

CAEP STANDARD ALIGNMENT: R5/RA5 Quality Assurance System & Continuous Improvement

CAEP COMPONENT ALIGNMENT: R5.2: Data Quality &RA5.2: Data Quality

Before beginning the process of validity or reliability, carefully read the following two documents available on the [Content Validity/Reliability Procedures webpage](https://www.marshall.edu/coepd-assessment/content-validity/content-validity-procedure/).

[<https://www.marshall.edu/coepd-assessment/content-validity/content-validity-procedure/>]

- CAEP Criteria for Evaluation of EPP-Created Assessments & Surveys
- CAEP Criteria for Evaluation of EPP-Created Surveys

Introduction:

The College of Education and Professional Development (COEPD) at Marshall University has established a procedure for content validity and reliability for all Education Preparation Provider (EPP) Created Assessments and Surveys, including key assessments, performance tasks, clinical evaluations, and national board-certified exams. The EPP adopted the procedure to evaluate its assessments in Spring 2022. The content validity and reliability procedures are used by both initial- and advanced-level programs. Procedures follow guidelines outlined in the CAEP Evaluation Frameworks for EPP-Created Assessments and Surveys to design, pilot, and judge the adequacy of the assessments created by the EPP.

The purpose of the content validity and reliability procedures is to provide guidance on collecting evidence and to document the adequate technical quality of assessment instruments and rubrics used to evaluate candidates in the COEPD.

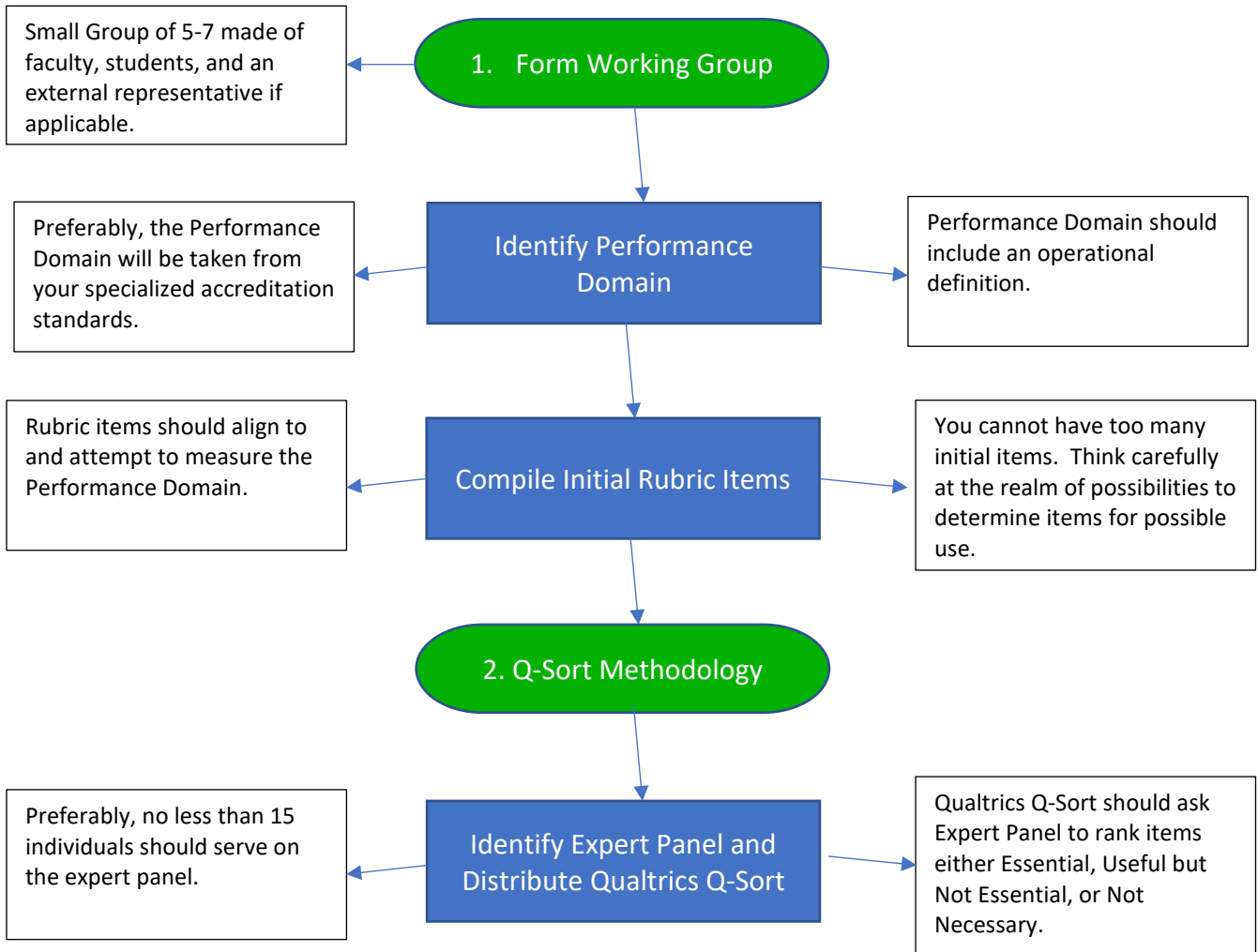
What is Validity?

