BUROS
CENTER FOR TESTING

UNIVERSITY OF Nebraska Lincoln
COLLEGE OF EDUCATION & HUMAN SCIENCES

CCCCO Assessment Training for Local Colleges: Day 1
October 19, 2022

Jessica L. Jonson, PhD
Maria Elena Oliveri, PhD

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

A11Y 11/23/22

Training Objectives

Requirements & expectations for local college assessment applications

- CCC Standards for Assessment Instrument Review: English as a Second Language (2022)

Criteria for review:

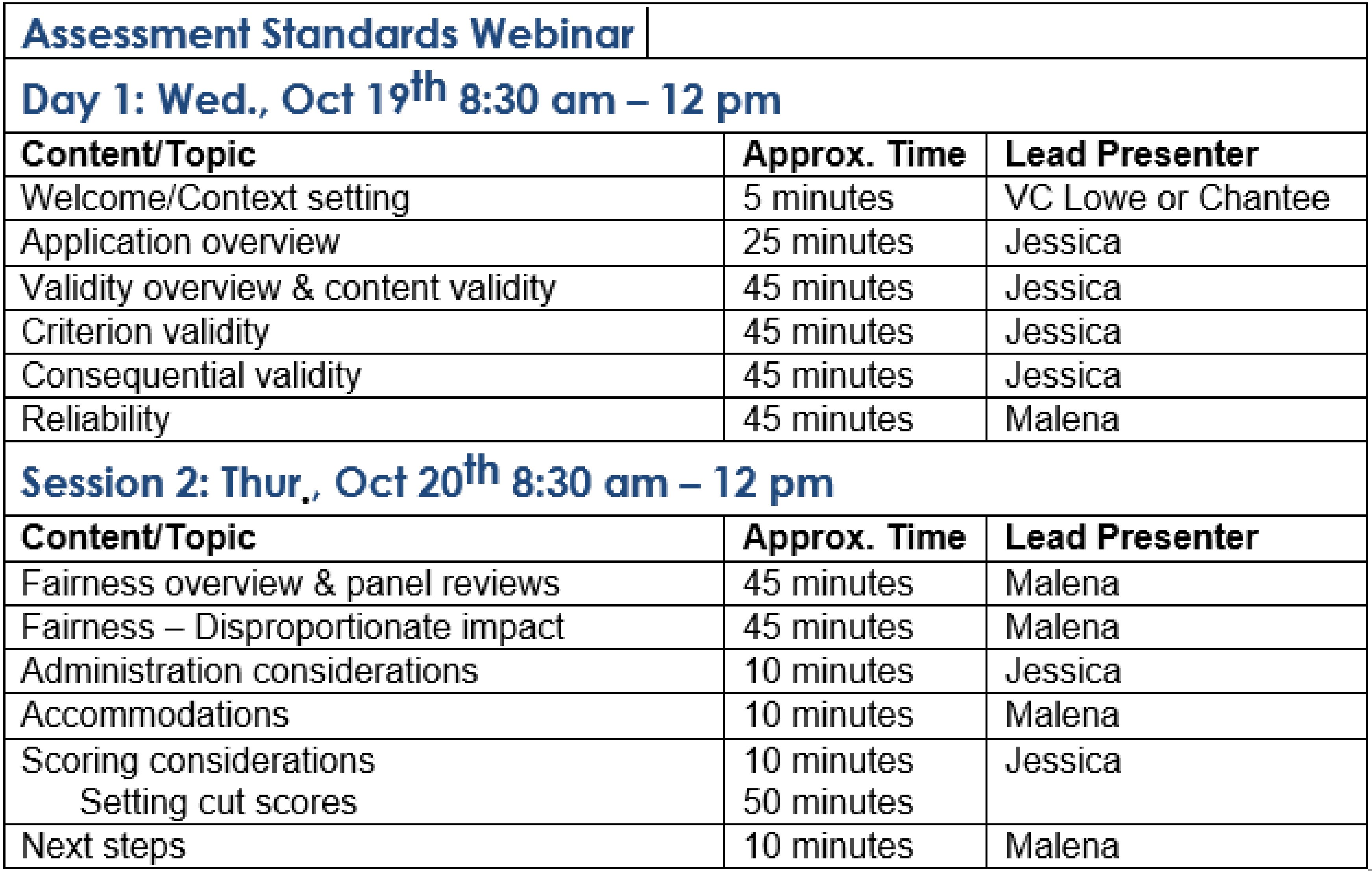
- Fairness
- Validity
- Reliability
- Accommodations
- Administration and Scoring

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Agenda: Training Sessions



Assessment Standards Webinar		
Day 1: Wed., Oct 19th 8:30 am – 12 pm		
Content/Topic	**Approx. Time**	**Lead Presenter**
Welcome/Context setting	5 minutes	VC Lowe or Chantee
Application overview	25 minutes	Jessica
Validity overview & content validity	45 minutes	Jessica
Criterion validity	45 minutes	Jessica
Consequential validity	45 minutes	Jessica
Reliability	45 minutes	Malena
Session 2: Thur., Oct 20th 8:30 am – 12 pm		
Content/Topic	**Approx. Time**	**Lead Presenter**
Fairness overview & panel reviews	45 minutes	Malena
Fairness – Disproportionate impact	45 minutes	Malena
Administration considerations	10 minutes	Jessica
Accommodations	10 minutes	Malena
Scoring considerations Setting cut scores	10 minutes 50 minutes	Jessica
Next steps	10 minutes	Malena

rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Training Structure

Conceptual overview of each criteria

Outline requirements/expectations for criteria from CCC Standards (2022)

Provide applied examples to provide further guidance

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln
COLLEGE OF EDUCATION
& HUMAN SCIENCES

Application Overview

Application Types
Application Statuses
Approval Decisions Types

BUROS
CENTER FOR TESTING

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Application Types

SECOND-PARTY: TEST/ASSESSMENT DEVELOPED AND MAINTAINED BY EXTERNAL VENDOR (COMMERCIALLY-AVAILABLE)

LOCALLY DEVELOPED: TEST/ASSESSMENT DEVELOPED LOCALLY FOR USE BY A SINGLE COLLEGE OR MULTICOLLEGE DISTRICT

LOCALLY MANAGED: EXISTING TEST/ASSESSMENT USED BY A LOCAL COLLEGE OR MULTICOLLEGE DISTRICT.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Application Status

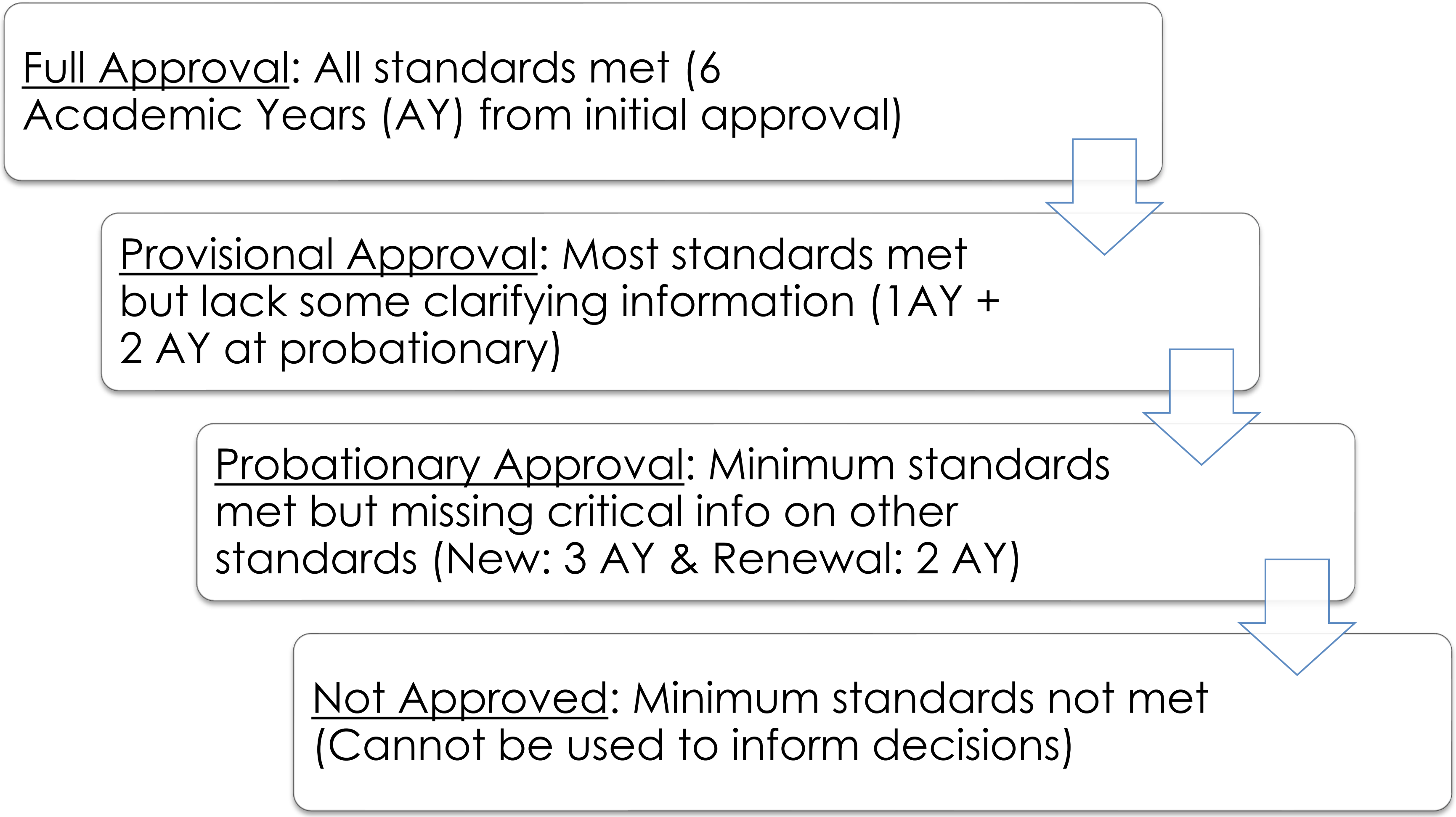
NEW

RENEWAL

RESUBMISSION

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Approval Decisions (Appendix D 2022 CCC Standards)



Full Approval: All standards met (6 Academic Years (AY) from initial approval)

Provisional Approval: Most standards met but lack some clarifying information (1AY + 2 AY at probationary)

Probationary Approval: Minimum standards met but missing critical info on other standards (New: 3 AY & Renewal: 2 AY)

Not Approved: Minimum standards not met (Cannot be used to inform decisions)

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln
COLLEGE OF EDUCATION & HUMAN SCIENCES

Types Of Application Status

NEW APPLICATION

- Test not previously reviewed
- Test previously approved but not renewed by 6 years
- Substantive changes to previously reviewed test

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln
COLLEGE OF EDUCATION & HUMAN SCIENCES

Types Of Application Status

RENEWAL APPLICATION

- Test previously approved in last 6 years
- No changes in test content, placement courses, or student demographics

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Types Of Application Status

RESUBMISSION

- Previously provisional/ probationary approvals seeking full approval
- Previously not approved seeking approval

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Minimum Standards By Application Status

NEW APPLICATIONS

- Fairness review
- Disproportionate impact study plan
- Content validation
- Criterion validation study plans
- Consequential validation study plan
- Reliability (at least one study)
- Accommodations plans
- Administration/Scoring documentation
- Cutscore setting study

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln
COLLEGE OF EDUCATION & HUMAN SCIENCES

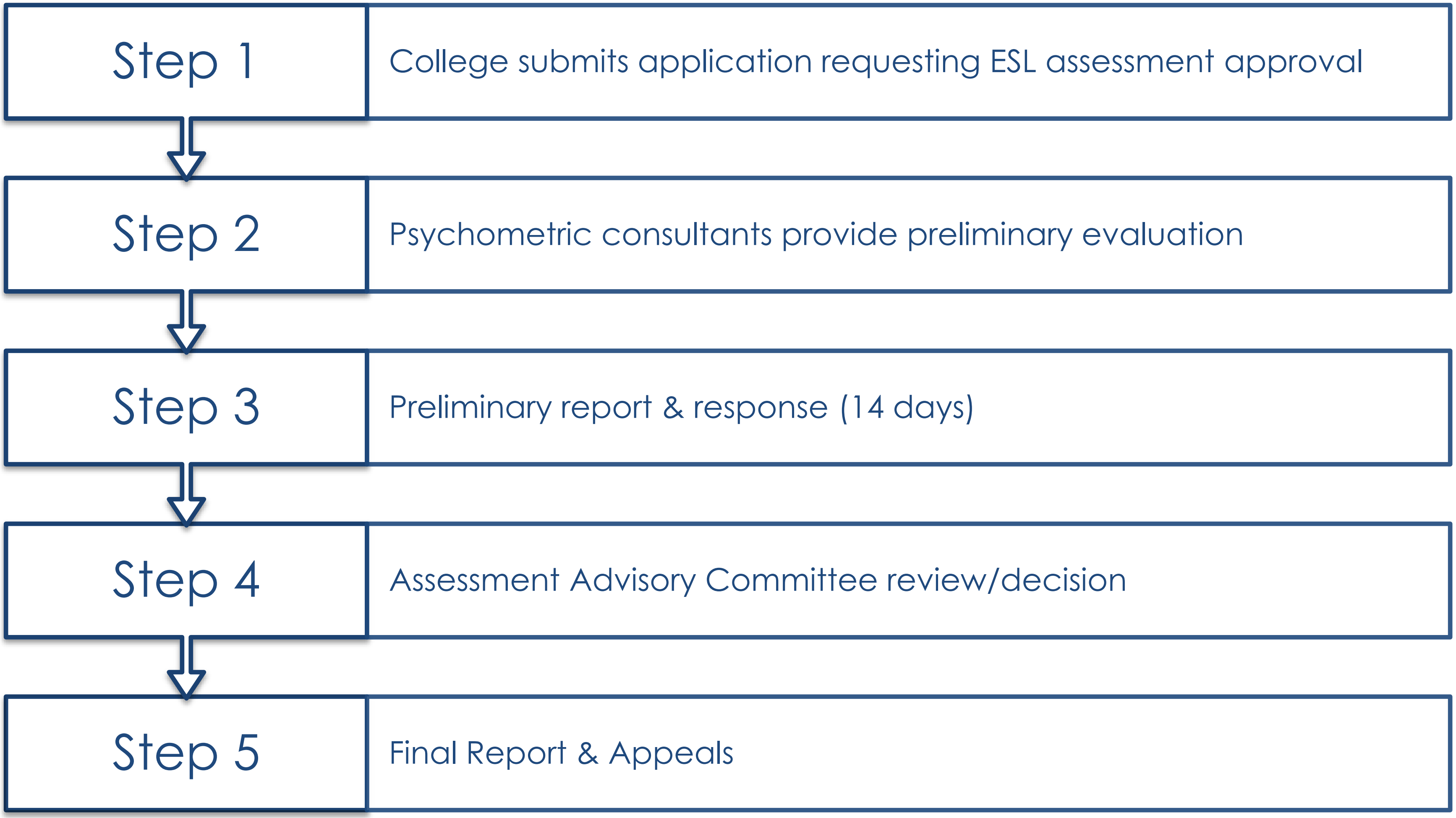
Minimum Standards By Application Status

RENEWAL APPLICATIONS

- Disproportionate impact study
- Two criterion validation studies
- Consequential validation study
- Reliability (at least one study)
- Accommodation documentation
- Administration/Scoring documentation
- Cutscore adjustments

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Application Review Cycle (Appendix B 2022 CCC Standards)



Step 1 College submits application requesting ESL assessment approval

Step 2 Psychometric consultants provide preliminary evaluation

Step 3 Preliminary report & response (14 days)

Step 4 Assessment Advisory Committee review/decision

Step 5 Final Report & Appeals

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Why Preliminary And Final Review?

Preliminary review: Psychometric review

Final review: Contextual review

Both are needed for a comprehensive review of quality, relevance, and appropriateness

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

QUESTIONS?

BUROS
CENTER FOR TESTING

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

BUROS
CENTER FOR TESTING

University of Nebraska Lincoln
COLLEGE OF EDUCATION & HUMAN SCIENCES

Validity Overview

Jessica L. Jonson, PhD

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

What Is Validity?

Validity is....

Extent to which evidence shows that an instrument or procedure appropriately measures the construct of interest (e.g., construct = English language proficiency) for a particular interpretation and use (e.g., course placement)

Construct = related knowledge, skills, or other attributes a test is intended to measure

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Validity: An Analogy

Provide a "snapshot not a film" of an individual's functioning that "describes a moment frozen in time, described from the viewpoint of the psychologist" (p. 637).

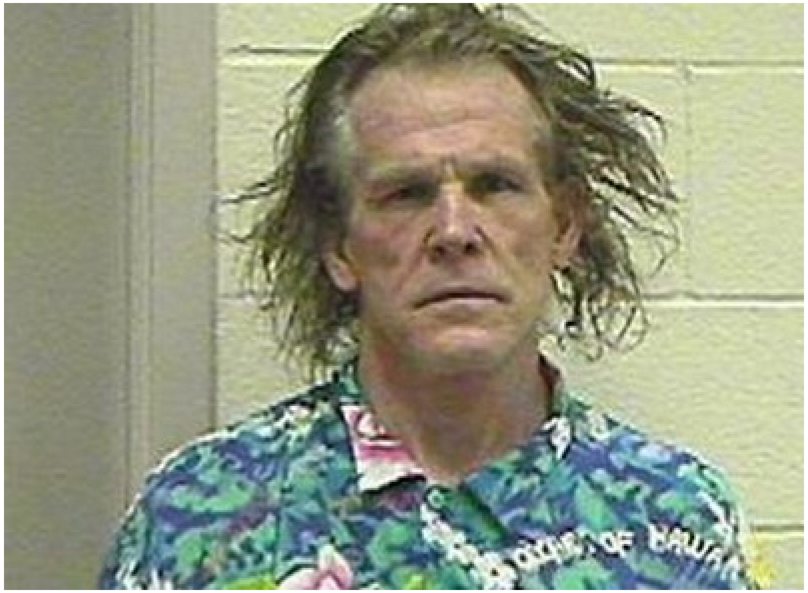
Cates, J. A. (1999). The art of assessment in psychology: Ethics, expertise, and validity. *Journal of Clinical Psychology, 55,* 631-641.

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | College of Education & Human Sciences

Validity: An Analogy



- "Good photos" represent subjects as they are.
- Even good photos are more appropriate for some purposes than others.
- Good tools (tests) and good techniques (test givers, test administration) essential for a "good picture" of test takers.

