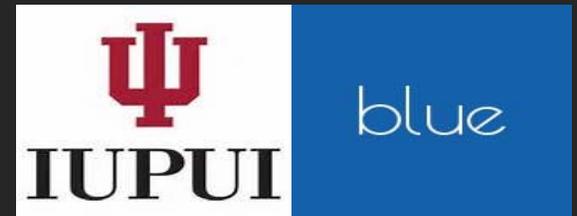**IUPUI and eXplorance**

Assessing the Reliability and Validity of Student Course Evaluations: Ongoing Dialogue Revisited

Presentation given at the *Bluenotes Global 2018 Conference*, Louisville, KY

August, 2018

Presentation Outline

- Quick Facts and Data About IUPUI

- Define Reliability and Validity

- Types of Reliability

- Correlation *vs* Reliability Coefficients

- Demonstrate How to Configure, Add and Use *Reliability Assessment Report Block* in Blue (to enhance the utility of reports)

- Types of Validity

- Consider Suggestions for Addressing Low Response Rates on Questionnaires
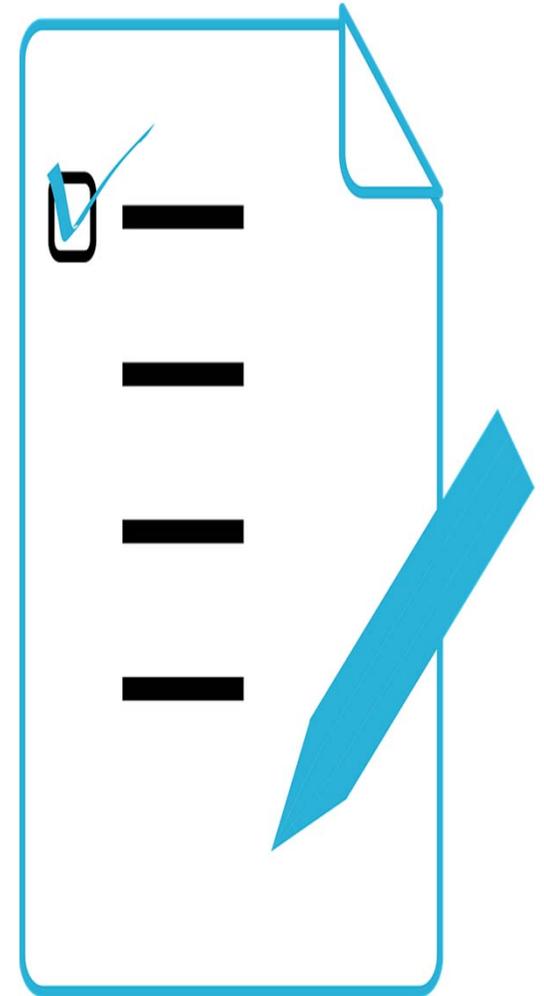
Learning Objectives

❖ Define the concepts of *reliability* and *validity*.

❖ Describe the role of *reliability assessment* in evaluating the technical properties and effectiveness of a questionnaire.

❖ Demonstrate how to set up, customize and use the *Reliability Assessment report block* in Blue.

❖ Interpret the *Reliability Assessment Index*, and describe why it is important to add reliability to reports obtained in Blue.

IUPUI: Quick Facts and Data

- *Indiana University-Purdue University Indianapolis*; Established in 1969

- 29,790 students [21,610 undergraduates and 8,180 graduate/professional students from 146 countries and 49 states in USA; 56% female, 44% male; 27% minorities]

- 18 distinct schools and 2 colleges that confer degrees thru IU & PU

  o 2 Purdue schools (Engineering & Technology, Science)

  o 16 IU schools (Business, Dentistry, Education, Herron Arts & Design, Health & Rehabilitation Science, Informatics & Computing, Law, Liberal Arts, Medicine, Nursing, Philanthropy, Physical Education & Tourism Management, Public & Environmental Affairs, Public Health, Social Work, University Graduate School; and

  o 2 Colleges (Honors College, University College)

- Offers more than 350 undergraduate, graduate, and professional programs

- Strong research focus

- Destination campus for Health and Life Sciences

Brief Introduction

- Questionnaire (online):

  o Important method of data collection used extensively in obtaining student feedback about courses and instructors

- Ideal requisites of a questionnaire:

  o Should be clear and easy to understand

  o Layout is easy to read and pleasant to eye

  o Sequence of questions easy to follow

  o Sensitive questions must be worded exactly

  o <u>Note</u>: For the present context, the terminologies *measuring instrument*, *scale*, *test* and *online survey* represent **questionnaire**; and *item* represents each question in a questionnaire.