Cronbach (1971) described validation as the process by which a test developer or test user collects evidence to support the types of inferences that are to be drawn from test scores. Crocker and Algina (2008) stated that the purpose of a content validation study is to “assess whether the items adequately represent a performance domain or construct of specific interest” (pg. 218). At minimum, content validation entails the following steps:

1. Defining the performance domain of interest;
2. Selecting a panel of qualified experts in the content domain;
3. Providing a structured framework for the process of matching items to the performance domain;
4. Collecting and summarizing the data from the matching process.

VALIDITY PROCEDURES

Example: EPP-Created Rubric



COMMITMENT TO STUDENTS: The creation of a learning environment and community to promote successful teaching and learning. Advanced candidates of the COEPD shall:

Example of a Pick, Group, and Rank Question Type in Qualtrics. Example provides Performance Domain (Commitment to Students), Operational Definition, and all items for grouping.

Items	Essential
Respect the rights of stakeholders.	Essential
Collaborate with stakeholders to improve learning.	
Promote evidence-based practices to meet the needs of learners.	
Use appropriate data to inform professional decisions.	Useful But Not Essential
Promote a positive learning environment.	
Adapt to novel or unexpected situations.	Not Necessary
Promote change to better meet the needs of students and their families.	
Encourage a climate of respect and self-responsibility.	

From the Qualtrics data report, conduct a CVR for each item.

3. Lawshe's Content Validity Ratio (CVR)

$CVR = (ne - n/2) / (n/2)$
 Ne= number of raters indicating item as essential
 N= total Q-sort members

Example of Using Microsoft Excel to find the CVR.

Example of Using Excel to Find CVR

	A	B	C	D
1	CVR FORMULA			
2	$CVR = (ne - N/2) / (N/2)$			
3				
4	Item	ne	n	CVR
5	Respects the Rights of Stakeholders	20	32	0.250

Identify Items Kept Using CVR and CVR Critical Value

Locate the CVR Critical Value on the COEPD Assessment [Content Validity Procedures Webpage](#). View your Panel Size and Obtain the required CVR Critical Value.

In the Chart below, you will see the CVR for all items. For the example, I used 30 participants as the N to get a CVR Critical Value of .333 in the chart to the left. For the chart items, you would only retain the items that are .333 or greater (as indicated in green). Those are the items we would keep in this example.

Panel Size	N _{Essential} (Minimum Number of Experts Required to Agree an Item Essential for Inclusion)	Proportion Agreeing Essential	CVR _{critical}
5	5	1	1.00
6	6	1	1.00
7	7	1	1.00
8	7	.875	.750
9	8	.889	.778
10	9	.900	.800
11	9	.818	.636
12	10	.833	.667
13	10	.769	.538
14	11	.786	.571
15	12	.800	.600
16	12	.750	.500
17	13	.765	.529
18	13	.722	.444
19	14	.737	.474
20	15	.750	.500
21	15	.714	.429
22	16	.727	.455
23	16	.696	.391
24	17	.708	.417
25	18	.720	.440
26	18	.692	.385
27	19	.704	.407
28	19	.679	.357
29	20	.690	.379
30	20	.667	.333
31	21	.677	.355
32	22	.688	.375
33	22	.667	.333
34	23	.676	.353
35	23	.657	.314
36	24	.667	.333
37	24	.649	.297
38	25	.658	.316
39	26	.667	.333
40	26	.650	.300

$CVI = (ne - N/2) / (N/2)$
 ne= Number of Panelists indicating "Essential"
 N= Total Number of Respondents

COMMITMENT TO STUDENTS (Minimum .33)

ITEM	ne	N	CVR
Respects the Rights of Stakeholders	20	32	.250
Collaborate with Stakeholders to Improve Learning	21	33	.313
Promote Evidence-Based Practices to Meet the Needs of Learners	30	32	.875
Use Appropriate Data to Inform Professional Decisions	29	31	.813
Promote a Positive Learning Environment	29	32	.813
Adapt to Novel or Unexpected Situations	10	30	-.375
Promote Change to Better Meet the Needs of Students and their Families	18	31	.125

Simplified Table of CVR critical including the Number of Experts Required to Agree an Item Essential.