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Validity: What Is Being Validated?

Don't validate: instrument or its scores

- ESL test or Overall language score

Do validate: inferences for particular test uses or applications

- ESL placement decisions for students at CCC

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln
COLLEGE OF EDUCATION
& HUMAN SCIENCES

CCC Standards: Five Types Of Validity Evidence

Test content: Analysis test content and measured construct

Response processes: Analysis of individual responses

Internal structure: Relationship test items and test components

Relations with other variables: Test score inference and important external variables and generalize to new situations

Consequences of testing: Soundness of proposed interpretations and intended uses and unintended consequence

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Validity Evidence Expectations And Review

Evidence that supports the particular interpretation and use for which test will be used

Evidence described in detail and rationale for how supports interpretation/use

Is evidence sound and sufficient for the purposes in which will be used?

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

QUESTIONS?

BUROS
CENTER FOR TESTING

University of Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Content Validation

Jessica L. Jonson, PhD

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Content Validation: Overview

Provide sufficient evidence that the test content is relevant and representative of the construct of interest (e.g., ESL knowledge and skills)

1. What content is covered on the test?
2. Does the test content align with ESL course expectations?

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Content Validation: Documentation Requirements (p. 18-19)

1. Describe the test and the knowledge and skills it assesses

- Format of the test and how it was developed
- ESL competencies (KSAs) measured by the test (table of specifications/test blueprint)
- Scores reported
- Representative test form (e.g., items, prompts, tasks, scoring rubric)

2. Conduct an alignment study

- Align assessment content with entry-level skills required for each ESL course (including transfer-level composition).

3. Evaluate and conclude if the test is representative and relevant for course placement decisions.

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Content Validation: Submission Requirements

New submissions: Required

- Probationary:
  - content description
  - alignment study

Renewal submissions: only if changes to test or ESL curriculum

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Content Validation: Test Format And Form

- Selected response (e.g., multiple-choice)
  - Items
- Performance assessment (e.g. writing)
  - Prompts/Tasks
  - Scoring rubrics/algorithms
- Computer-adapted
  - Describe item bank
  - CAT algorithms (rules/restrictions)

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Content Validation:
Table of Specifications/Test Blueprint

Table of specifications/Test blueprint: List competencies measured and number of (or which) items measure each

Preferred practice: identify items measure different levels of ESL competency (beginning, intermediate, advanced)

If test format:

- Performance assessment: scoring rubrics provide descriptions of different levels of performance for a list of characteristics

- Item banks: specify not only how many items in the bank for each competency but also the number (or range) items each test take receives according to the item selection algorithm

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Examples for Different Test Items

- Objective test Items: require students to provide or select the correct response
  - Multiple-choice, true/false, short answer

- Subjective test items: permit students to provide an original answer
  - Essay, performance, problem-solving
  - Often scored using criteria or rubric

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Test Blueprint: Objective Test Item Example

General Competency	Knowledge/Skills	Number of items	Beginner	Intermediate	Advanced
Beginning Literacy/Phonics	1a. Ask for, give, follow, or clarify directions to a place or location, including reading signs	14	6	4	4
	1b. Identify different kinds of housing, areas of the home, and common household items	1	0	1	0
	1c. Interpret clock time	9	3	4	2
Vocabulary	2a. Understand or use appropriate language for informational purposes (e.g., to identify, describe, ask for information, state needs, agree or disagree)	8	4	2	2
	2b. Identify, evaluate and access schools and other learning resources	6	2	1	3
	2c. Identify safety measures that can prevent accidents and injuries	5	3	1	2

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Test Blueprint: Scoring Rubric Example

Criteria	Score 5-6 (English 1A)	Score 4 (ESL 151)	Score 2-3 (ESL 101)	Score 1 (ESL 100)
a. Response to text and prompt	Exhibits an insightful response to the text; effectively addresses all tasks	Exhibits and adequate response to the text; addresses most aspects of the task	May not respond adequately to the text; may ignore some aspects of the task	Demonstrates a failure to comprehend the tasks at hand
b. Organization, development and support	Is well organized and substantially developed with effective examples and evidence	Is unevenly organized and generally developed with some effective examples and evidence	May lack coherent structure and effective examples	Lacks any structure or development; may be inappropriately brief
c. Style (diction and syntax)	Makes sophisticated syntactic choices and uses precise diction	May lack syntactic variety and exhibit inexact diction	Often lacks precise word choice and syntactic variety	Lacks control of syntax and vocabulary
d. Writing conventions	Usually employs correct grammar, punctuation, and spelling	Contains a few grammar errors, but generally observes conventions	Contains errors that interfere with meaning	Contains numerous grammatical errors that interfere with meaning

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | College of Education & Human Sciences

Content Alignment Study: Individuals And Materials Needed

Who:

Faculty teaching ESL courses (but were not involved in the development of the test)

What:

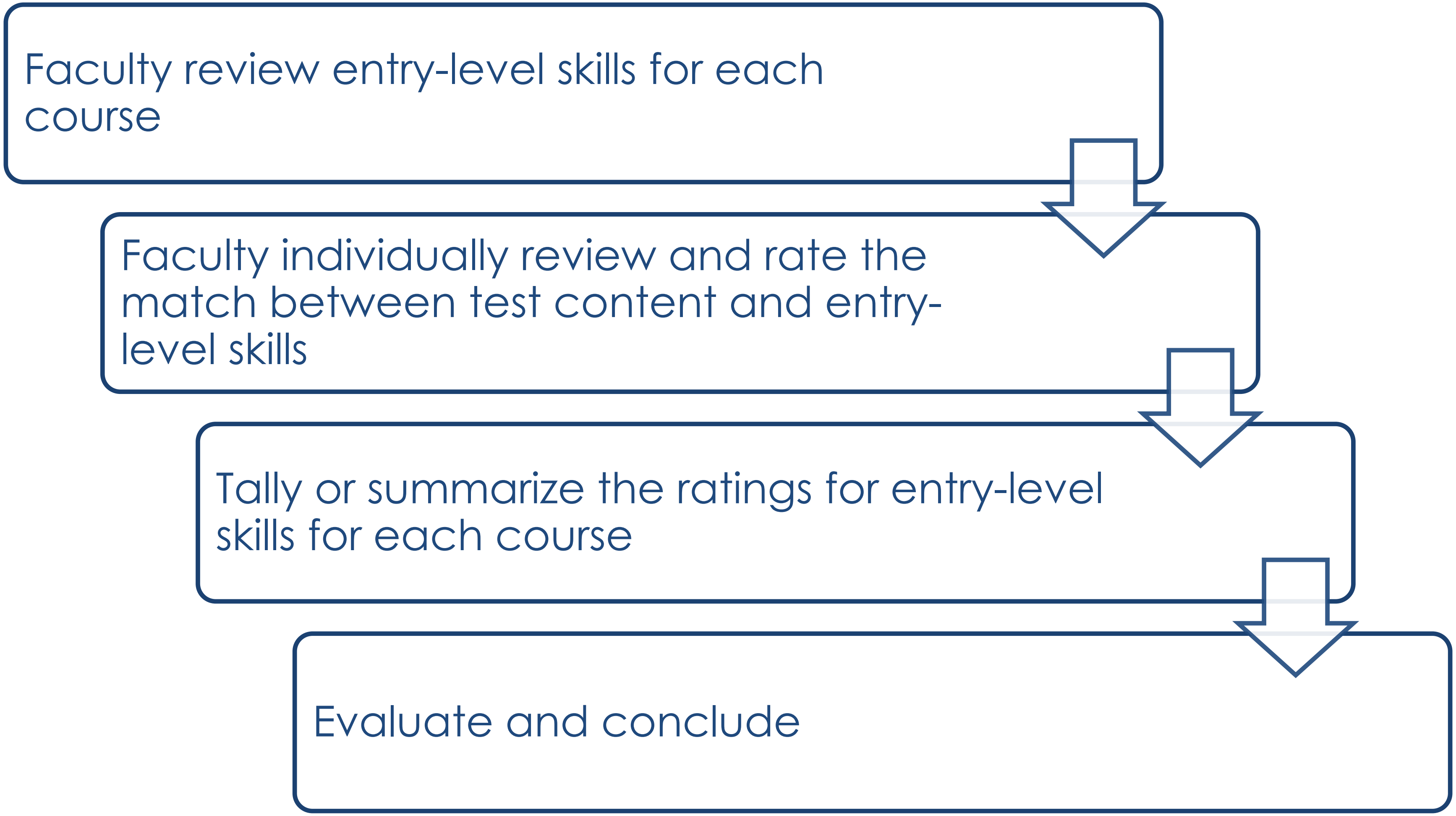
- Listing of entry-level skills needed for each ESL course and transfer-level composition
- Knowledge or skills measured by the test (blueprint/table of specifications)
- Test items, a representative sample of items (CAT), prompts/scoring rubric (Performance assessment)

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | College of Education & Human Sciences

Content Alignment Study: Process



Faculty review entry-level skills for each course

Faculty individually review and rate the match between test content and entry-level skills

Tally or summarize the ratings for entry-level skills for each course

Evaluate and conclude

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln
COLLEGE OF EDUCATION
& HUMAN SCIENCES

Content Alignment Study – Multiple Items Example

Test Item	ESL 1 (Entry-skills)			ESL 2 (Entry-skills)					Transfer Comp (Entry-skills)		
	1	2	3	1	2	3	4	5	1	2	3
1	Y			Y							
2		Y			Y					Y	
3											
4			Y			Y	Y				
5	Y		Y	Y		Y		Y	Y		Y
6		Y			Y					Y	
7											
8	Y	Y		Y	Y				Y	Y	
9											
10	Y										
Total	4	3	2	3	3	2	1	1	2	3	1

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln COLLEGE OF EDUCATION & HUMAN SCIENCES

Alignment Example: Scoring Rubric

	Congruent with…			
Scoring Rubric	ESL 100 entry skills	ESL 101 entry skills	ESL 151 entry skills	English 1A entry skills
Recommend ESL 100	5.0	4.0	2.0	1.0
Recommend ESL 101	3.4	5.0	4.0	3.6
Recommend ESL 151	2.6	3.6	5.0	4.0
Recommend English 1A	1.4	3.2	4.0	5.0

Scoring scale:

1= *no match* between the scoring rubric and the course entry skills
2= *little match* between the scoring rubric and the course entry skills
3= *moderate match* between the scoring rubric and the course entry skills
4= *good match* between the scoring rubric and the course entry skills
5= *strong match* between the scoring rubric and the course entry skills

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln
COLLEGE OF EDUCATION & HUMAN SCIENCES

Content Validation: Analysis Considerations

- Are all entry-level skills addressed?
- Are those entry-level skills sufficiently addressed (at least one item)?
  - If not, should new content be developed?
- Is there test content that does not align with entry-level skills?
  - If so, should it be removed?
- Is the test content representative and relevant for all ESL courses?
  - If not, which courses is it appropriate for and which is it not?
  - As a result, what is the local college's plan for using the test results in placement process?

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln COLLEGE OF EDUCATION & HUMAN SCIENCES

Content Validation: Common Errors/Omissions

- No description of faculty, details about process, or summary of results
- Describes overall skills measured instead of item-level review
- Results not summarized for each course
- No interpretation or conclusions
- For performance assessment, no review aligning rubric criteria and entry-level skills

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

QUESTIONS?

BUROS
CENTER FOR TESTING

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

BUROS
CENTER FOR TESTING

University of Nebraska Lincoln
COLLEGE OF EDUCATION & HUMAN SCIENCES

Criterion Validation Studies

Jessica L. Jonson, PhD

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Criterion Validation: Overview

Extent to which inference from a test score relates to alternate measure of construct or an outcomes measure (AKA criterion variable).

- Criterion variable collected at the same time as the test (concurrent) or in the future (predictive)

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln COLLEGE OF EDUCATION & HUMAN SCIENCES

Criterion Validation: Documentation Requirements (p. 20-21)

1. Describe the study sample
   - Demographically representative of ESL student population at local college (don’t forget cultural/linguistic groups)
   - Representation from all ESL proficiency levels and cohorts
   - Census or random sample not a convenience sample
   - Sufficient size (n=10 per group, n=30 overall): Gather over multiple years if population is small

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Criterion Validation: Documentation Requirements (p. 20-21)

2. Describe study methods
   - What, when, and how data was collected (recent data – last 3 years)
     - Test score and recommended placement level
     - Criterion variable: either score, recommended placement level, or both
     - Whether student initial course enrollment matched recommended placement by the test
   - Rationale for selected criterion variables. Two different criterion variables are required.

     1. One at the time of testing. Possibilities include:
     - student self-assessment of proficiency
     - other multiple measures used in placement decisions
     - test scores from another ESL proficiency measure

     2. One after initial enrollment. Possibilities include:
     - instructor assessment of proficiency
     - mid-term/final course grade
     - mid-term/final course exam score

   - How data was analyzed

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Criterion Validation: Documentation Requirements(p. 20-21)

3. Summarize the results and actions taken

- Provide a demographic representation of the study sample
- Provide descriptive statistics and distribution of test scores/placement levels and criterion scores/levels in the data set
- Report results for all courses in the ESL sequence and the transfer-level composition
- If correlation coefficients must be .35 or higher (or comparable effect size if alternate statistical analysis)
- When sample sizes permit, report results separately for cultural/linguistic groups (minimum n=10 per group)

4. Evaluate and conclude

- Based on the results, make recommendations about the use of the test scores for the placement decisions of students from different demographic groups and for specific course/proficiency levels

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

CCC Standards: Representing Cultural/Linguistic Groups

Representation of cultural/linguistic (C/L) groups that constitute 2% or more of your ESL student population

- If demographic data is unavailable, ask ESL faculty to identify the key C/L groups
- English language abilities can vary across C/L groups
- Minimum n=10 per C/L group (encourage census data when possible)
- Keep in mind: Language differences in cultural groups (e.g., Spanish speakers)

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | College of Education & Human Sciences

Criterion Validation: Submission Requirements

New submissions: Two studies are required

- Probationary: Detailed plan for conducting these studies

Renewal submissions: Two studies are required.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Criterion Validation: Types Of Comparisons

At the time of testing (concurrent)

Test variable	Criterion variable
Recommended course placement based on test score	Student self-assessment of placement based on survey of entry-level skills for courses
Recommended course placement based on test score	Recommended course placement based on another multiple measure used in the placement decision
Score from the test	Score from a test of another measure

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Irvine Valley – Example Student Self-Assessment

Review the levels below and choose the best level that describes your English skills. Use the sentence “Today, I believe I can…”. Choose ONLY 1 box.

Proficient	□ Write 5 to 7 page essays in academic English with little or no help □ Read college level texts in English, including a 300-page novel or nonfiction with little or no dictionary help. □ Fully understand a college lecture in English on academic topics such as Biology, History, or Sociology.
Advanced	□ Write 2 to 3 page essays in academic English with some help. □ Read short college- level texts in English, including a 200-page novel or nonfiction with dictionary help. □ Understand most of a college lecture in English on academic topics such as Biology, History, and Sociology
Low Intermediate	□ Write a group of sentences in English with help. □ Read a group of sentences in English, but sometimes I do not know all of the words. □ Understand a slow-paced conversation in English on a familiar topic or in a practical everyday situations such as shopping or the weather.
Low Beginner	□ Write some words and a couple of sentences. I know my English ABCs, numbers, and some words, but I need a lot of help. □ Read and understand some familiar words in simple sentences. □ Understand some simple questions, directions, or greetings.

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln
COLLEGE OF EDUCATION
& HUMAN SCIENCES

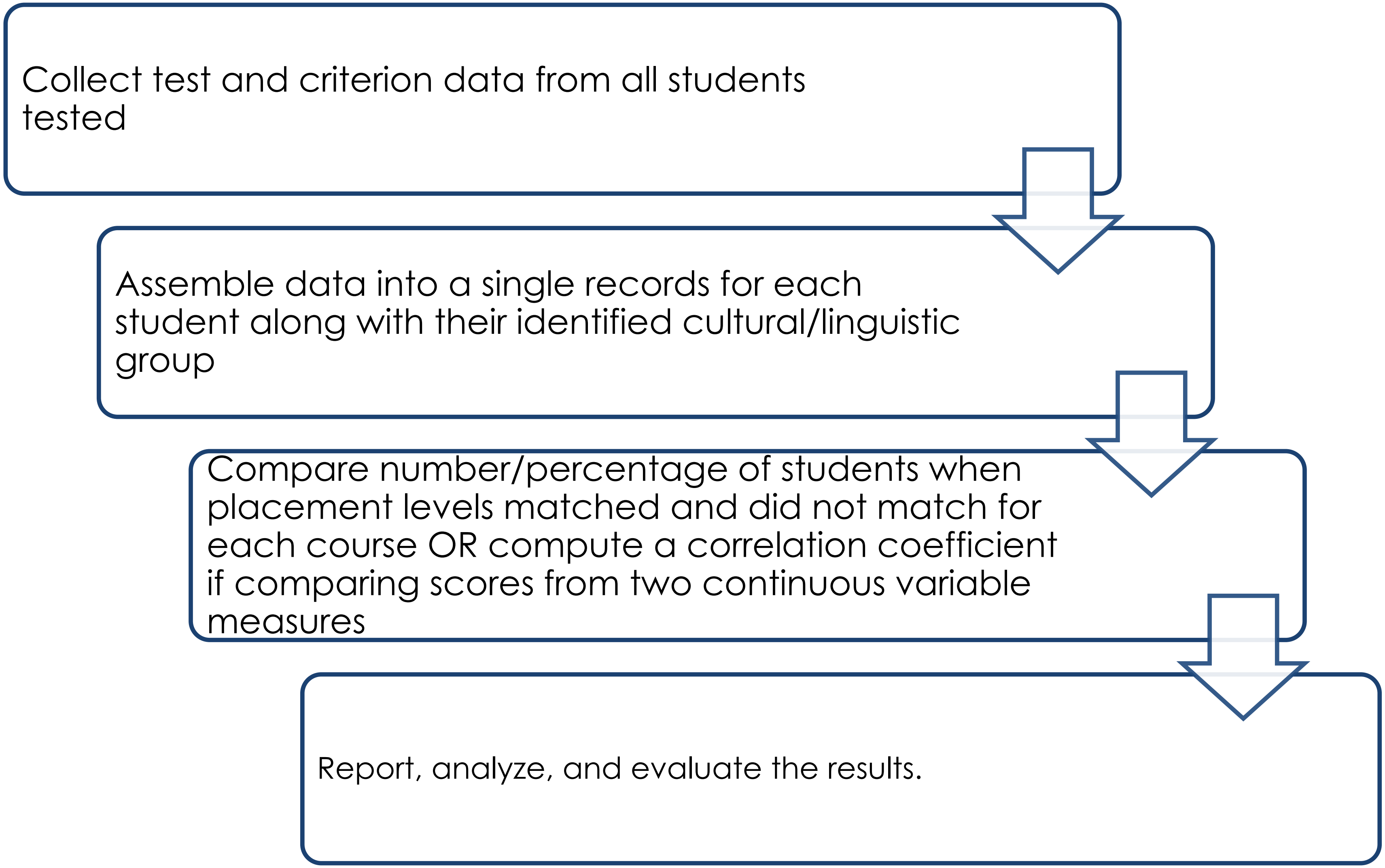
Criterion Validation: Types Of Comparisons

