}$	Y - $\overline{Y}$
1	17	22	-1.1	
2	21	17	2.9	-1.4
3	21	17	2.9	-1.4
4	18	19	-0.1	0.6
5	18	19	-0.1	0.6
6	16	13	-2.1	-5.4
7	19	18	0.9	-0.4
8	22	18	3.9	-0.4
9	18	23	-0.1	4.6
10	19	20	0.9	1.6
11	20	19	1.9	0.6
12	15	17	-3.1	-1.4
13	17	18	-1.1	-0.4
14	12	17	-6.1	-1.4
Mean	18.1	18.4		

To get X – $\overline{X}$, we take each second person's first half total minus the average, which is 18.1. For example, 21 – 18.1 = 2.9.

To get Y – $\overline{Y}$, we take each second half total minus the average, which is 18.4. For example, 18 – 18.4 = -0.4

Dr. K. A. Korb
University of Jos

$$r_{xy} = \frac{\Sigma(X - \overline{X})\,(Y - \overline{Y})}{\sqrt{[\Sigma(X - \overline{X})^2]\,[\Sigma(Y - \overline{Y})^2]}}$$

$$\Sigma(X - \overline{X})\,(Y - \overline{Y}) = 12.66$$

S/No	1 Half Total	2 Half Total	X - $\overline{X}$	Y - $\overline{Y}$	(X - $\overline{X}$)(Y - $\overline{Y}$)
1	17	22	-1.1	3.6	-3.96
2	21	17	2.9	-1.4	-4.06
3	21	17	2.9	-1.4	-4.06
4	18	19	-0.1	0.6	-0.06
5	18	19	-0.1	0.6	-0.06
6	16	13	-2.1	-5.4	11.34
7	19	18	0.9	-0.4	-0.36
8	22	18	3.9	-0.4	-1.56
9	18	23	-0.1	4.6	-0.46
10	19	20	0.9	1.6	1.44
11	20	19	1.9	0.6	1.14
12	15	17	-3.1	-1.4	4.34
13	17	18	-1.1	-0.4	0.44
14	12	17	-6.1	-1.4	8.54
Mean	18.1	18.4		Sum	12.66

Next we multiply (X – $\overline{X}$) times (Y – $\overline{Y}$).

After we have multiplied, we sum up the products.

Dr. K. A. Korb
University of Jos

$$r_{xy} = \frac{\Sigma(X - \overline{X})(Y - \overline{Y})}{\sqrt{[\Sigma(X - \overline{X})^2][\Sigma(Y - \overline{Y})^2]}}$$

$$\sqrt{[\Sigma(X - \overline{X})^2][\Sigma(Y - \overline{Y})^2]} = 82.72$$

> To calculate the denominator, we have to square $(X - \overline{X})$ and $(Y - \overline{Y})$

S/No	1 Half Total	2 Half Total	$X - \overline{X}$	$Y - \overline{Y}$	$(X - \overline{X})^2$	$(Y - \overline{Y})^2$
1	17	22	-1.1	3.6	1.21	12.96
2	21	17	2.9	-1.4	8.41	1.96
3	21	17	2.9	-1.4	8.41	1.96
4	18	19	-0.1	0.6	0.01	0.36
5	18	19	-0.1	0.6	0.01	0.36
6	16	13	-2.1	-5.4	4.41	29.16
7	19	18	0.9	-0.4	0.81	0.16
8	22	18	3.9	-0.4	15.21	0.16
9	18	23	-0.1	4.6	0.01	21.16
10	19	20	0.9	1.6	0.81	2.56
11	20	19	1.9	0.6	3.61	0.36
12	15	17	-3.1	-1.4	9.61	1.96
13	17	18	-1.1	-0.4	1.21	0.16
14	12	17	-6.1	-1.4	37.21	1.96
				Sum	90.94	75.24

> Next we sum the squares across the participants. Then we multiply the sums.
> 90.94 * 75.24 = 6842.33.
> Finally, $\sqrt{6842.33}$ = 82.72.

Dr. K. A. Korb
University of Jos

$$r_{xy} = \frac{\Sigma(X - \overline{X})(Y - \overline{Y})}{\sqrt{[\Sigma(X - \overline{X})^2][\Sigma(Y - \overline{Y})^2]}}$$

$$\Sigma(X - \overline{X})(Y - \overline{Y}) = 12.66$$

$$\sqrt{[\Sigma(X - \overline{X})^2][\Sigma(Y - \overline{Y})^2]} = 82.72$$

- Now that we have calculated the numerator and denominator, we can calculate r_{xy}

$$r_{xy} = \frac{12.66}{82.72}$$
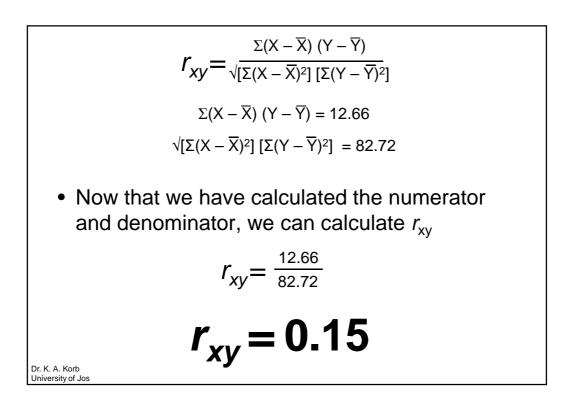
$$r_{xy} = 0.15$$

Dr. K. A. Korb
University of Jos

$$r_{SB} = \frac{2r_{hh}}{1 + r_{hh}}$$

- Now that we have calculated the Pearson correlation between our two halves ($r_{xy} = 0.15$), we substitute this value for r_{hh} and we can calculate r_{SB}

$$r_{SB} = \frac{2 * 0.15}{1 + 0.15}$$

$$r_{SB} = \frac{0.3}{1.15}$$

$$r_{xy} = 0.26$$

The measure did not have good reliability in my sample!