4. Create or Review Rubric

If you are reviewing an existing rubric, please begin with Step 1 above to review and modify rubric items.

Using the identified items,
create your evaluation rubric.

Using the same group of individuals as before, prepare an evaluation rubric while considering the following: 1) Basis for judging candidate performance is well defined, 2) performance levels are qualitatively defined by specific criteria aligned with indicators, 3) performance levels represent a developmental sequence, 4) feedback provided to candidates is actionable, and 5) performance levels are defined in actionable, performance-based, or observable behavior terms.

5. Expert Panel Distribution

Distribute the Assignment, Assignment Instructions, and the Evaluation Rubric to the Expert Panel used with the Q-Sort. Looking for Construct Validity, the panel members will rate the representativeness and clarity of each item and item descriptor as it relates to the overall construct. Provide each expert with either a paper copy response form or a form created in Qualtrics. The example below is using Qualtrics to collect responses. A template for paper copies can be found the COEPD Resources Microsoft Team under Assessment.

Q3

DISPOSITION	MEETS EXPECTATIONS	DEVELOPING
Promote Evidence-Based Practices to Meet the Needs of Learners	Candidate evaluates, or considers the results of a current evaluation of, the needs of the learner and utilizes evidence-based practices that are aligned with those needs.	Candidate fails to consider re: of a current evaluation of t learner's needs OR the candi fails to utilize evidence-basi practices aligned to the need the learner.

Please select how Representative and Clear the item is in measuring the Construct.

Construct: Commitment to Students
Definition: The creation of a learning environment and community to promote successful teaching and learning. Advanced-level candidates of the COEPD shall:
Item: Promote Evidence-Based Practices to Meet the Needs of Learners:Right

	Item is Representative of the Construct	Item Needs Minor Revisions to be Representative of the Construct	Item Needs Major Revisions to be Representative of the Construct	Item is Not Representative of the Construct
Representativeness Promote Evidence-Based Practices to Meet the Needs of Learners:Right	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Clarity Promote Evidence-Based Practices to Meet the Needs of Learners:Right	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please note that the rubric performance levels and performance descriptors are above each item question for easy viewing by the panel members.

6. Content Validity Index

From the data obtained from the Expert Panel distribution, complete the Content Validity Index (CVI).

CVI = No. of Experts who rated the item (representative/clear) a 3 or 4 divided by the number of total experts.

$$\text{CVI} = \frac{\text{Number of experts who rated the item as 3 or 4}}{\text{Number of total experts}}$$

CVI of .80 or above is acceptable

$$\text{CVI} = \frac{\text{Number of Experts who rated item as 3 or 4}}{\text{Total Number of Experts}}$$

Retain CVI \geq .80

COMMITMENT TO STUDENTS: The creation of a learning environment and community to promote successful teaching and learning. Advanced-level candidates of the COEPD shall:

Promoted Evidence-Based Practices to Meet the Needs of Learners						
	# Rated as 4	# Rated as 3	# Rated as 2	# Rated as 1	Total Ratings	CVI
Representative	9	1			10	1
Clear	11	2			13	1

Number of Experts who rated item as 3 or 4

Total Number of Experts

RELIABILITY PROCEDURES

7. Obtain Cronbach's Alpha

- The COEPD will assess the reliability of an assessment, or the confidence that a test score earned is a good representation of actual knowledge of the content.
- To determine if an assessment item is reliable, the EPP will use Cronbach's Alpha, the most common measure of reliability.
- **Cronbach's Alpha** is a measure used to assess the reliability, or internal consistency, of a set of scale or test items and indicates whether an item measures the same construct.