After initial enrollment (predictive)

Test variable	Criterion variable
Recommended course placement based on test score	Instructor assessment of student proficiency
Recommended course placement based on test score	Midterm/Final course grade
Recommended course placement based on test score	Midterm/Final course exam score

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Criterion Validation: Process At Time Of Testing



Collect test and criterion data from all students tested

Assemble data into a single records for each student along with their identified cultural/linguistic group

Compare number/percentage of students when placement levels matched and did not match for each course OR compute a correlation coefficient if comparing scores from two continuous variable measures

Report, analyze, and evaluate the results.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Criterion Validation: Example At Time Of Testing

Results table for all students tested

		Student Self-Assessment		
	Placement Level	ESL 1 (n = #)	ESL 2 (n = #)	Transfer Comp (n = #)
Test Score	ESL 1 (n = #)	60%	30%	10%
	ESL 2 (n = #)	25%	70%	5%
	Transfer Comp (n=#)	0%	0%	100%

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Criterion Validation: Example At Time of Testing

Results tables for different cultural/linguistic groups

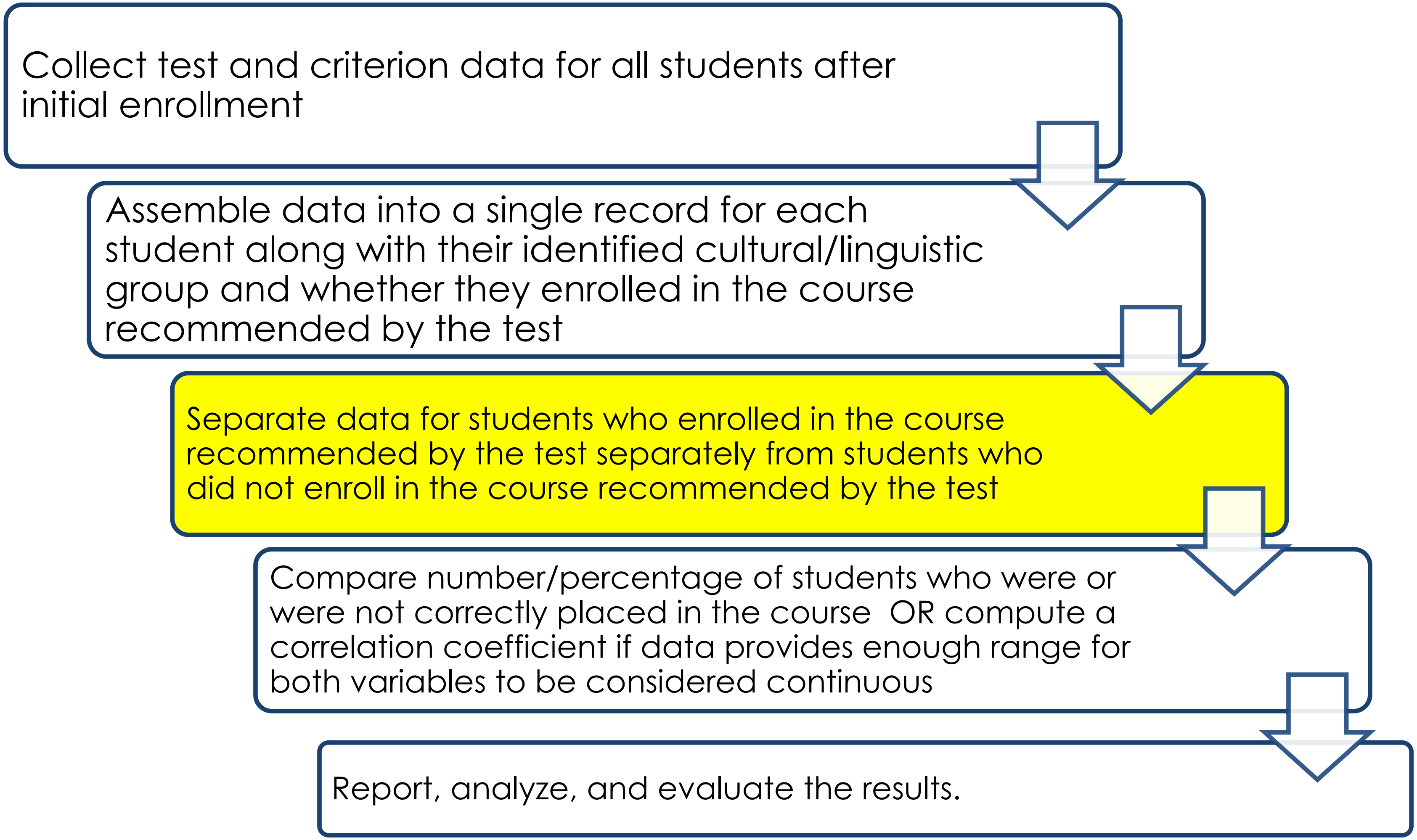
Cultural/Linguistic Group 1

		Student Self-Assessment		
	Placement Level	ESL 1 (n =#)	ESL 2 (n =#)	Transfer Comp (n =#)
Test Score	ESL 1 (n = #)	70%	25%	5%
	ESL 2 (n = #)	15%	75%	10%
	Transfer Comp (n=#)	5%	5%	90%

Cultural/Linguistic Group 2

		Student Self-Assessment		
	Placement Level	ESL 1 (n =#)	ESL 2 (n =#)	Transfer Comp (n =#)
Test Score	ESL 1 (n = #)	80%	20%	0%
	ESL 2 (n = #)	20%	60%	20%
	Transfer Comp (n=#)	10%	15%	75%

Criterion Validation: Process After Initial Enrollment



Collect test and criterion data for all students after initial enrollment

Assemble data into a single record for each student along with their identified cultural/linguistic group and whether they enrolled in the course recommended by the test

Separate data for students who enrolled in the course recommended by the test separately from students who did not enroll in the course recommended by the test

Compare number/percentage of students who were or were not correctly placed in the course OR compute a correlation coefficient if data provides enough range for both variables to be considered continuous

Report, analyze, and evaluate the results.

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln COLLEGE OF EDUCATION & HUMAN SCIENCES

Criterion Validation:
Example Instructor Assessment Of Proficiency

Instructor rating scale:

How prepared is the student related to your course prerequisite skills in order to succeed in your course?

1. Unprepared for the course. Student probably should have been placed into a lower level course.
2. Adequately prepared for the course. Student was placed into the appropriate level course.
3. Over-prepared for the course. Student probably should have been placed into a higher level course.

Reporting results:

Instructor Rating	ESL1 (n=#)	ESL2 (n=#)	Transfer Comp (n=#)
Over-prepared	22%	5%	4%
Adequately prepared	58%	85%	82%
Under-prepared	20%	10%	14%

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Criterion Validation: Course/Exam Grades

Reporting results:

Course/Exam Grades	ESL1 (n=#)	ESL2 (n=#)	Transfer Comp (n=#)
A/B	22%	85%	82%
C	58%	5%	4%
D/F	20%	10%	14%

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Criterion Validation: Analysis Considerations

- Discuss results instead of using hard and fast cutoffs (e.g., 75%)
- Results may lead to cautions about the use of the test for certain courses or students or reconsideration of cut scores
- Poor results could be due poor measurement of your criterion variable
- Only one source of validity – consider results along with other sources

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Criterion Validation: Common Errors/Omissions

- Data for courses are combined rather than reported individually.
- Too few courses or students included, limit generalizability.
- Only data is reported with little to no attention as to whether cut scores should be revisited or use of the test should be reconsidered.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

QUESTIONS?

BUROS
CENTER FOR TESTING

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

BUROS
CENTER FOR TESTING

University of Nebraska Lincoln | College of Education & Human Sciences

Consequential Validation

Jessica L. Jonson, PhD

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Consequential Validity: Overview

Examination of the intended and unintended consequences of test use.

Key consequence for CCC: Students with a goal of transferring to a 4-year institution or an associates degree should enter and complete a transfer-level composition course or ESL course equivalent to transfer-level English composition within 3 years of declaring a transfer- or degree-seeking goal (title 5, 55522.5)

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Consequential Validation: Documentation Requirements (p.21-23)

1. Describe the study sample
   - Demographically representative of ESL student population at local college (don't forget cultural/linguistic groups)
   - Representation from all ESL proficiency levels and cohorts
   - Census or random sample, not a convenience sample
   - Sufficient size (good rule of thumb n=30): Gather over multiple years if population is small

2. Describe study methods

- What, when, and how data was collected (recent data – last 3-5 years)
- How data was analyzed

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Consequential Validation: Documentation Requirement (p.21-23

3. Summarize the results and actions taken

- Provide a demographic representation of the study sample
- Provide descriptive statistics and distribution for all study variables:
  - across the entire sample,
  - for students who did and did not enroll in the initial course recommended by the test, and
  - for relevant cultural/linguistic groups.
- Report results for all courses in the ESL sequence and the transfer level composition
- When sample sizes permit, report results separately for cultural/linguistic groups

4. Evaluate and conclude

- Based on the results, make recommendations about the use of the test scores for the placement decisions of students from different demographic groups and for specific course/proficiency levels

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | College of Education & Human Sciences

Consequential Validation: Submission Requirements

New submissions: Study required for full approval

- Probationary: Detailed, appropriate plan for conducting study is provided.

Renewal submissions: Study is required

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | College of Education & Human Sciences

Consequential Validity: Study Variables

Key variables:

- test score and recommended placement level
- Whether the student successfully completed transfer-level composition 3 years after initial enrollment in ESL sequence

Other helpful data

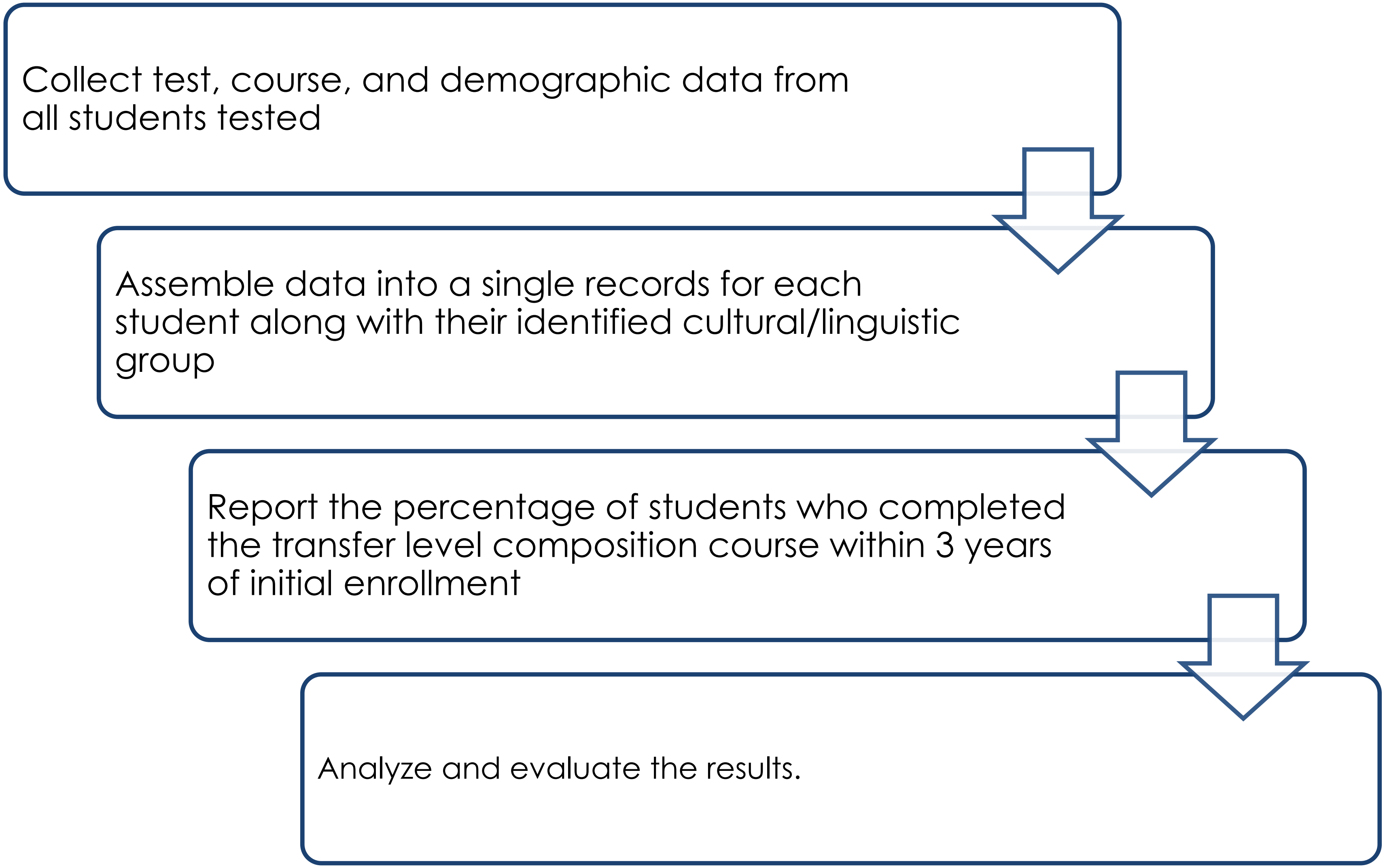
1. If the student is degree seeking?
2. Cultural/linguistic group
3. ESL cohort
4. Final placement recommendation
5. Initial ESL course the student enrolled
6. Term and course grade for each enrolled ESL course culminating with transfer-level composition course

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Consequential Validation: Process



Collect test, course, and demographic data from all students tested

Assemble data into a single records for each student along with their identified cultural/linguistic group

Report the percentage of students who completed the transfer level composition course within 3 years of initial enrollment

Analyze and evaluate the results.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Consequential Validity: Example Analyses

- Percentage of students who did and did not enroll in the course recommended by the test
  - Does this differ by cultural/linguistic group or another demographic?
- Percentage of students who have and have not successfully completed transfer level composition within 3 years times of declaring a transfer- or degree-seeking goal?
  - For each course level initially enrolled, report percentages for each cohort and cultural/linguistic group

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | College of Education & Human Sciences

Consequential Validation: Results Tables

% degree-seeking students completed transfer level composition in 3 years

		Initial Course Enrollment			
	% enrolled in course recommended by the test	ESL 1 (n=#)	ESL 2 (n=#)	ESL 3 (n=#)	Overall (n=#)
Cultural/linguistic group 1 (n= #)	80%	78%	85%	98%	79%
Cultural/linguistic group 2 (n= #)	65%	60%	75%	92%	70%

		Initial Course Enrollment			
	% enrolled in course recommended by the test	ESL 1 (n=#)	ESL 2 (n=#)	ESL 3 (n=#)	Overall (n=#)
Cohort – Year 1(n= #)	82%	66%	92%	100%	85%
Cohort – Year 2 (n= #)	75%	75%	88%	93%	89%
Cohort – Year 3 (n= #)	80%	78%	90%	95%	90%

permission of the Buros Center for Testing.

Consequential Validation: Analysis Considerations

- What differences are found between cultural/linguistic groups in terms of initial course enrollments & completion rates?
- Is there consistent results across cohorts? If not, why might those results have been different?
- Are completion rates a concern? What follow-up analyses should be done to determine if the concerns are a result of the validity of test score interpretation?
- If concerns about the validity of test score interpretation, should cut scores be revisited? Additional reliability, validity, or fairness analyses conducted for different cultural/linguistic groups or another demographic?

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

QUESTIONS?

BUROS
CENTER FOR TESTING

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

BUROS
CENTER FOR TESTING

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Reliability

Maria Elena Oliveri, PhD

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Overview

Reliability is…how consistently (accurately or precisely) a "test" measures a given construct.

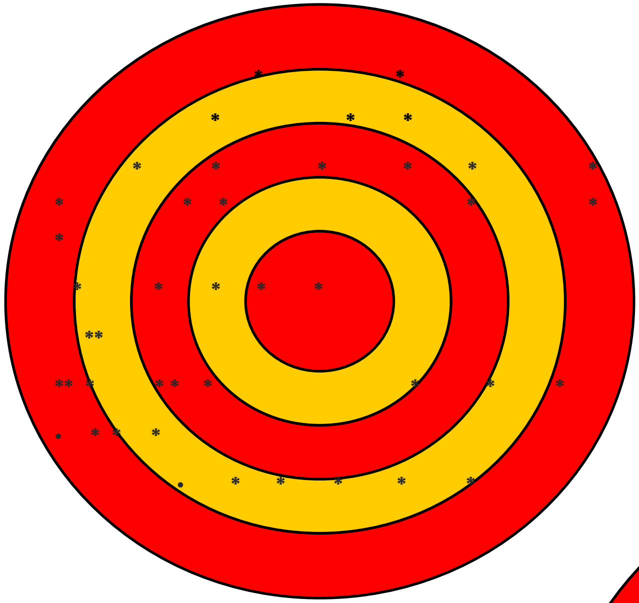
Validity is … degree of relevance of score-based inferences.

Consistency allows for confidence in score-based inferences.