Commonly used terms… (continued)

"My car is unreliable."

"You have a valid point."

In scientific research terminology …

" The study findings were not reliable."

"The conclusion of the study was not valid."

Definition of Key Terms

- **Reliability:**

  ✓ Refers to "consistency" of scores or answers provided by an instrument. (Stable or consistent responses)

    o "*determine whether student ratings are consistent enough to be used as a source of evidence for making judgments about teaching effectiveness.*

    o "*the degree to which results obtained by a measurement and procedure can be* replicated."

- Scores or responses obtained from an instrument can be considered reliable but not valid.

    – Reliability is an insufficient foundation for establishing validity

- An instrument should be reliable and valid, depending on the context in which an instrument is used.

Definition of Key Terms (continued)

- **Validity:**
  - ✓ refers to the ability of an instrument to accurately *measure what it is intended to measure*.
  - ✓ Validity refers to whether evidence supports the *interpretation of a score for its **intended purpose**.*
  - ➢ Validity of any assessment depends on <u>proper interpretation</u> and <u>use</u>.

  - ➢ … appropriateness, correctness, meaningfulness, and usefulness of the specific inferences that educators make based on the data they collect.

Reliability and Validity

(a) So unreliable as to be invalid

(b) Fair reliability and fair validity

(c) Fair reliability but invalid

(d) Good reliability but invalid

(e) Good reliability and good validity

The bulls-eye in each target represents the information that is desired. Each dot represents a separate score obtained with the instrument. A dot in the bulls-eye indicates that the information obtained (the score) is the information the researcher desires.

Source: Fraenkel & Wallen, 2006; © The McGraw Hill Companies, Inc.

Types of Reliability

- For researchers, types of reliability include:
  - Test-Retest Reliability (or Stability)
  - Alternate-Form Reliability (or Equivalence)
  - Internal Consistency Reliability (or Homogeneity)
  - Inter-Rater Reliability
  - Intra-Rater Reliability

Reliability measured in aspects of:

Stability
- Done to ensure that same results are obtained when used consecutively for two or more times
- Test-retest methods is used

Internal Consistency
- To ensure all components of an instrument measure the same attribute (Homogeneity)
- Split-half method

Equivalence
- Used when two observers study a single phenomenon simultaneously
- Inter-rater reliability

Correlation coefficient (r_{xy})

- Measures the degree of relationship between two sets of scores (or ratings)

- r_{xy} can range from -1 to +1

- r_{xy} = 0 indicates absence of any relationships

Correlation Coefficient	Strength of Relationship
+/- 0.7 to 1.00	Strong
+/- 0.3 to 0.69	Moderate
+/- 0.0 to 0.29	None to weak

Reliability coefficient (r_{xx})

- Measures the degree of **consistency** among sets of scores (or ratings)

- r_{xx} can range from **0** to **+1** (Note: Positive values are preferred…)

- r_{xx} = **0** indicates lack of consistency among scores

Reliability Coefficient	Strength of Relationship
0.7 to 1.00	Strong
0.3 to 0.69	Moderate
0.0 to 0.29	None to weak

- A reliability coefficient (Cronbach's alpha) of **0.70 or higher** is considered "**acceptable**" (and indicates that the items have *relatively high internal consistency*.)

Reliability Assessment (eXplorance Blue)

- *Reliability assessment* is an indicator of how accurate the results of the report are.

- *Reliability assessment* formula used in Blue is based on the *size of audience* and the *number of responses received*.

- Range of Index (**Blue**): **0 to 2.2**, where …

  - the **closer to 0** the higher the more confident that the results are reliable; and

  - the **closer to 2.2** reflects less confidence that the results are reliable.

Reliability Assessment Configuration

Reliability Assessment Settings (Blue)	
Default Level	**Threshold**
Insufficient	> 0.500
Sufficient	0.500
Good	0.250

Notes:

- Only *positive values* can be used as *threshold values* in Blue.

- *Higher threshold values* are *less valid*.

- A **threshold value** of **0** is considered *perfect*, a *response rate of 100%.*

- The *highest threshold value*, a *response rate of 0%,* is **2.2**

Reliability Assessment in Blue

Which is more reliable?