Use SPSS Software to determine Cronbach's Alpha

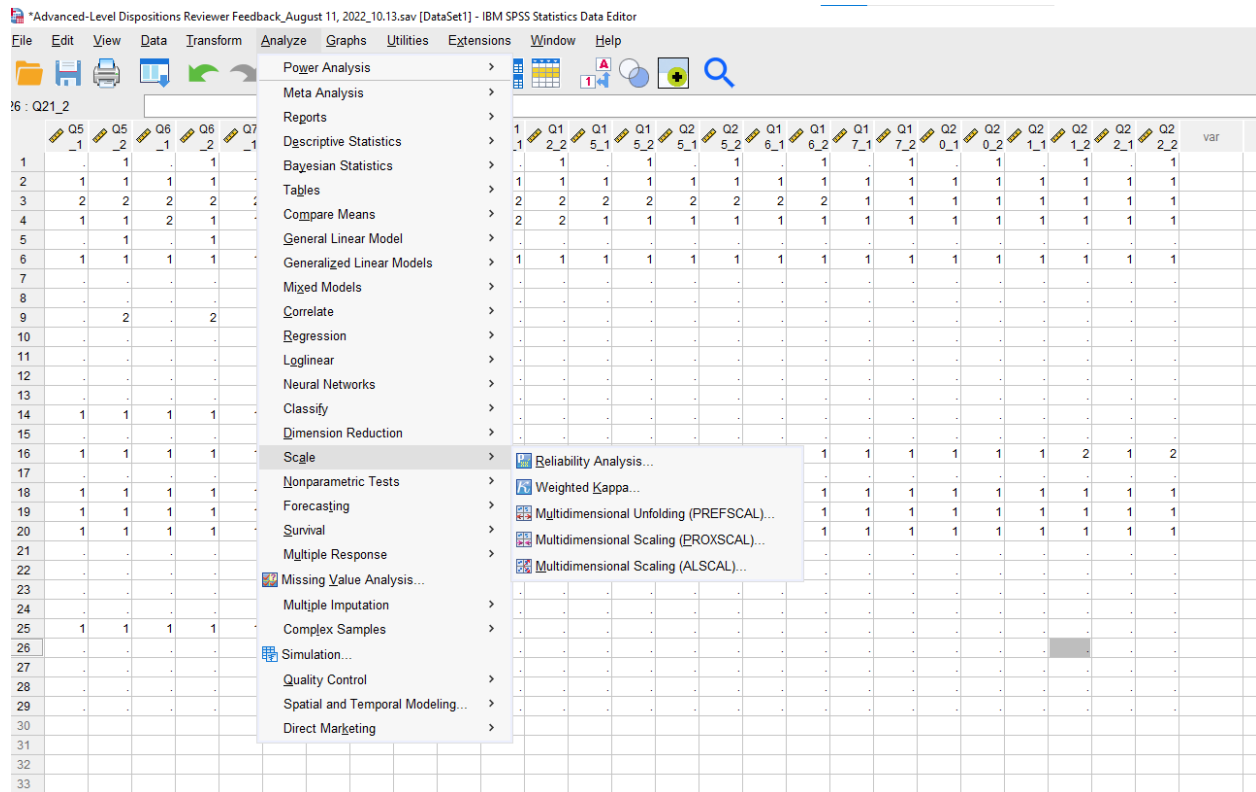
1. Download your data from Qualtrics by going selecting the Data & Analysis Tab > Export & Input > Export Data > SPSS > Download
 2. Open the data set in SPSS
- You may delete all columns except the questions responses (below image)

*Advanced-Level Dispositions Reviewer Feedback_August 11, 2022_10.13.sav [DataSet1] - IBM SPSS Statistics Data Editor