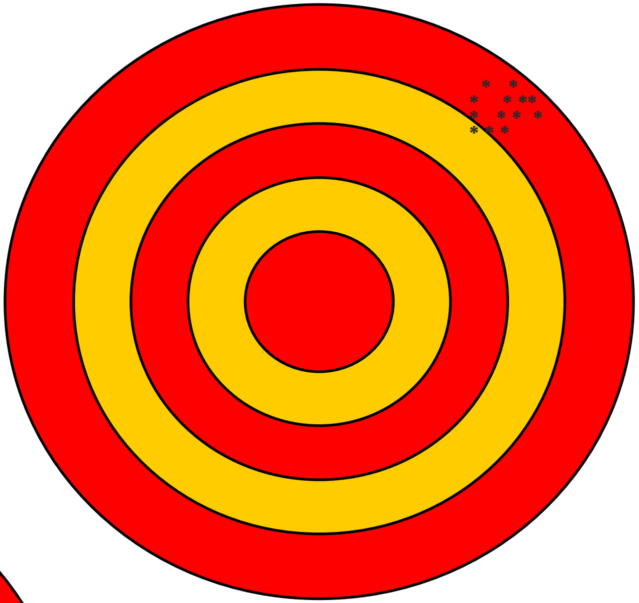
BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

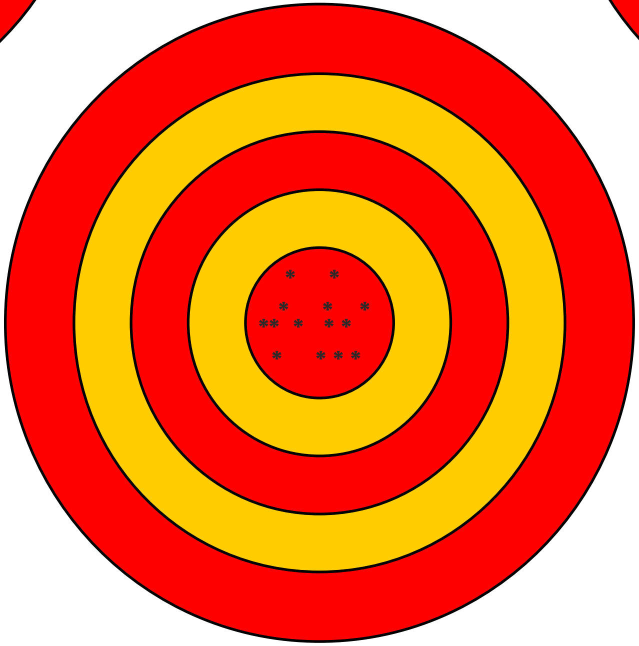
University of Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES



NEITHER
Reliable nor Valid



Reliable
but NOT Valid



BOTH
Reliable and Valid

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln
COLLEGE OF EDUCATION & HUMAN SCIENCES

Tests Are Not Reliable

A test is not reliable – reliability evidence is attributable to the **scores** from the test

- Depends on circumstances test is given
- Also sample dependent

Ideally…

Looking for reliability evidence from a sample of students similar to EL population at CCC in a context in which placement decisions will be made.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability Overview - Continued

No assessment instrument is free of error, which requires that the reliability of the assessment instrument and the degree of error associated with test scores be documented.

Error can stem from multiple sources - the reliability evidence provided should consider the error sources that are most relevant and of greatest concern for the assessment instrument.

Report reliability estimate and standard error of measurement(SEM) for all reported scores. If corrected estimates are reported, uncorrected should be as well.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Doc. Requirements (pp.23-24)

Reliability information is required.

- Report the percentage (or number) of students in the study sample and provide a demographic comparison of the study sample with the demographic representation of the local college ESL student population. Include a sufficient and representative sample of ESL students from cultural/linguistic groups that constitute approximately 2% or more of the ESL student population at the local college.
  - Encourage representation from full range ESL proficiency levels and from all available ESL cohorts
- Describe the study methods (for each type of relevant measurement error); describe what data were collected, when (in last 3 years), how data were collected & analyzed.
- Summarize the results and actions based on the results.
  - Report SEM across the score scale and confidence intervals at cut points.
- Provide conclusions and summarize recommendations.

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Reliability: Submission Requirements

New submissions: Reliability information addressing is required. Including:

- Internal consistency
- Test-retest
- Any other relevant sources (inter-rater, inter-prompt, inter-form)

Probationary: At least one reliability study

Renewal submissions: Reliability information is required. Including:

- Internal consistency
- Any other relevant sources (inter-rater, inter-prompt, inter-form)

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Types of Reliability Estimates

Test-Retest/Stability(1 form, 2 occasions)

Administer same test to same group across two time points and correlate scores

Internal consistency (1 form, 1 occasion)

Compare responses across items within test from 1 occasion

Equivalent forms (2 or more forms/prompts, 1 occasion)

Use one set of questions on same construct and divide into two equivalent sets administered to same sample and compare scores

Human scoring

Compare scores across multiple raters

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

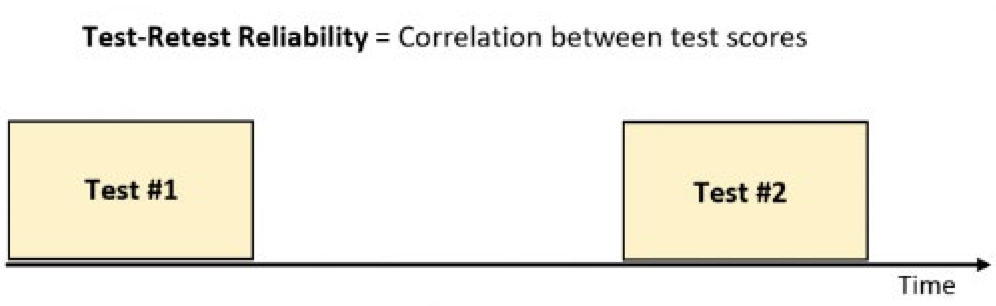
University of Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Reliability: Process

- Conduct a reliability study:
  - administer the same test on two occasions (test-retest approach);
  - internal consistency.
- Subscores: Evaluate reliability of subtest scores if subtest scores are used to make placement decisions.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Test-Retest



Test-Retest Reliability = Correlation between test scores

Test #1

Test #2

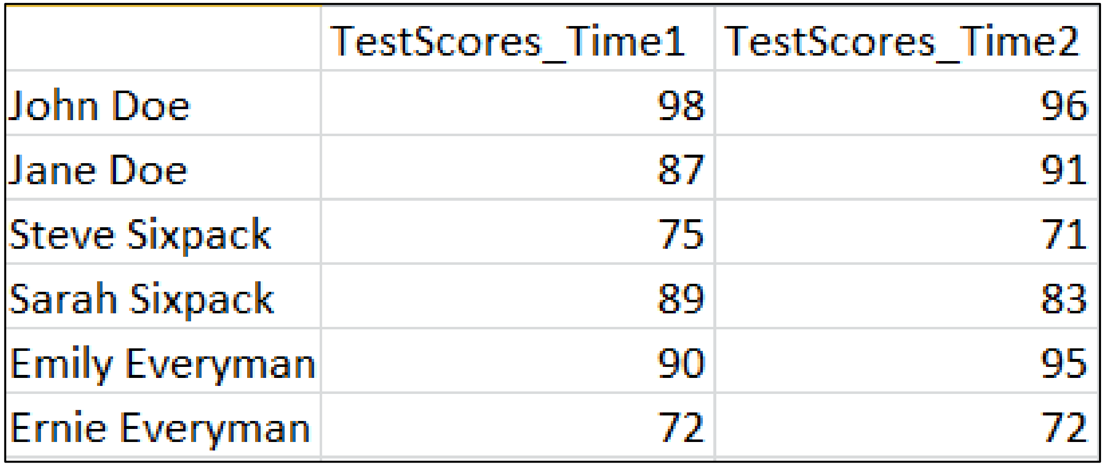
Time

To calculate the test-retest reliability, you can use the Pearson Correlation Coefficient, which takes on a value between -1 and 1 where:

- -1 indicates a perfectly negative linear correlation between two scores.
- 0 indicates no linear correlation between two scores.
- 1 indicates a perfectly positive linear correlation between two scores.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Test-Retest (Excel Example) and Spreadsheet



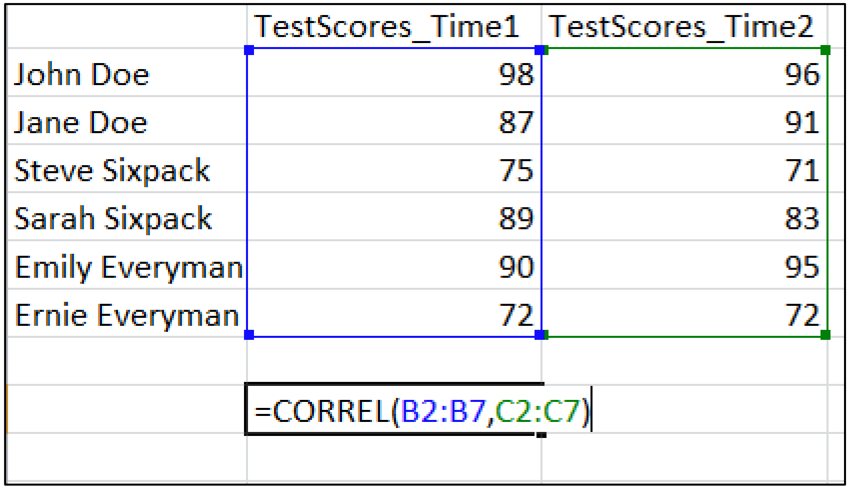
	TestScores_Time1	TestScores_Time2
John Doe	98	96
Jane Doe	87	91
Steve Sixpack	75	71
Sarah Sixpack	89	83
Emily Everyman	90	95
Ernie Everyman	72	72

Step 1) Collect test data from the same students at time 1 and time 2. The test at time 2 should be administered at least two weeks after the test at time 1.

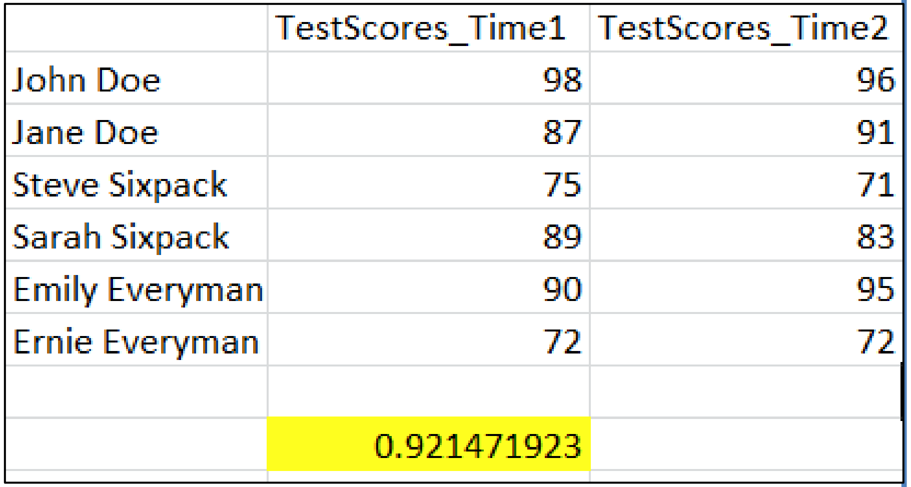
Step 2) Pair the scores from students at time 1 and time 2.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Test-Retest (Excel Example) and Spreadsheet (cont.)



	TestScores_Time1	TestScores_Time2
John Doe	98	96
Jane Doe	87	91
Steve Sixpack	75	71
Sarah Sixpack	89	83
Emily Everyman	90	95
Ernie Everyman	72	72
	=CORREL(B2:B7,C2:C7)	



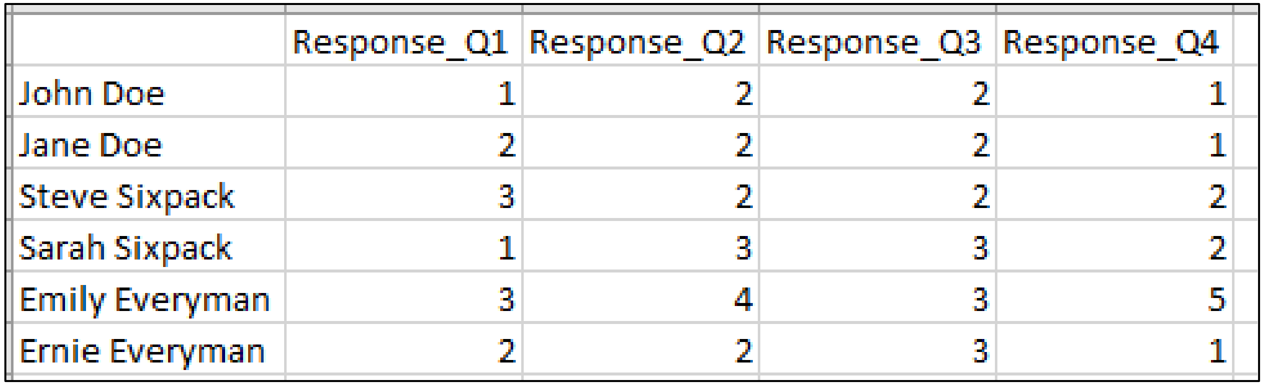
	TestScores_Time1	TestScores_Time2
John Doe	98	96
Jane Doe	87	91
Steve Sixpack	75	71
Sarah Sixpack	89	83
Emily Everyman	90	95
Ernie Everyman	72	72
	0.921471923	

Step 3) Use the CORREL function, calculate the correlation between the test scores at times 1 and 2.

Step 4) Interpret and report the results. **The threshold for acceptable test-retest reliability is .75.**

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Internal Consistency

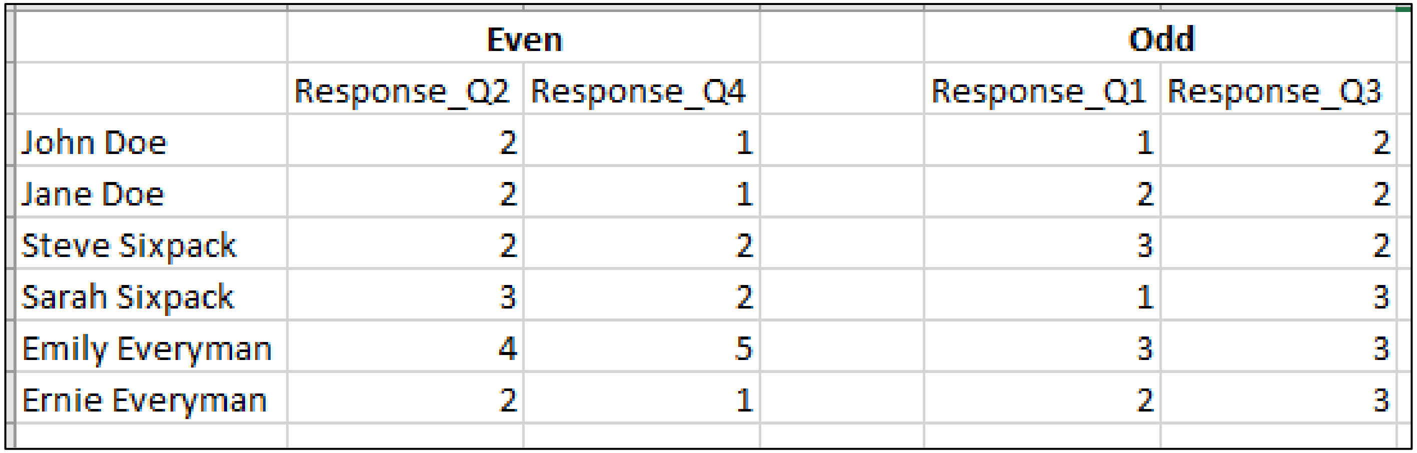


	Response_Q1	Response_Q2	Response_Q3	Response_Q4
John Doe	1	2	2	1
Jane Doe	2	2	2	1
Steve Sixpack	3	2	2	2
Sarah Sixpack	1	3	3	2
Emily Everyman	3	4	3	5
Ernie Everyman	2	2	3	1

Step 1) Administer a test to a group of students and collect their response data. If necessary, convert the responses into a numerical format.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Internal Consistency (cont.)



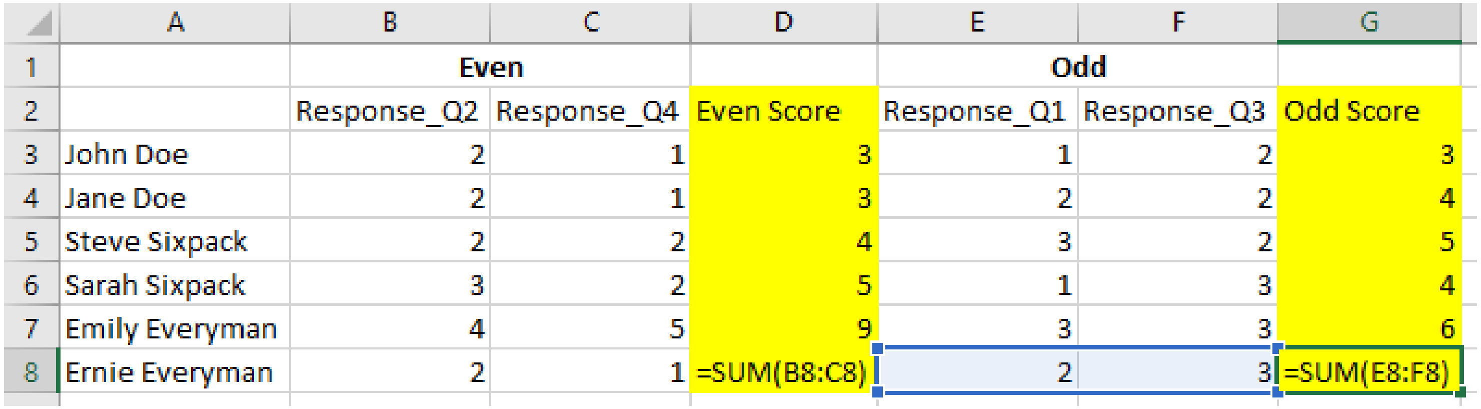
	Even			Odd	
	Response_Q2	Response_Q4		Response_Q1	Response_Q3
John Doe	2	1		1	2
Jane Doe	2	1		2	2
Steve Sixpack	2	2		3	2
Sarah Sixpack	3	2		1	3
Emily Everyman	4	5		3	3
Ernie Everyman	2	1		2	3

Step 2) Divide the test into two parts. For example, split the test by even and odd numbered items or at random.

Note. It is easier for later calculations if each half of the test has the same number of items.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Internal Consistency (cont.)