- 1 response out of 5 (20%),

  or

- 50 responses out of 1000 (5%)?

Background:

- Course evaluation results were used as a piece of evidence for determining academic promotion, tenure, expectations, etc.

- Question was, how 'reliable' are results to inform action?

- Response rates do not indicate size of classroom

Reliability Assessment in Blue

Options available:

- Simple verbiage can be presented on the report alongside the value to represent the level

- Captions can be configured for emphasis (e.g., green for success, red for insufficient)

- Ranges per level can be adjusted/recalibrated

- Different levels can be customized to be used for different reports

- Instructions can be included to indicate how to meet thresholds for their class size in the future

Reliability Assessment in Blue

Help Center Guide

https://support.explorance.com/hc/en-us/articles/115002468583

Illustrative Example of Results...

Reliability Assessment (using Blue "Default Threshold Values")

Class Number	Audience Invited Count (N)	Responses (n)	Response Rate (%)	Reliability Assessment	Comment
23298	35	35	100%	0.00	Good
23602	23	15	65%	0.20	Good
27978	18	10	56%	0.31	Sufficient
25389	16	2	13%	1.36	Insufficient
40200	31	29	94%	0.03	Good
25896	18	4	22%	0.86	Insufficient
23582	65	23	35%	0.30	Sufficient
27633	50	24	48%	0.23	Good
23126	18	7	39%	0.51	Insufficient
36045	23	10	43%	0.39	Sufficient

Reliability Assessment Block Added to Report: Example #1

Instructor	Course	
Joy C Penney	ME 50105-32698 Hybrid Electric Transportation	
Project Audience: 50	Responses Received: 45	Response Ratio: 90%

Reliability Assessment

With 45 responses from a survey population of 50
the data presented in this report is considered to be Good
Number of responses needed to be considered sufficient: 14
Number of responses needed to be considered good: 31

Reliability Assessment Block Added to Report: Example #2

Instructor	Course	
John Doe	MET 34800-26255 Engineering Materials	
Project Audience: 18	Responses Received: 10	Response Ratio: 56%

Reliability Assessment

With 10 responses from a survey population of 18
the data presented in this report is considered to be Sufficient
Number of responses needed to be considered sufficient: 10
Number of responses needed to be considered good: 15

Reliability Assessment Block Added to Report: Example #3

Instructor	Course	
Sara K. Marsh	**ME 34400-22098 Intro to Engineering Materials**	
Project Audience: 90	**Responses Received: 43**	**Response Ratio: 48%**

Reliability Assessment

With 43 responses from a survey population of 90

the data presented in this report is considered to be Good

Number of responses needed to be considered sufficient: 16
Number of responses needed to be considered good: 42

Reliability Assessment Block Added to Report: Example #4

Instructor	Course	
Mary Brown	**MSTE 31200-25294 Business of Motorsports**	
Project Audience: 44	**Responses Received: 13**	**Response Ratio: 30%**

Reliability Assessment

With 13 responses from a survey population of 44

the data presented in this report is considered to be Insufficient
Number of responses needed to be considered sufficient: 14
Number of responses needed to be considered good: 29

Reliability Assessment Block Added to Report: Example #5

| Instructor	Course	
Eve Pressley	ECE 26300-23628 Intro to Computing in Elect Engr.	
Project Audience: 56	Responses Received: 18	Response Ratio: 32%

Reliability Assessment

With 18 responses from a survey population of 56
the data presented in this report is considered to be **Sufficient**
Number of responses needed to be considered sufficient: 15
Number of responses needed to be considered good: 33