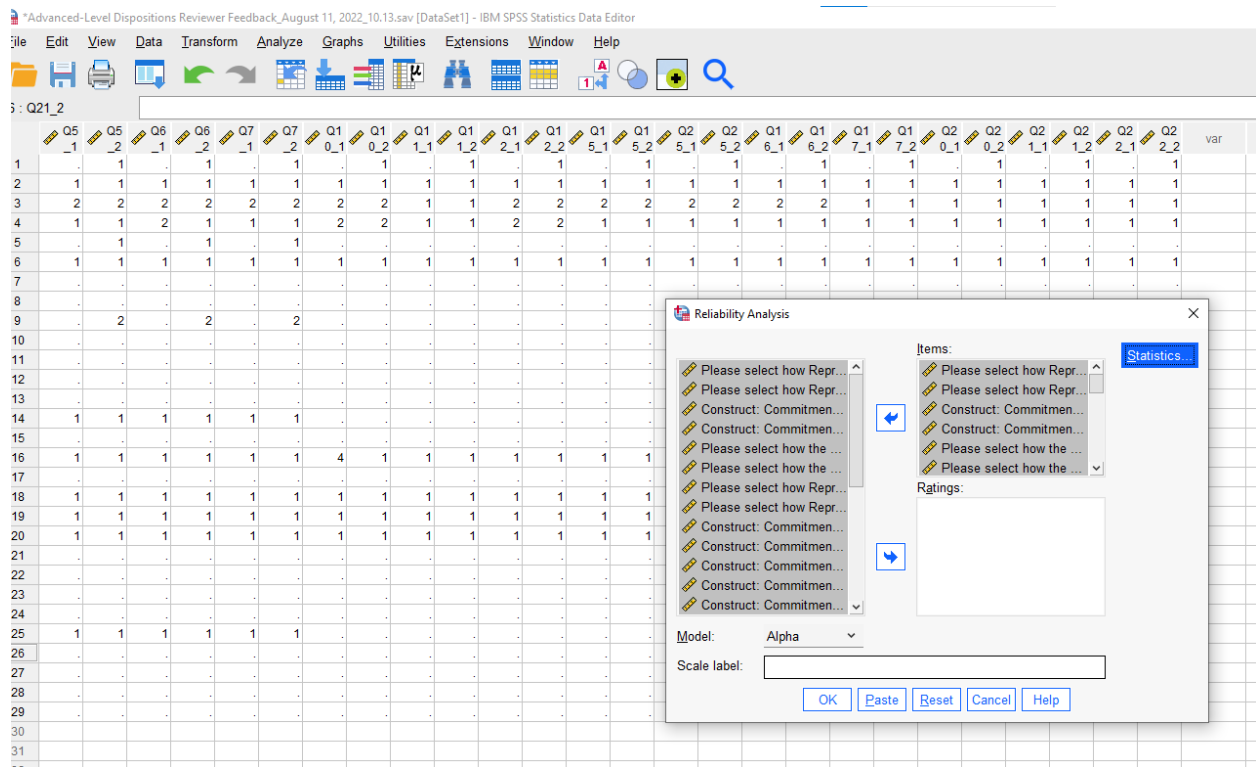
	Q5_1	Q5_2	Q6_1	Q6_2	Q7_1	Q7_2	Q1_0_1	Q1_0_2	Q1_1_1	Q1_1_2	Q1_2_1	Q1_2_2	Q1_5_1	Q1_5_2	Q2_5_1	Q2_5_2	Q1_6_1	Q1_6_2	Q1_7_1	Q1_7_2	Q2_0_1	Q2_0_2	Q2_1_1	Q2_1_2	Q2_2_1	Q2_2_2	
1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
3	2	2	2	2	2	2	2	2	1	1	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	
4	1	1	2	1	1	1	2	2	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
5	.	1	.	1	.	1	
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
7	
8	
9	.	2	.	2	.	2	
10	
11	
12	
13	
14	1	1	1	1	1	1	
15	
16	1	1	1	1	1	1	4	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	2	1	2	
17
18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
21
22
23
24
25	1	1	1	1	1	1	



3. Select Analyze > Scale > Reliability Analysis (below image)



4. Move question items to the Items box and then select Statistics (below image)



5. Select Item, Scale, Scale if Item Deleted, and Correlations (below image)
 6. Select Continue, then Submit

Advanced-Level Dispositions Reviewer Feedback_August 11, 2022_10.13.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

26 : Q21_2

	Q5_1	Q5_2	Q6_1	Q6_2	Q7_1	Q7_2	Q1_0_1	Q1_0_2	Q1_1_1	Q1_1_2	Q1_2_1	Q1_2_2	Q1_5_1	Q1_5_2	Q2_5_1	Q2_5_2	Q1_6_1	Q1_6_2	Q1_7_1	Q1_7_2	Q2_0_1	Q2_0_2	Q2_1_1	Q2_1_2	Q2_2_1	Q2_2_2	var	
1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	1	
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
3	2	2	2	2	2	2	2	2	1	1	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	
4	1	1	2	1	1	1	2	2	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
5	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	.	1	1	
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
7	1
8	1	
9	.	2	.	2	.	2	.	2	.	2	.	2	.	2	.	2	.	2	.	2	.	2	.	2	.	2	1	
10	1	
11	1	
12	1	
13	1	
14	1	1	1	1	1	1	1	
15	1	
16	1	1	1	1	1	1	1	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
17	1	
18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
21	1	
22	1	
23	1	
24	1	
25	1	1	1	1	1	1	1	
26	1	
27	1	
28	1	
29	1	
30	1	
31	1	
32	1	
33	1	
34	1	
35	1	
36	1	
37	1	