	A	B	C	D	E	F	G
1		Even			Odd		
2		Response_Q2	Response_Q4	Even Score	Response_Q1	Response_Q3	Odd Score
3	John Doe	2	1	3	1	2	3
4	Jane Doe	2	1	3	2	2	4
5	Steve Sixpack	2	2	4	3	2	5
6	Sarah Sixpack	3	2	5	1	3	4
7	Emily Everyman	4	5	9	3	3	6
8	Ernie Everyman	2	1	=SUM(B8:C8)	2	3	=SUM(E8:F8)

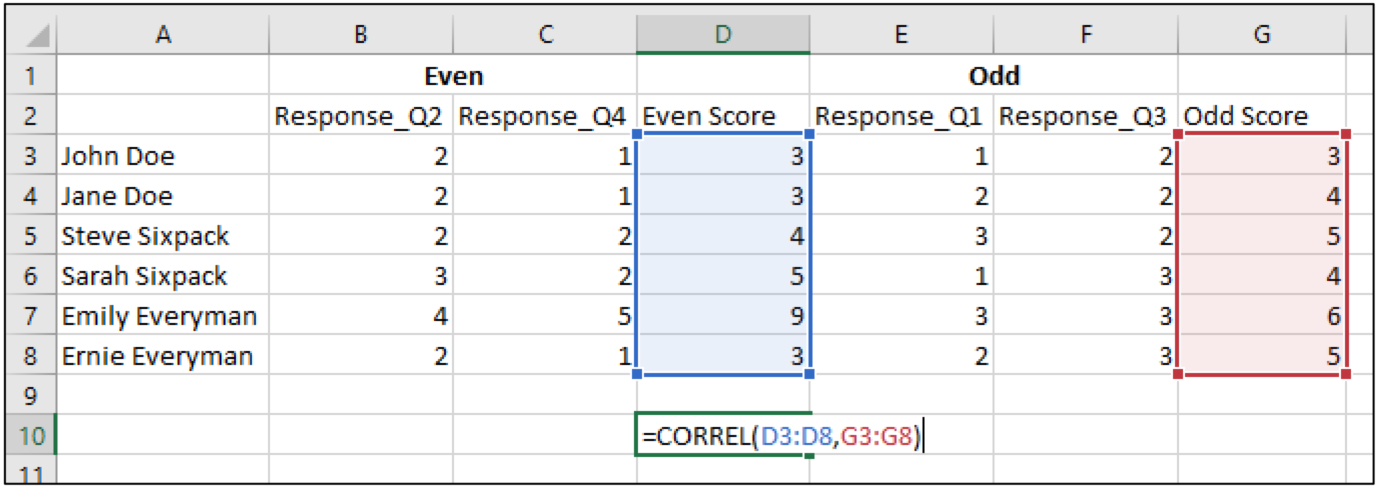
Step 3) Calculate the Even and Odd scores for each student using the SUM function.

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln
COLLEGE OF EDUCATION & HUMAN SCIENCES

Reliability: Internal Consistency (cont.)

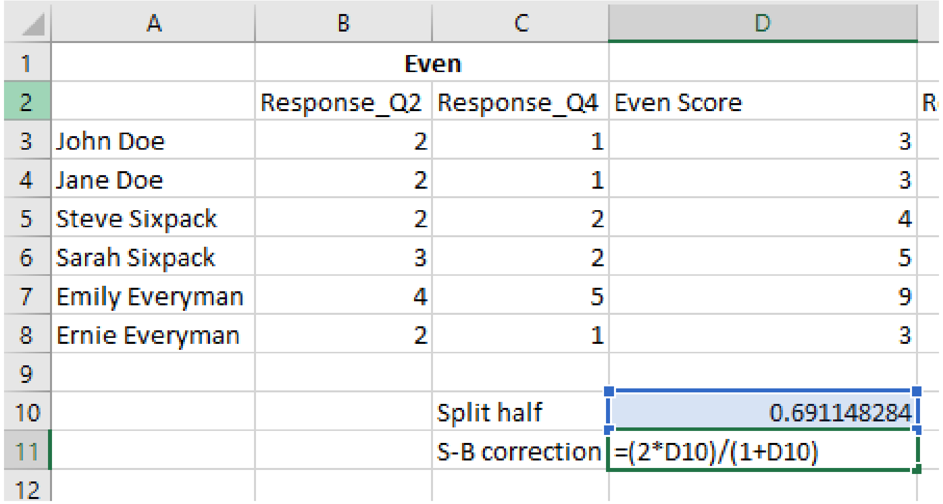


	A	B	C	D	E	F	G
1		Even			Odd		
2		Response_Q2	Response_Q4	Even Score	Response_Q1	Response_Q3	Odd Score
3	John Doe	2	1	3	1	2	3
4	Jane Doe	2	1	3	2	2	4
5	Steve Sixpack	2	2	4	3	2	5
6	Sarah Sixpack	3	2	5	1	3	4
7	Emily Everyman	4	5	9	3	3	6
8	Ernie Everyman	2	1	3	2	3	5
9							
10				=CORREL(D3:D8,G3:G8)			
11							

Step 4) Find the correlation between the even and odd scores using the CORREL function.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Internal Consistency



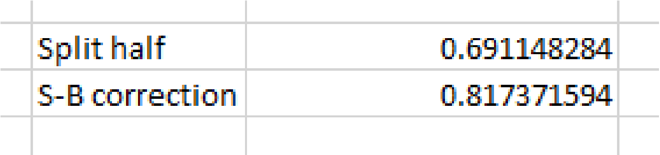
	A	B	C	D	
1		Even			
2		Response_Q2	Response_Q4	Even Score	R
3	John Doe	2	1	3	
4	Jane Doe	2	1	3	
5	Steve Sixpack	2	2	4	
6	Sarah Sixpack	3	2	5	
7	Emily Everyman	4	5	9	
8	Ernie Everyman	2	1	3	
9					
10			Split half	0.691148284	
11			S-B correction	=(2*D10)/(1+D10)	
12					

Step 5) The reliability estimate will likely be lower than you were expecting. Reliability is a function of test length, and you just cut yours in half! You can correct for this by using the Spearman-Brown formula: $r_{\text{predicted}} = \frac{2r}{1+r}$, where *r* is the split-half reliability estimate.

Note. This version of the formula only works when your test was split into equal halves.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Internal Consistency



Split half	0.691148284
S-B correction	0.817371594

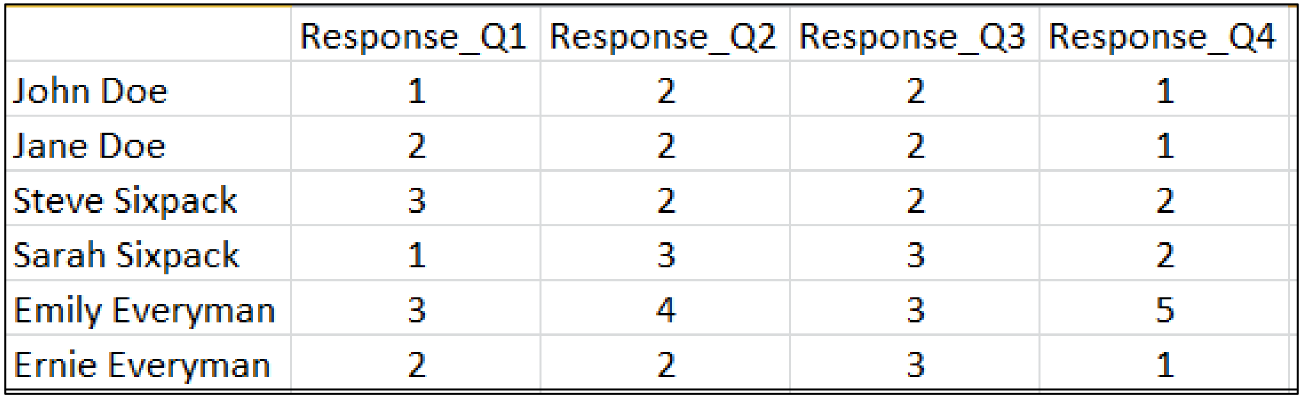
Step 6) Interpret the corrected reliability. **The acceptable threshold for an estimate of internal-consistency reliability is .80.**

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Internal Consistency

α

- Cronbach alpha coefficient



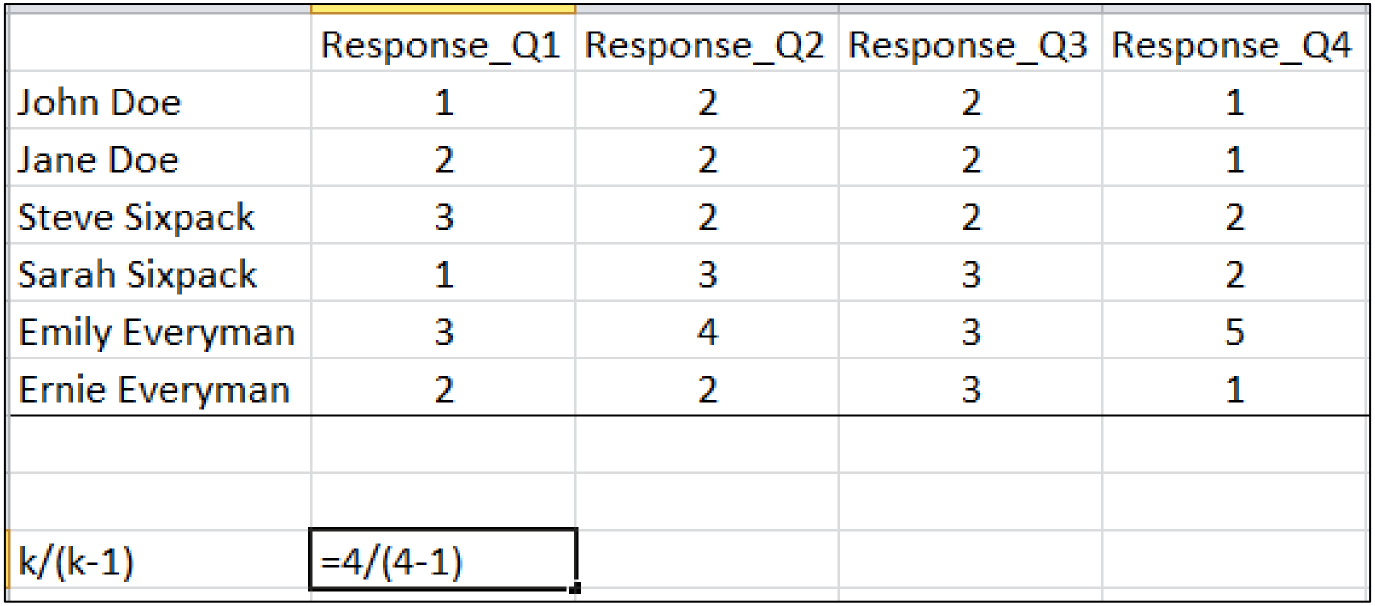
	Response_Q1	Response_Q2	Response_Q3	Response_Q4
John Doe	1	2	2	1
Jane Doe	2	2	2	1
Steve Sixpack	3	2	2	2
Sarah Sixpack	1	3	3	2
Emily Everyman	3	4	3	5
Ernie Everyman	2	2	3	1

Step 1) For a single test, collect test takers' raw responses. If necessary, convert the responses into a numerical format.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Internal Consistency (cont.)

α



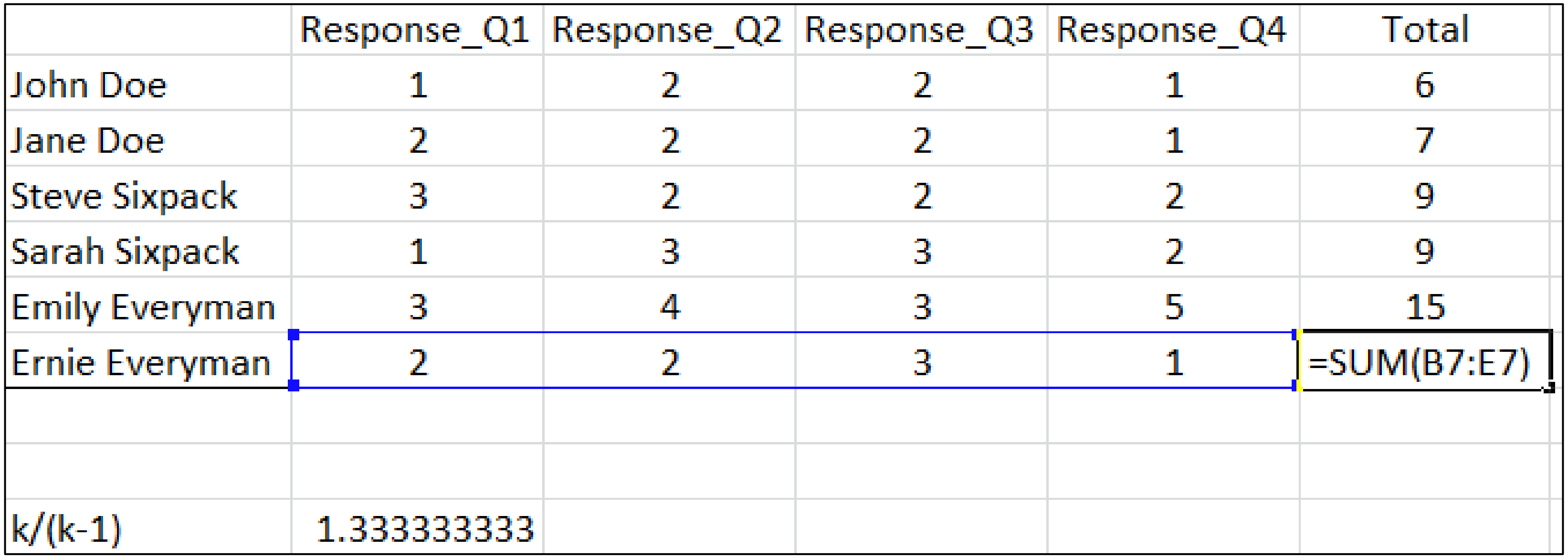
	Response_Q1	Response_Q2	Response_Q3	Response_Q4
John Doe	1	2	2	1
Jane Doe	2	2	2	1
Steve Sixpack	3	2	2	2
Sarah Sixpack	1	3	3	2
Emily Everyman	3	4	3	5
Ernie Everyman	2	2	3	1
k/(k-1)	=4/(4-1)			

Step 2) Calculate **k/(k-1)**, where k is the total number of questions on the test.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Internal Consistency (cont.)

α



	Response_Q1	Response_Q2	Response_Q3	Response_Q4	Total
John Doe	1	2	2	1	6
Jane Doe	2	2	2	1	7
Steve Sixpack	3	2	2	2	9
Sarah Sixpack	1	3	3	2	9
Emily Everyman	3	4	3	5	15
Ernie Everyman	2	2	3	1	=SUM(B7:E7)
k/(k-1)	1.333333333				

Step 3) Calculate each examinee's total score by summing their responses using the SUM function.

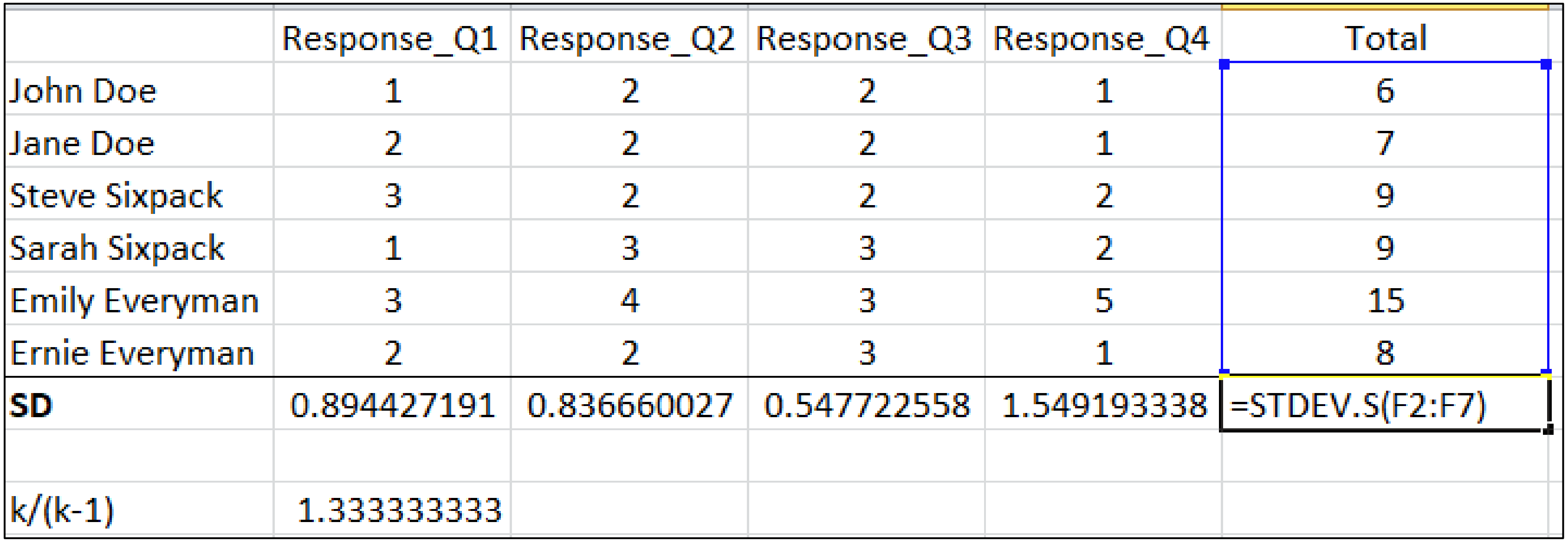
BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | College of Education & Human Sciences

Reliability: Internal Consistency (cont.)