Challenges

❖ What challenges related to online student ratings has your University experienced (or would you anticipate experiencing)?
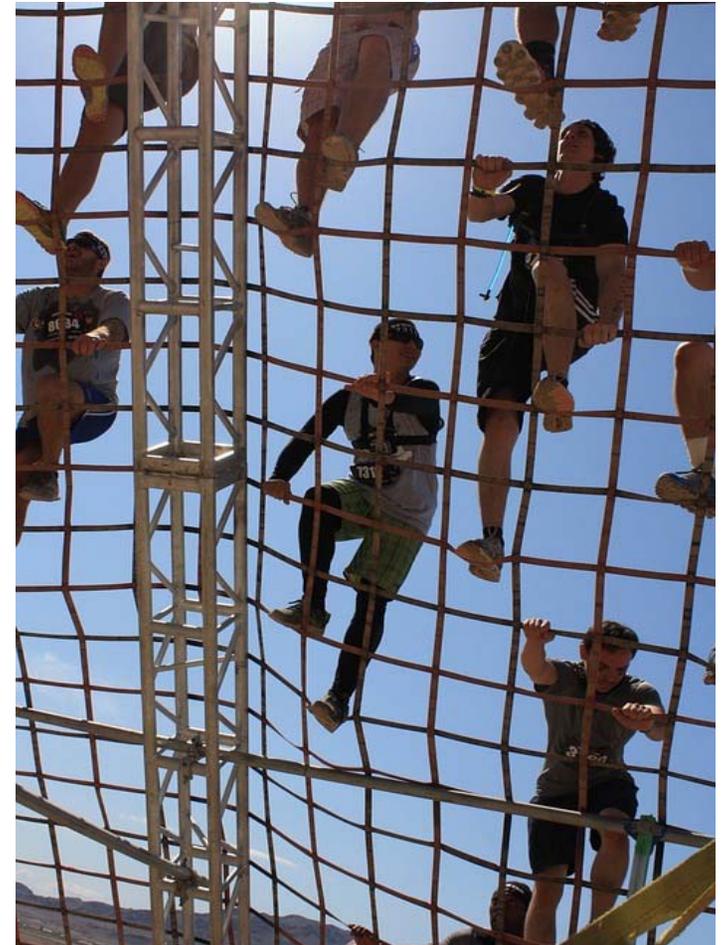
Response Rates

- **Low response rates** for online surveys

- **Non-response bias** might occur...

- Sufficient response rates are important in obtaining **reliability** and **validity** evidence that is acceptable.

Tips to Increase Response Rates for Online Course Evaluations

1. Communicate with your students (to increase engagement)

2. Promote importance & usefulness of your online course evaluations

3. Make it easy for students to provide their feedback (e.g., use uniform positioning of Likert-type response scales)

4. Revisit email subject and customize content

5. Offer incentives to spur course evaluation participation and completion

6. In-class evaluations (allow students to complete the evaluations on their mobile devices)

What you can do if your response rate is low

What you can do if your response rate is low

1. Make sure you have lots of respondents

- Larger numbers of respondents provide more accurate response estimates

What you can do if your response rate is low

2. Look for non-response bias

- Were some students more likely to respond than others?

- How might differences impact results?

What you can do if your response rate is low

3. Weighting survey results

- Numerical adjustment of survey responses

- Statisticians swear by it

- Is it necessary?