Reliability Analysis: Statistics

Descriptives for

- Item
- Scale
- Scale if item deleted

Inter-Item

- Correlations
- Covariances

Summaries

- Means
- Variances
- Covariances
- Correlations

ANOVA Table

- None
- E test
- Friedman chi-square
- Cochran chi-square

Interrater Agreement: Fleiss' Kappa

- Display agreement on individual categories
- Ignore string cases
- String category labels are displayed in uppercase

Asymptotic significance level (%): 95

Missing

- Exclude both user-missing and system missing values
- User-missing values are treated as valid

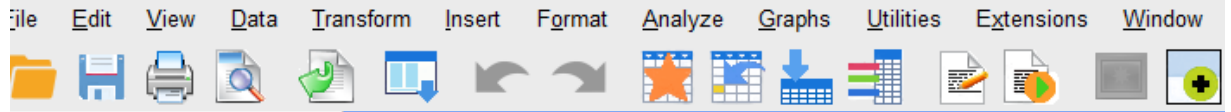
- Hotelling's T-square
- Tukey's test of additivity
- Intraclass correlation coefficient

Model: Two-Way Mixed Type: Consistency

Confidence interval: 95 % Test value: 0

Continue Cancel Help

7. View Cronbach's Alpha. A score near 1.00 shows high reliability. A score closer to 0 shows low reliability. As a rule of thumb, if the reliability of items is .80 or above, the assessment has very good reliability. If the reliability of items is .50 or below, the assessment would not be considered reliable. View image on next page to view Cronbach's Alpha under Reliability Statistics.



Output

- Reliability
 - Title
 - Notes
 - Warnings
 - Scale: ALL VARIAB
 - Title
 - Case Process
 - Reliability Stat
 - Item Statistics
 - Inter-Item Con
 - Item-Total Sta
 - Scale Statistic

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.938	.958	18

Item Statistics

	Mean	Std. Deviation	N
Please select how Representative and Clear each Disposition is in Measuring the Construct. Construct: Commitment to Students Definition: The creation of a learning environment and community to promote successful teaching and learning. Advanced-level candidates of the COEPD shall: Disposition: Promote Evidence-Based Practices to Meet the Needs of Learners - Representativeness Promote Evidence-Based Practices to Meet the Needs of Le	1.13	.354	8
Please select how Representative and Clear each Disposition is in Measuring the Construct. Construct: Commitment to Students Definition: The creation of a learning environment and community to promote successful teaching and learning. Advanced-level candidates of the COEPD shall: Disposition: Promote Evidence-Based Practices to Meet the Needs of Learners - Clarity Promote Evidence-Based Practices to Meet the Needs of Le	1.13	.354	8

8. Pilot Your Rubric

Distribute Assessment and Use Rubric to Score Submissions

9. Calibrate “Norm” the Rubric

- Calibrating the rubric simply means that the rubric items and scales are clear enough to lead to a consensus among faculty who are grading students work.
 - This is important when various faculty members are evaluating assessments among multiple course sections.
- Calibrating helps to ensure that the rubric uses is consistent, and for the most part, free from bias.
- Helps to ensure that assessment data reflects the expectations of the program, not an individual faculty member or instructor.

Steps for Norming Rubric

1. A facilitator prepares materials for scorers to begin calibrating the assessment rubric. Materials include the assessment instructions, the grading rubric, and a student artifact.
2. Using the rubric, scorers read the assignment instructions, view the student artifact, and score the artifact using the assessment rubric. Scorers should note words and phrases in the performance descriptors that best describe the quality of work.
3. One at a time, scorers share scores for each rubric category while a recorder completes a group score sheet. Do not provide an explanation for the score at this point. Once all scores are shared and recorded by the facilitator, the scorers will discuss differences in the scores, where the differences occurred, and why scorers may have evaluated the artifact differently.
4. Scorers justify their evaluation by pointing to specific language in the rubric and evidence in the student artifact. Discuss each piece of student work and resolve issues that may be present because of rubric language, or the evidence provided in the student artifact. Scorer consensus should be reached.