α



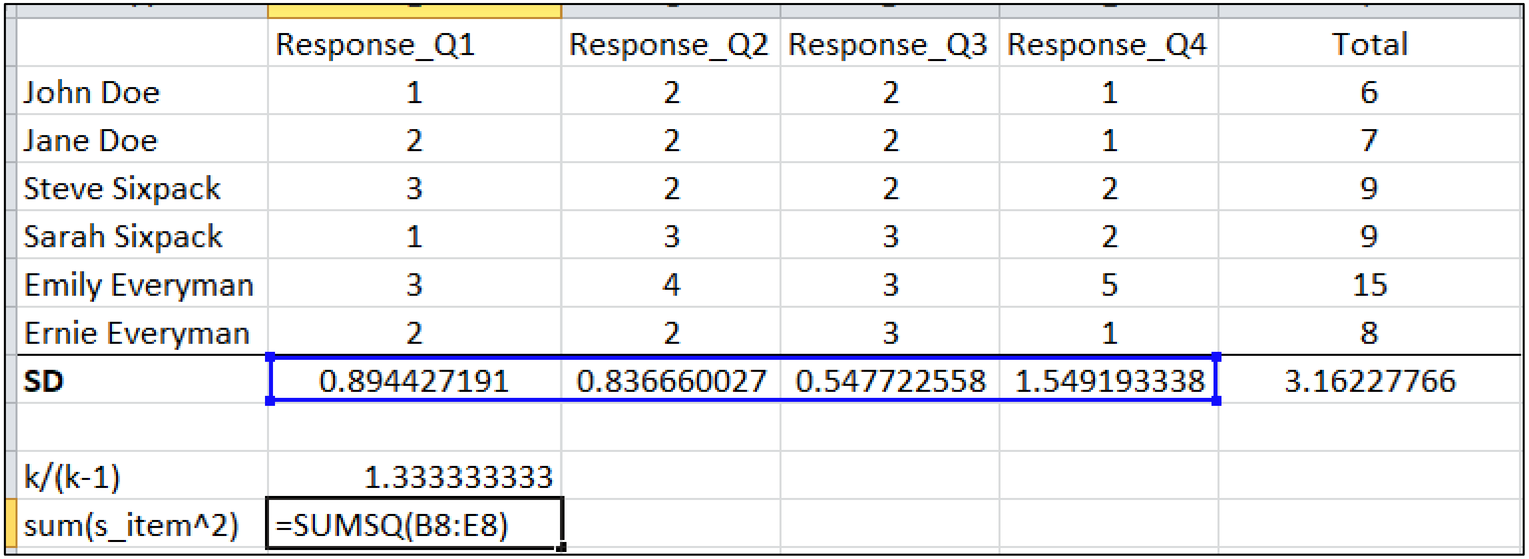
	Response_Q1	Response_Q2	Response_Q3	Response_Q4	Total
John Doe	1	2	2	1	6
Jane Doe	2	2	2	1	7
Steve Sixpack	3	2	2	2	9
Sarah Sixpack	1	3	3	2	9
Emily Everyman	3	4	3	5	15
Ernie Everyman	2	2	3	1	8
SD	0.894427191	0.836660027	0.547722558	1.549193338	=STDEV.S(F2:F7)
k/(k-1)	1.333333333				

Step 4) Calculate the standard deviation of the responses to each item and the total scores using the STDEV.S function.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Internal Consistency (cont.)

α



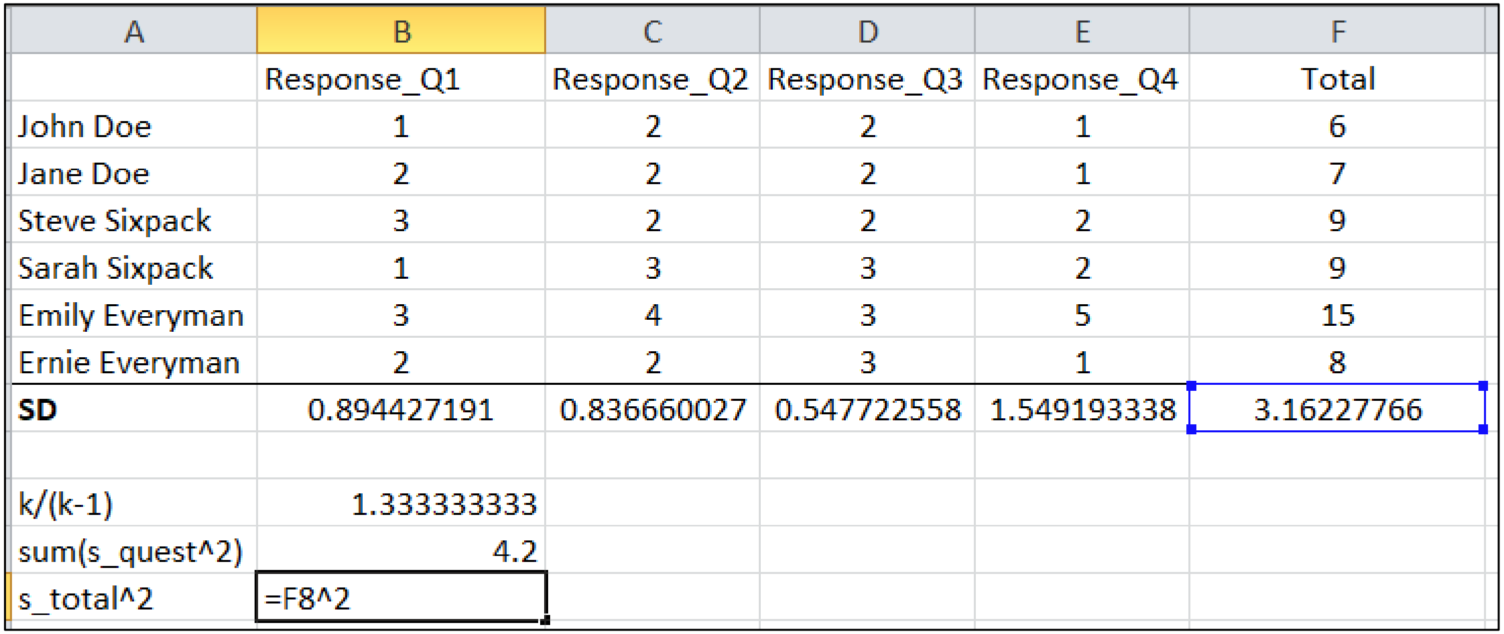
	Response_Q1	Response_Q2	Response_Q3	Response_Q4	Total
John Doe	1	2	2	1	6
Jane Doe	2	2	2	1	7
Steve Sixpack	3	2	2	2	9
Sarah Sixpack	1	3	3	2	9
Emily Everyman	3	4	3	5	15
Ernie Everyman	2	2	3	1	8
SD	0.894427191	0.836660027	0.547722558	1.549193338	3.16227766
k/(k-1)	1.333333333				
sum(s_item^2)	=SUMSQ(B8:E8)				

Step 5) Calculate the sum of the squared standard deviations for the questions using the SUMSQ function.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Internal Consistency (cont.)

α



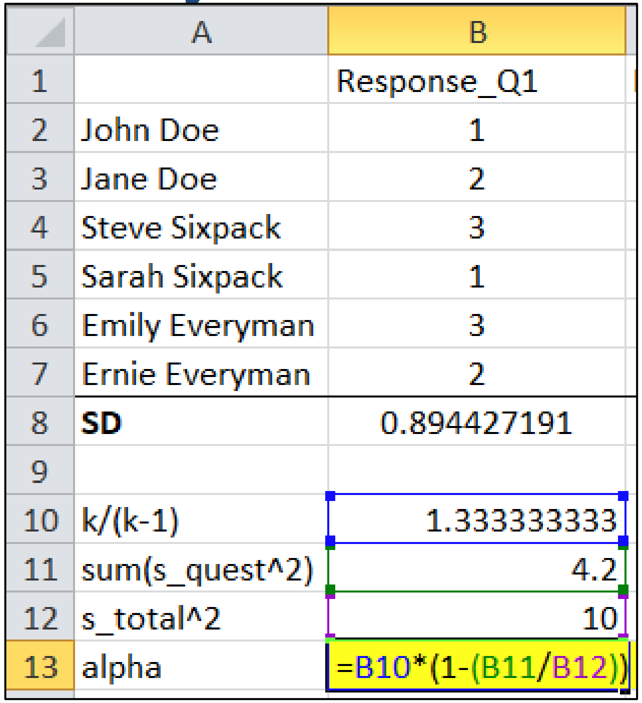
A	B	C	D	E	F
	Response_Q1	Response_Q2	Response_Q3	Response_Q4	Total
John Doe	1	2	2	1	6
Jane Doe	2	2	2	1	7
Steve Sixpack	3	2	2	2	9
Sarah Sixpack	1	3	3	2	9
Emily Everyman	3	4	3	5	15
Ernie Everyman	2	2	3	1	8
SD	0.894427191	0.836660027	0.547722558	1.549193338	3.16227766
k/(k-1)	1.333333333				
sum(s_quest^2)	4.2				
s_total^2	=F8^2				

Step 6) Square the standard deviation of the total scores.

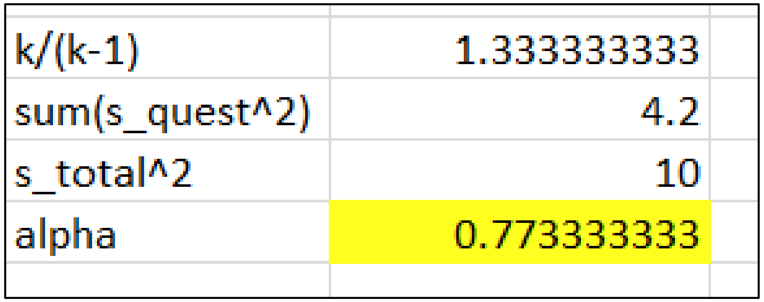
Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Internal Consistency (cont.)

α



	A	B
1		Response_Q1
2	John Doe	1
3	Jane Doe	2
4	Steve Sixpack	3
5	Sarah Sixpack	1
6	Emily Everyman	3
7	Ernie Everyman	2
8	**SD**	0.894427191
9		
10	k/(k-1)	1.333333333
11	sum(s_quest^2)	4.2
12	s_total^2	10
13	alpha	=B10*(1-(B11/B12))



k/(k-1)	1.333333333
sum(s_quest^2)	4.2
s_total^2	10
alpha	0.773333333

Step 7) Combine all the parts. Use the equation:

$$\alpha = \left(\frac{k}{(k-1)}\right) * \left(1 - \left(\left(\sum s_i^2\right) / s_t^2\right)\right)$$

Step 8) Interpret the result. **The threshold for acceptable internal-consistency reliability is .80**.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

CCC Standards: Reliability Estimate Minima

Internal consistency = .80 or higher

Equivalent form/inter-prompt = .75 or higher

Inter-rater (prefer methods correcting for chance)

- Intraclass correlation = .75+
- Interscorer correlation = .70+
- Percent agreement = 90% or higher (1 point difference)
- Cohens Kappa = .40 or higher
- Report how inconsistencies between scorers were resolved

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Standard Error Of Measurement (SEM)

Estimate of how repeated measures of a person's score (SD) is distributed around their "true" score to determine the degree of test score precision.

- Lower values preferred (less error)

EXAMPLE: Student score = 30

- SEM 3.5 30 +/- 3.5 = 26.5 to 33.5
- SEM 10 30 +/- 10 = 20 to 30

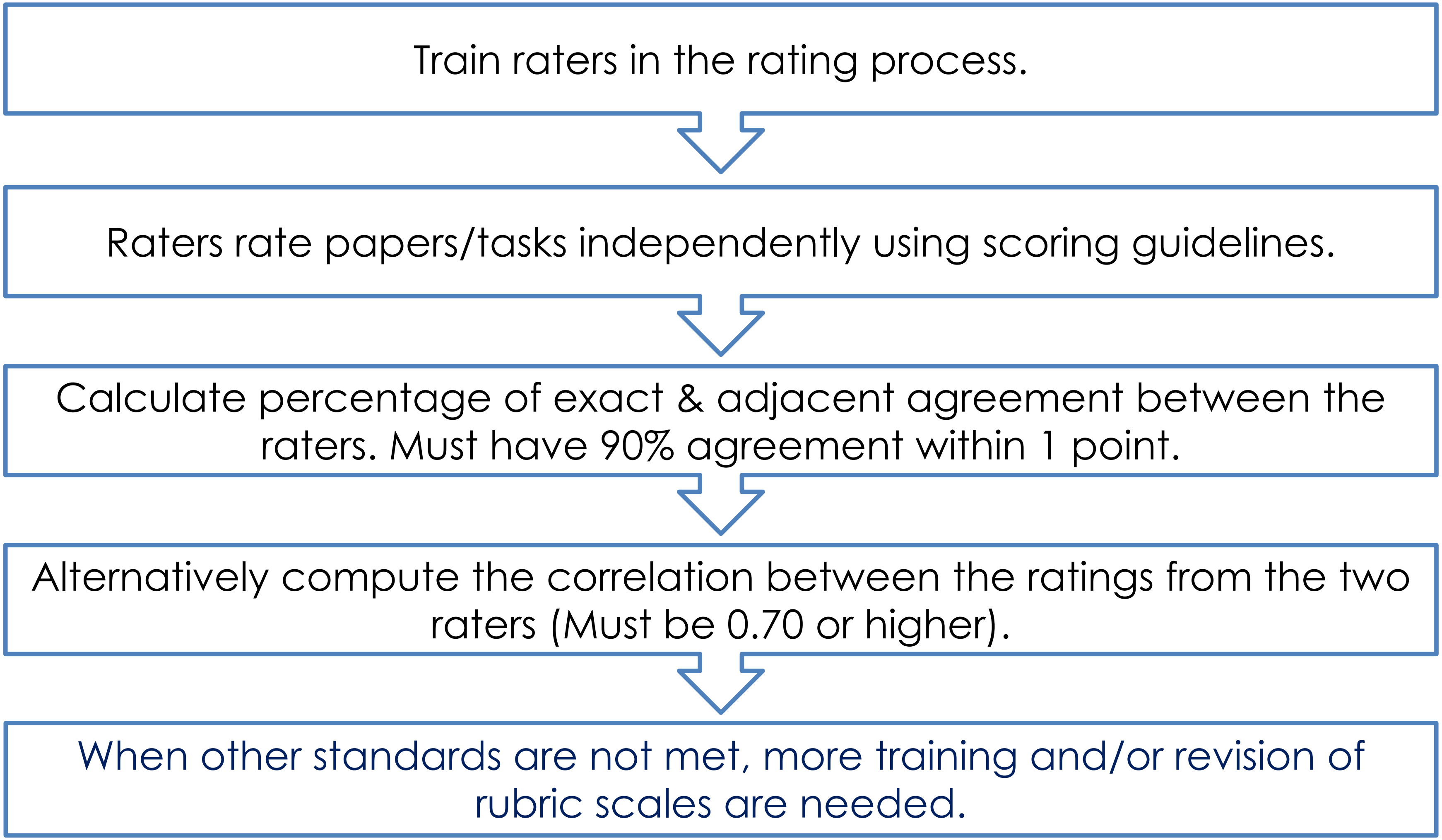
Important to know SEM across score distribution especially at consequential cut points.

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Reliability Studies: Rater Agreement



Train raters in the rating process.

Raters rate papers/tasks independently using scoring guidelines.

Calculate percentage of exact & adjacent agreement between the raters. Must have 90% agreement within 1 point.

Alternatively compute the correlation between the ratings from the two raters (Must be 0.70 or higher).

When other standards are not met, more training and/or revision of rubric scales are needed.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Performance Tests

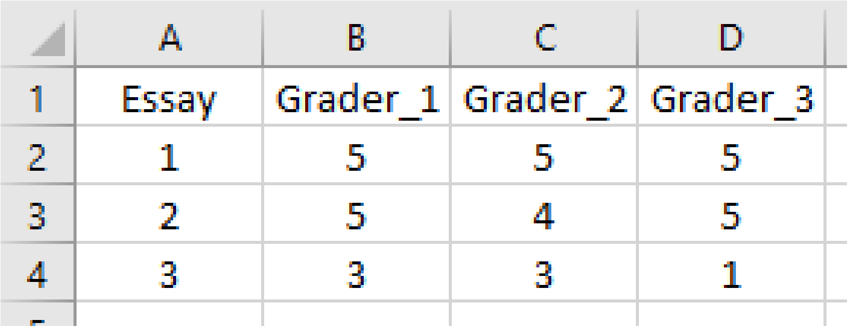
- Example – Percentage Agreement

	1	2	3	4	5	6
1	7	2	1	0	0	0
2	3	10	3	2	1	0
3	0	2	12	2	0	1
4	0	2	2	13	1	1
5	0	1	1	4	12	3
6	0	0	0	1	3	11

- A random sample of 100 papers were rated by the two independent readers.
- 90% agreement within one point difference between the two readers
- The CCC standard is met (>90% within one point on a 6-point scale)

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Percent Agreement

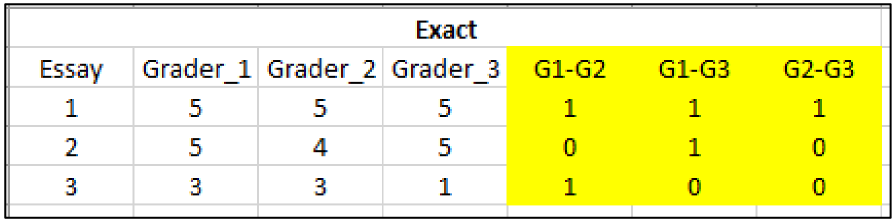


	A	B	C	D
1	Essay	Grader_1	Grader_2	Grader_3
2	1	5	5	5
3	2	5	4	5
4	3	3	3	1

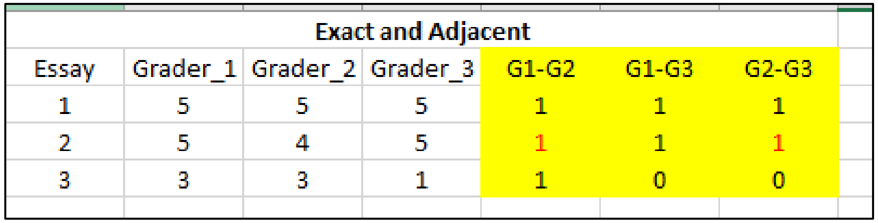
Step 1) Collect grader/rater ratings of a performance test (e.g., an essay or a presentation).

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Percent Agreement (cont.)



Exact						
Essay	Grader_1	Grader_2	Grader_3	G1-G2	G1-G3	G2-G3
1	5	5	5	1	1	1
2	5	4	5	0	1	0
3	3	3	1	1	0	0



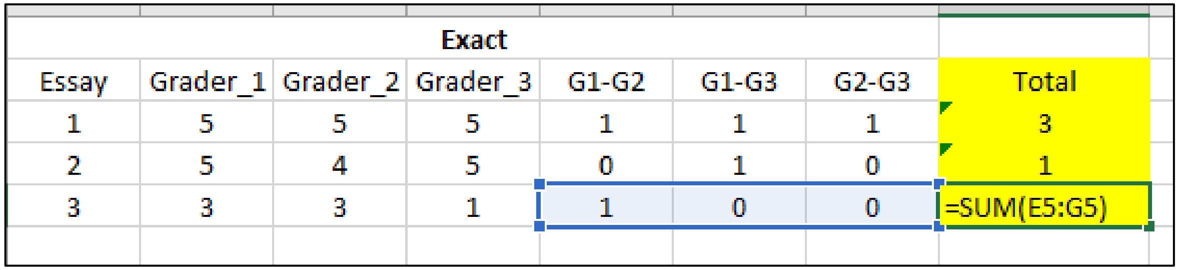
Exact and Adjacent						
Essay	Grader_1	Grader_2	Grader_3	G1-G2	G1-G3	G2-G3
1	5	5	5	1	1	1
2	5	4	5	1	1	1
3	3	3	1	1	0	0

Step 2) For each combination of graders, make a new column. In each of these columns for each essay, add a 1 if the grader pair gave the same rating, or a 0 if the grader pair gave a different rating.