Final note: "Quality" matters

Good survey data is accurate, timely, and accessible.
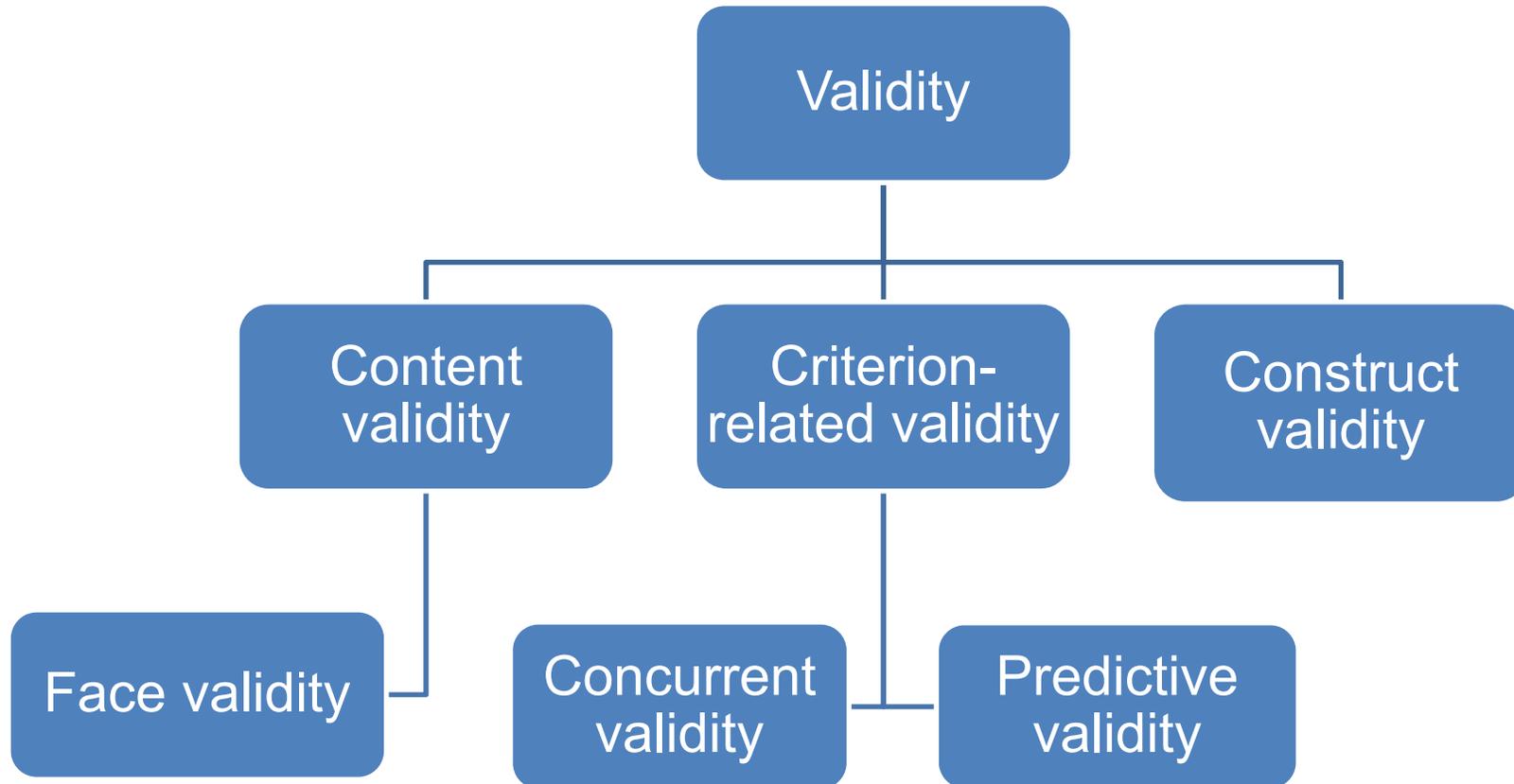
- Think about survey design

Non-response bias is not directly the result of the response rate.

- Who didn't respond to the questionnaire?
- Are they different from those who did respond?

Don't be complacent, but do be aware.

Types of Validity

Summary of Validity

	CONTENT	CRITERION		CONSTRUCT
		CONCURRENT	PREDICTIVE	
What it measures	Whether the instrument covers a representative sample of the domains to be measured	Ability of the measurement instrument to estimate present results or outcomes	Ability of the measurement instrument to predict future results or outcomes	Extent to which the instrument measures a theoretical construct (e.g., teaching effectiveness)
How it is accomplished	Asks experts to assess the instrument to establish that the items are representative of the outcome	Correlate student ratings from the instrument with a concurrent behavior	Correlate student ratings from the measurement scale with behavior in future	Correlate student ratings of instruction with results from an established instrument

Other Related Terminology …

- **Internal Validity**
  - ➢ Refers to validity of the measurement itself
  - ➢ Reflects the extent to which a causal conclusion based on a study is warranted; as evidenced by the extent to which a study minimizes systematic error (or 'bias').

- **External Validity**
  - ➢ Ability to generalize the findings to the target population.

- **Practicality**
  - ➢ "Is the questionnaire or measurement instrument easy to construct, administer, score and interpret?"

Conclusion

- Validated questionnaire:

  ✓ It is a questionnaire that has undergone a validation procedure to show that it accurately measures what it is intended to measure, *regardless of who responds, when they respond, and to whom they respond or when self-administered* and whose reliability has also bee examined, thereby:

    o Reducing bias and ambiguities

    o Better quality of data and credible information

Q & A Session…

Question & Answer

and

Thank You!

Contact Information

❖ **IUPUI Testing Center**

➢ **Howard Mzumara, Ph.D.**

Director, IUPUI Testing Center

Institutional Research & Decision Support, IUPUI

(317) 278-2214

hmzumara@iupui.edu

tc.iupui.edu

❖ **eXplorance Inc.**

➢ **Zelbrey Bedard**

Vice President, Product Management (Blue)

(514) 938-2111 ext.: 226

zbedard@explorance.com

explorance.com