Examples

	Distinguished (4)	Proficient (3)	Basic (2)	Unsatisfactory (1)
Awareness of the impact of motivation and schema on the learning process	Lesson begins with a clear description of method to activate prior knowledge and build schema toward the learning objective(s); Includes questions that generate intellectual curiosity and engage students with the lesson content through higher order critical thinking and problem solving	Lesson begins with a clear description of method to activate prior knowledge and build schema toward the learning objective(s); Includes questions that generate intellectual curiosity and engage students with the lesson content	Lesson begins with a clear description of method to activate prior knowledge and build schema toward the learning objective(s); Includes questions, but questions lack student engagement and/or stimulation of intellectual curiosity	Lesson begins with a method to introduce the lesson content, but introduction does not activate prior knowledge and/or build schema toward the learning objective(s); Questions are not included or do not engage students in intellectual curiosity

SCENARIO: 5 Faculty members are convened to norm a rubric. A component of the rubric is "Awareness of the Impact of Motivation and Schema on the Learning Process."

The faculty members are given the exact student artifact and asked to score the artifact using this rubric.

The Facilitator has asked that each evaluator provide their score for the artifact and Has recorded the following scores...

	A	B	C	D	E	F
		R1	R2	R3	R4	R5
Awareness of the impact of motivation and schema on the learning process		4	4	4	2	4

How should the facilitator proceed?

The facilitator should begin by asking Rater 4 (R4) to their reasoning for giving the rating of "2." The rating could be for numerous reasons (unclear verbiage, misinterpretation, etc. The facilitator should ask for reasons behind the rating of "4."

From that point, R4 may be willing to increase their score.

10. Cohen's Kappa

- Cohen's Kappa assess the extent to which two raters agree.
- Range from -1 to +1
- Cohen suggested the Kappa result be interpreted as follows:
 - values ≤ 0 as indicating no agreement
 - 0.01–0.20 as none to slight
 - 0.21–0.40 as fair
 - 0.41–0.60 as moderate
 - 0.61–0.80 as substantial
 - and **0.81–1.00 as almost perfect agreement**

Steps for Cohen's Kappa

Artifact Rater Scores

	Rater1	Rater2	VAR
1	4.00	4.00	
2	4.00	4.00	
3	4.00	3.00	
4	2.00	2.00	
5	3.00	3.00	
6	3.00	3.00	
7	4.00	4.00	
8	3.00	3.00	
9	4.00	3.00	
10	3.00	3.00	
11	3.00	2.00	
12	4.00	4.00	
13	4.00	4.00	
14	2.00	2.00	
15	3.00	3.00	
16			

- Artifacts should be randomly assigned to raters.
- Artifact 1: Kim and Paula
- Artifact 2: Sandra and Lisa
- Artifact 3: Tom and Feon
- Artifact 4: Feon and Sandra
- Artifact 5: Lisa and Tom
- Artifact 6: Paula and Lanai
- And so on....
- Keep a master list of who the raters are for each artifact, But don't share that list with the raters.

Artifact Rater Scores

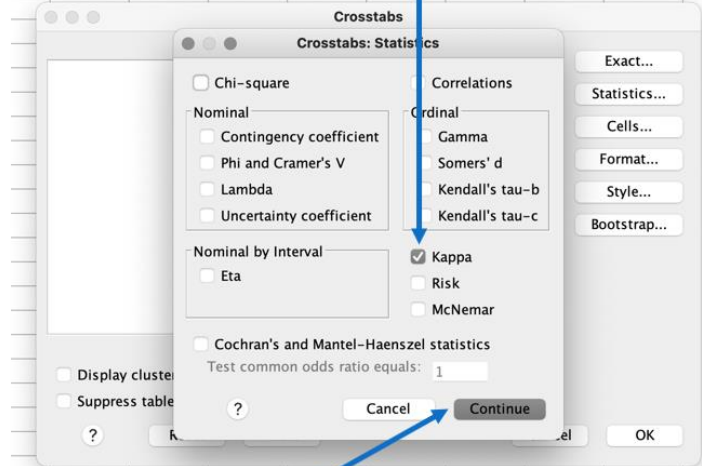
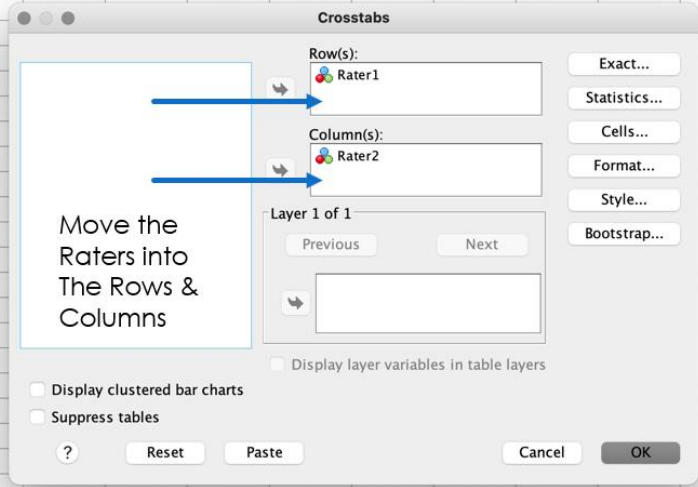
	Rater1	Rater2	V&F
1	4.00	4.00	
2	4.00	4.00	
3	4.00	3.00	
4	2.00	2.00	
5	3.00	3.00	
6	3.00	3.00	
7	4.00	4.00	
8	3.00	3.00	
9	4.00	3.00	
10	3.00	3.00	
11	3.00	2.00	
12	4.00	4.00	
13	4.00	4.00	
14	2.00	2.00	
15	3.00	3.00	
16			