Note. This will calculate *Exact* percent agreement. Exact and Adjacent percent agreement can be calculated by determining if the grader pair selected the same rating *or an adjacent rating*. The bottom picture shows what exact and adjacent scoring would look like.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Percent Agreement (cont.)

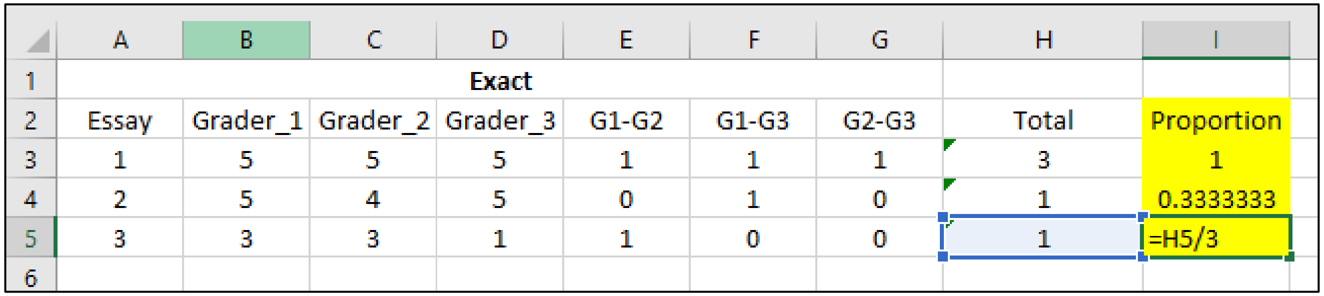


			Exact				
Essay	Grader_1	Grader_2	Grader_3	G1-G2	G1-G3	G2-G3	Total
1	5	5	5	1	1	1	3
2	5	4	5	0	1	0	1
3	3	3	1	1	0	0	=SUM(E5:G5)

Step 3) For each essay, calculate the total number of times the grader gave the same rating by using the SUM function.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Percent Agreement (cont.)

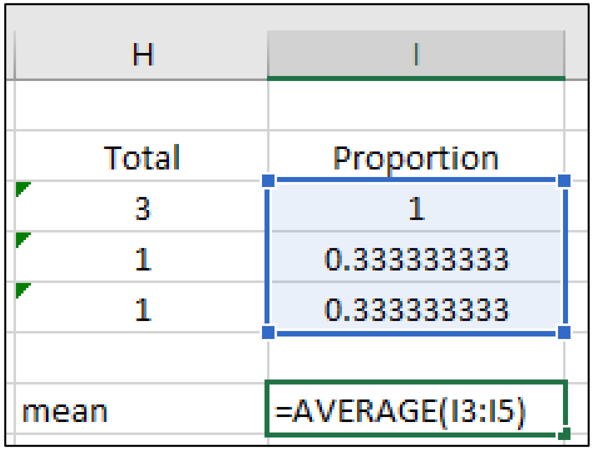


	A	B	C	D	E	F	G	H	I
1				Exact					
2	Essay	Grader_1	Grader_2	Grader_3	G1-G2	G1-G3	G2-G3	Total	Proportion
3	1	5	5	5	1	1	1	3	1
4	2	5	4	5	0	1	0	1	0.3333333
5	3	3	3	1	1	0	0	1	=H5/3
6									

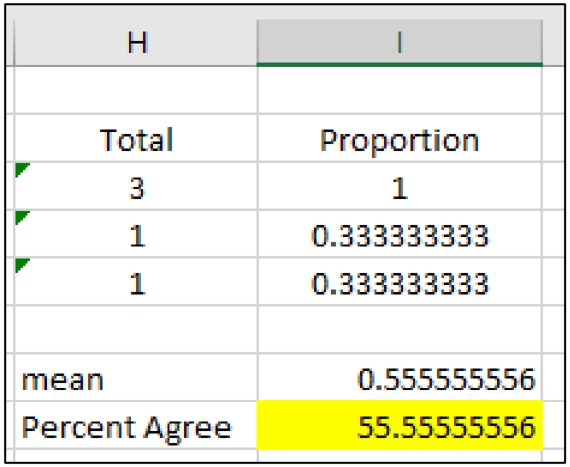
Step 4) Calculate the proportion of agreements across all grader pairs.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Percent Agreement (cont.)



H	I
Total	Proportion
3	1
1	0.333333333
1	0.333333333
mean	=AVERAGE(I3:I5)



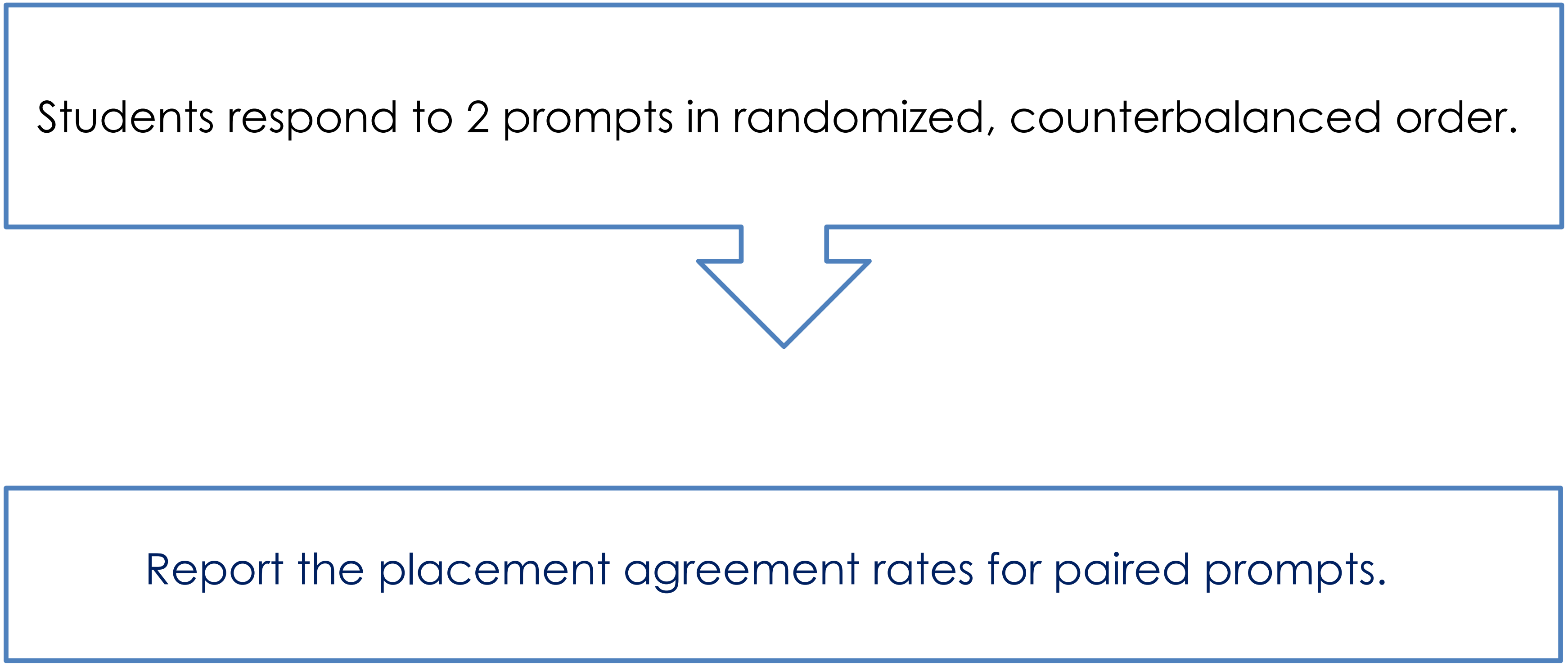
H	I
Total	Proportion
3	1
1	0.333333333
1	0.333333333
mean	0.555555556
Percent Agree	55.55555556

Step 5) Calculate the average of the proportion of agreements for all essays. Then, multiply the result by 100 to get the percent agreement.

Step 6) Interpret the results. A threshold of 90% is required for acceptable performance test reliability.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability Studies: Rater Agreement



Students respond to 2 prompts in randomized, counterbalanced order.

Report the placement agreement rates for paired prompts.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Example

- Two faculty members were trained in the application of the writing rubric to consistently rate written essays for the English Written Sample Assessment. A random sample of 60 papers were taken and rated by two independent readers. Rater agreement was calculated by correlating the ratings from the two independent raters.

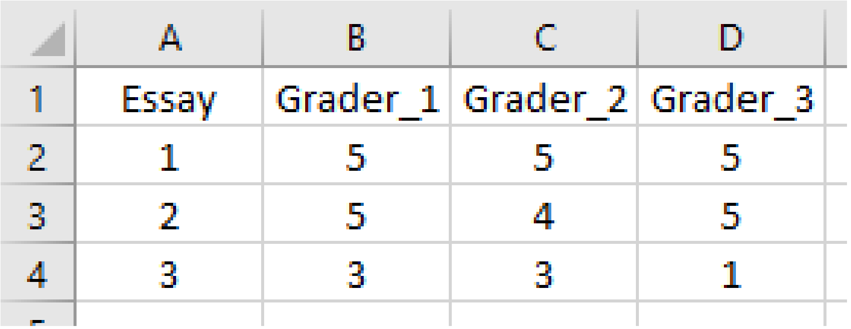
- The resulting correlation between the two raters was .73.

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | College of Education & Human Sciences

Reliability: Intraclass Correlation



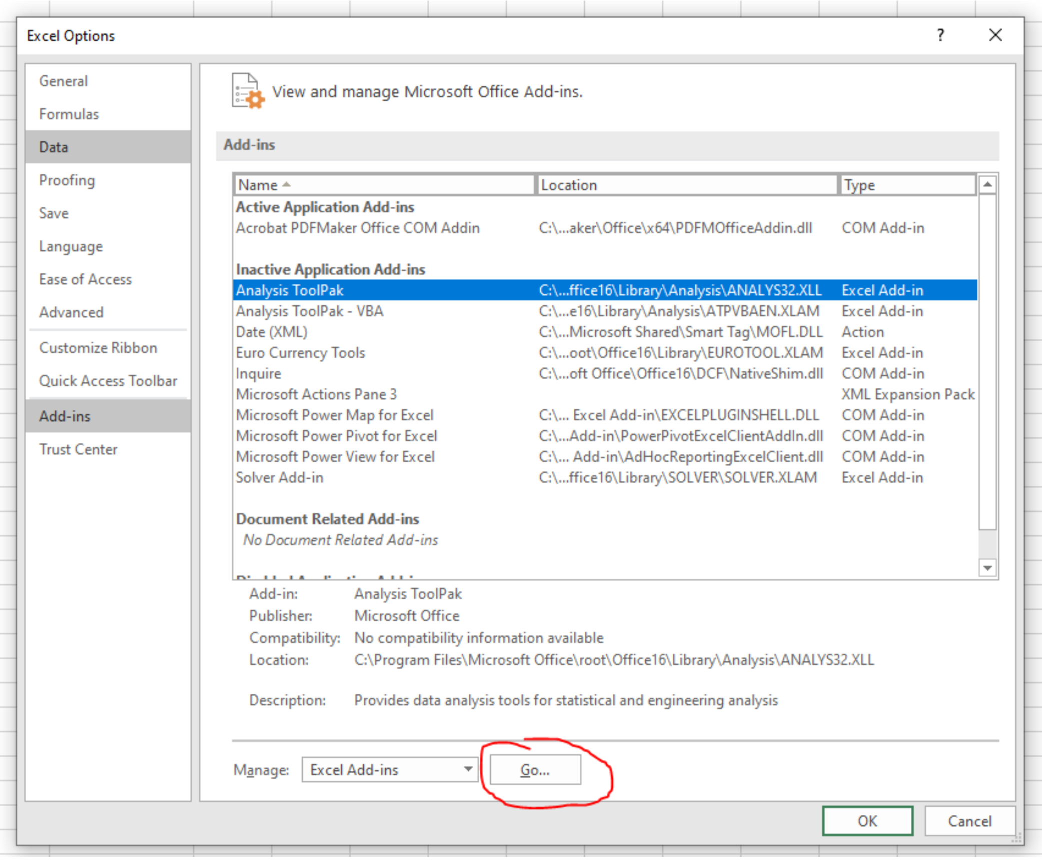
	A	B	C	D
1	Essay	Grader_1	Grader_2	Grader_3
2	1	5	5	5
3	2	5	4	5
4	3	3	3	1

Step 1) Collect Grader/rater ratings of a performance test (e.g., an essay or a presentation).

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Intraclass Correlation (cont.)

Step 2) First, you will need the *Analysis ToolPak* add in for Excel. Go to File --> Options --> Add-ins, then find Analysis ToolPak and click Go...



Excel Options

View and manage Microsoft Office Add-ins.

Add-ins

Name	Location	Type
Active Application Add-ins		
Acrobat PDFMaker Office COM Addin	C:\...aker\Office\x64\PDFMOfficeAddin.dll	COM Add-in
Inactive Application Add-ins		
Analysis ToolPak	C:\...ffice16\Library\Analysis\ANALYS32.XLL	Excel Add-in
Analysis ToolPak - VBA	C:\...e16\Library\Analysis\ATPVBAEN.XLAM	Excel Add-in
Date (XML)	C:\...Microsoft Shared\Smart Tag\MOFL.DLL	Action
Euro Currency Tools	C:\...oot\Office16\Library\EUROTOOL.XLAM	Excel Add-in
Inquire	C:\...oft Office\Office16\DCF\NativeShim.dll	COM Add-in
Microsoft Actions Pane 3		XML Expansion Pack
Microsoft Power Map for Excel	C:\... Excel Add-in\EXCELPLUGINSHELL.DLL	COM Add-in
Microsoft Power Pivot for Excel	C:\...Add-in\PowerPivotExcelClientAddIn.dll	COM Add-in
Microsoft Power View for Excel	C:\... Add-in\AdHocReportingExcelClient.dll	COM Add-in
Solver Add-in	C:\...ffice16\Library\SOLVER\SOLVER.XLAM	Excel Add-in
Document Related Add-ins		
No Document Related Add-ins		

Add-in: Analysis ToolPak
Publisher: Microsoft Office
Compatibility: No compatibility information available
Location: C:\Program Files\Microsoft Office\root\Office16\Library\Analysis\ANALYS32.XLL

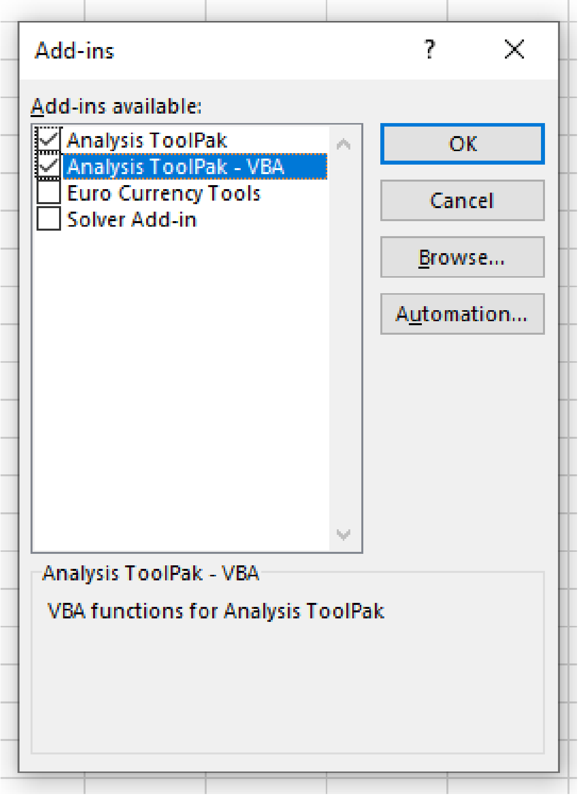
Description: Provides data analysis tools for statistical and engineering analysis

Manage: Excel Add-ins

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Intraclass Correlation (cont.)

Step 3) Make sure Analysis ToolPak and Analysis ToolPak – VBA are select and click OK



Add-ins

Add-ins available:

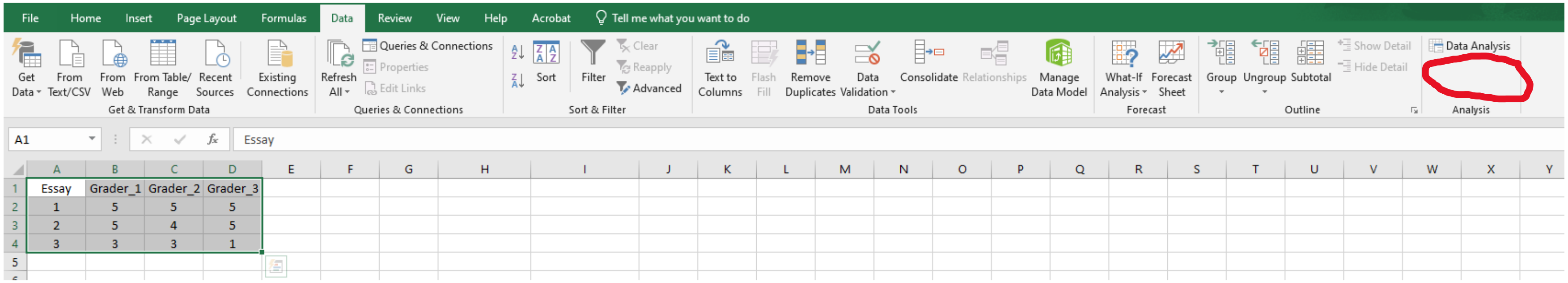
☑ Analysis ToolPak
☑ Analysis ToolPak - VBA
□ Euro Currency Tools
□ Solver Add-in

Analysis ToolPak - VBA

VBA functions for Analysis ToolPak

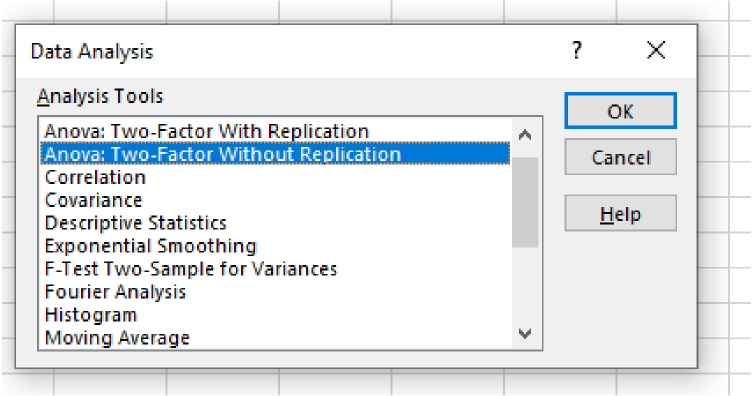
Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Intraclass Correlation (cont.)



Essay	Grader_1	Grader_2	Grader_3
1	5	5	5
2	5	4	5
3	3	3	1

Step 4) Navigate to the data tab then highlight your data. There is a new option in the data tab from the add-in: Data Analysis. Click it.



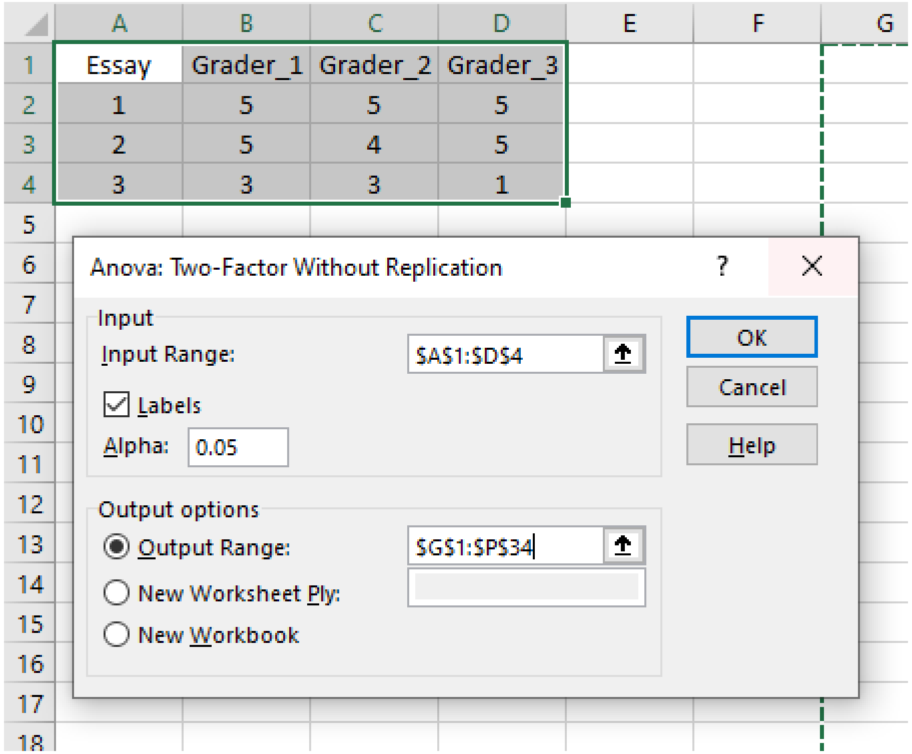
Step 5) Click Anova: Two-Factor Without Replication

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln
COLLEGE OF EDUCATION & HUMAN SCIENCES

Reliability: Intraclass Correlation (cont.)



Essay	Grader_1	Grader_2	Grader_3
1	5	5	5
2	5	4	5
3	3	3	1

Anova: Two-Factor Without Replication

Input

Input Range: $A$1:$D$4

☑ Labels

Alpha: 0.05

Output options

● Output Range: $G$1:$P$34

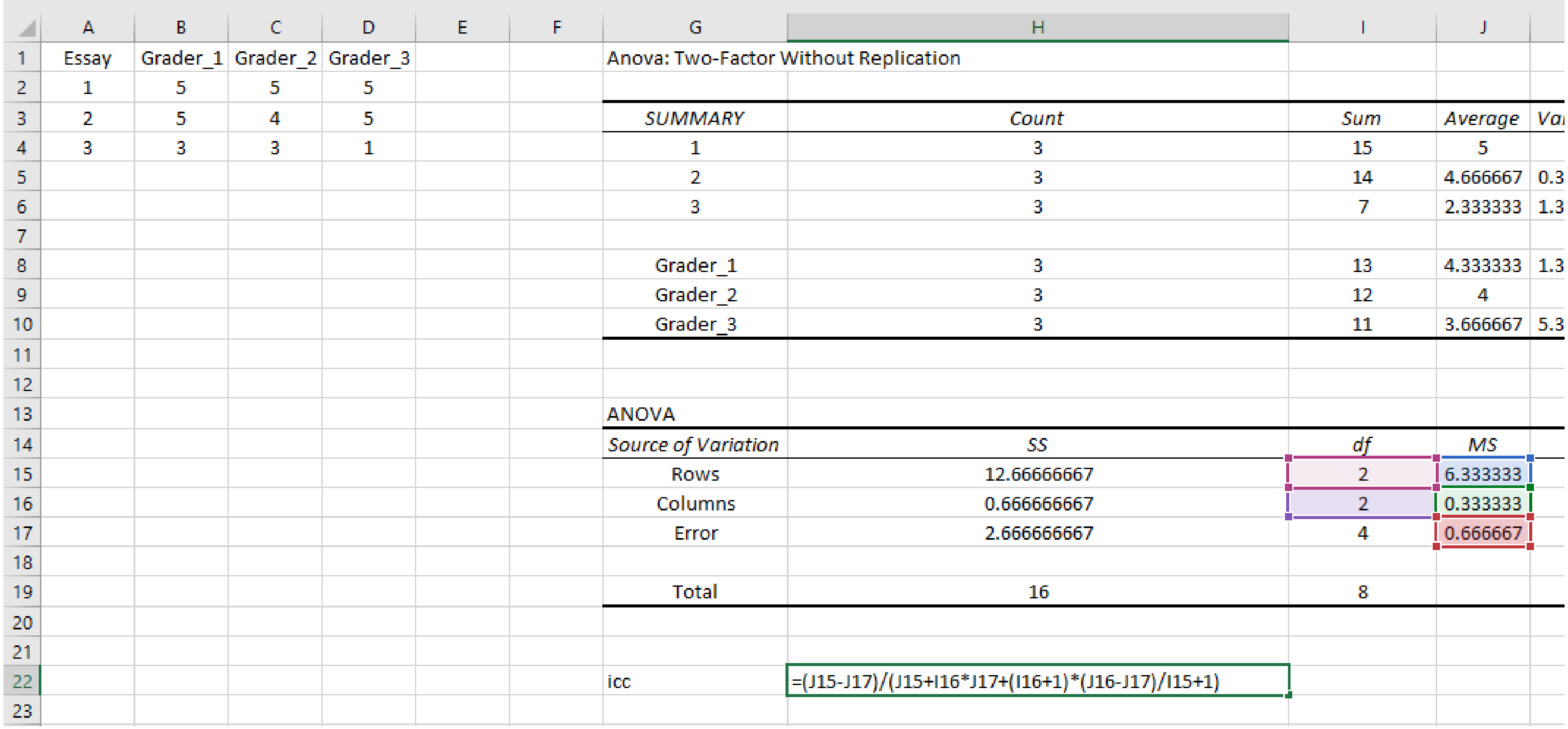
○ New Worksheet Ply:

○ New Workbook

Step 6) Select your data in the input range (if not there by default). Select where you want your output saved. Click OK.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Intraclass Correlation (cont.)



	A	B	C	D	E	F	G	H	I	J	
1	Essay	Grader_1	Grader_2	Grader_3			Anova: Two-Factor Without Replication				
2	1	5	5	5							
3	2	5	4	5			*SUMMARY*	*Count*	*Sum*	*Average*	*Va*
4	3	3	3	1			1	3	15	5	
5							2	3	14	4.666667	0.3
6							3	3	7	2.333333	1.3
7											
8							Grader_1	3	13	4.333333	1.3
9							Grader_2	3	12	4	
10							Grader_3	3	11	3.666667	5.3
11											
12											
13							ANOVA				
14							*Source of Variation*	*SS*	*df*	*MS*	
15							Rows	12.66666667	2	6.333333	
16							Columns	0.666666667	2	0.333333	
17							Error	2.666666667	4	0.666667	
18											
19							Total	16	8		
20											
21											
22							icc	=(J15-J17)/(J15+I16*J17+(I16+1)*(J16-J17)/I15+1)			
23											

Step 7) From the resulting ANOVA table, you can calculate an ICC, as shown above.

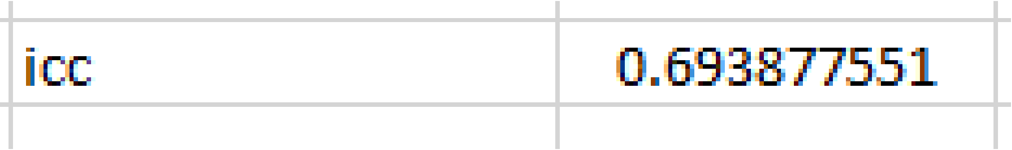
Note. There are multiple types of ICCs.

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln COLLEGE OF EDUCATION & HUMAN SCIENCES

Reliability: Intraclass Correlation (cont.)

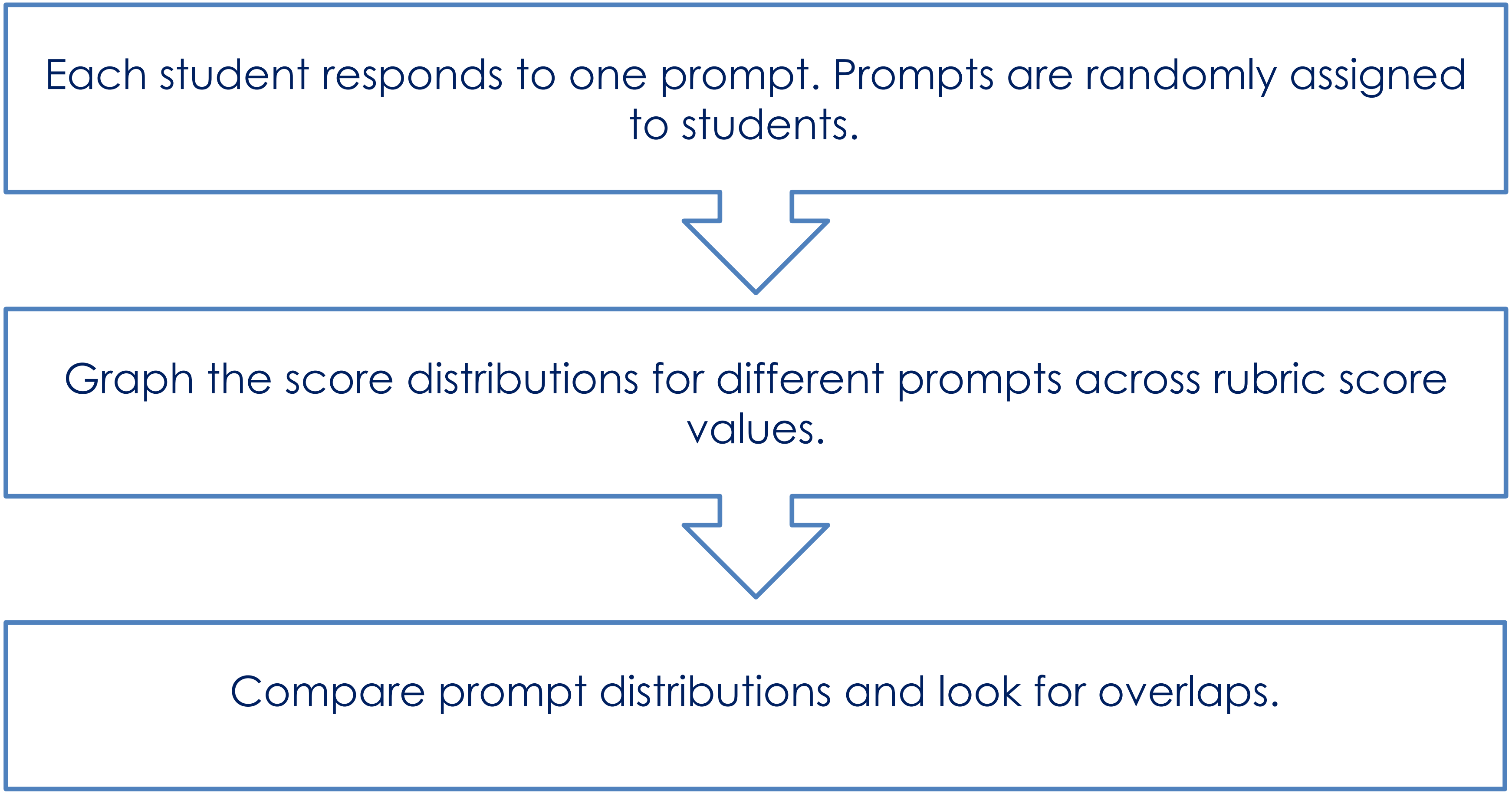


icc	0.693877551

Step 8) Interpret the results. **A threshold of .75 of higher is required for ICCs.**

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability Studies: Interprompt Agreement (Randomized=Alternate)

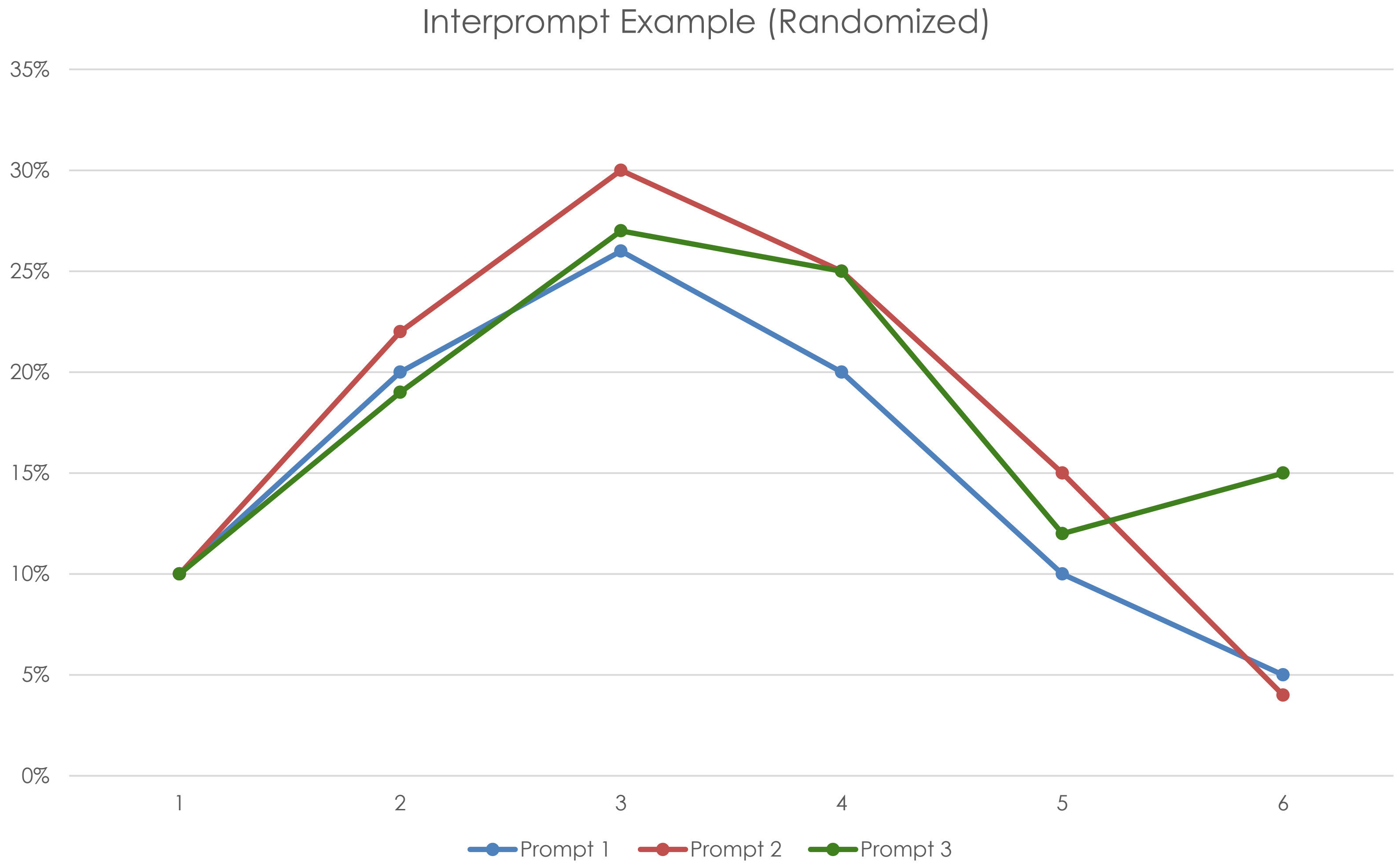


Each student responds to one prompt. Prompts are randomly assigned to students.

Graph the score distributions for different prompts across rubric score values.

Compare prompt distributions and look for overlaps.

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.



Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

Reliability: Analysis Considerations

- Practice effects
- Clarity in item writing
- Diversity (heterogeneity) among group members
- Objectivity in scoring
- Fatigue, differences in motivation
- Differences in test-taking environments (having distractors)

BUROS
CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

UNIVERSITY OF Nebraska Lincoln | COLLEGE OF EDUCATION & HUMAN SCIENCES

Reliability: Errors/Omissions

- Values reported do not meet minimal criteria stated in the Standards.
- Rubric scoring procedures or prompts change but a new study was not conducted
- Equivalency evidence for different prompts is not provided

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | College of Education & Human Sciences

QUESTIONS?

Session 2 Training: Agenda

Thur, Oct 20th 8:30 am – 12 pm

- Fairness – Overview and panel reviews (45 minutes)
- Fairness – Disproportionate impact (45 min)
- Administration considerations (10 minutes)
- Accommodations (10 minutes)
- Scoring considerations (10 minutes)
  - Setting cut scores (50 minutes)
- Next steps (10 minutes)

BUROS CENTER FOR TESTING

Copyright 2022 by University of Nebraska-Lincoln for use by CCCCO. All rights reserved. No part of this work may be used, accessed, reproduced, distributed, or stored in any form or by any means without the written permission of the Buros Center for Testing.

University of Nebraska Lincoln | College of Education & Human Sciences