Select Statistics

Move the Raters into The Rows & Columns

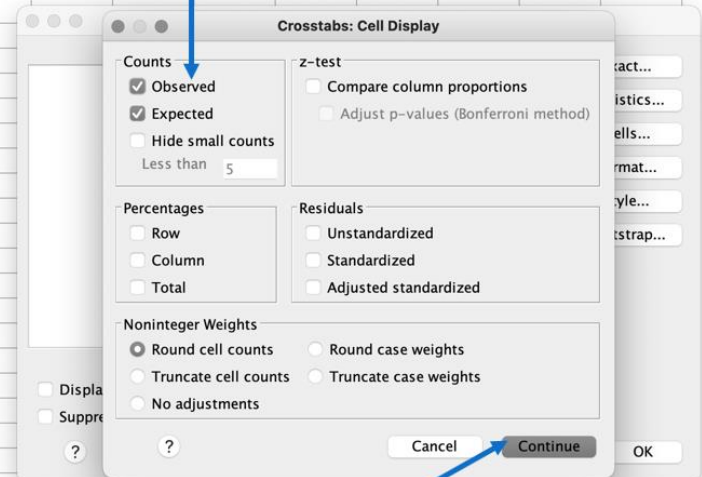
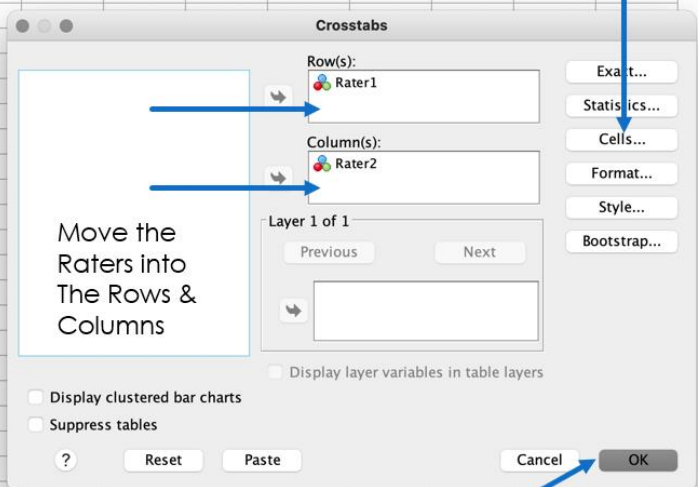
Select Kappa



Select Continue

Select Cells

Select Observed & Expected



Select Continue

Select Continue

➔ Crosstabs

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
Rater1 * Rater2	15	100.0%	0	0.0%	15	100.0%

Rater1 * Rater2 Crosstabulation

Rater1			Rater2			Total
			Basic	Proficient	Distinguished	
Basic	Count		2	0	0	2
	Expected Count		.4	.9	.7	2.0
Proficient	Count		1	5	0	6
	Expected Count		1.2	2.8	2.0	6.0
Distinguished	Count		0	2	5	7
	Expected Count		1.4	3.3	2.3	7.0
Total	Count		3	7	5	15
	Expected Count		3.0	7.0	5.0	15.0

Cohen's Kappa = .683
Substantial Agreement

Symmetric Measures

		Value	Asymptotic Standard Error ^a	Approximate T ^b	Approximate Significance
Measure of Agreement	Kappa	.683	.162	3.645	.000
N of Valid Cases		15			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

You've now established that your EPP-Created Assessments are Valid and Reliable!