

Item Analysis Made Easy

Presenters

Tim Bristol, PhD, RN, CNE, ANEF, FAADN

Karin J. Sherrill, RN, MSN, CNE, CHSE, ANEF, FAADN



Purpose

- Making decisions about test items after giving an exam can get emotional. One could follow the numbers, listen to the learners, or both in order to make the best decisions. We'll discuss the importance of applying both content validity and item analysis to make decisions that are fair to everyone.

Objectives

- Differentiate content validity from item analysis.
- Discuss quick steps in applying statistical data to improving exams.
- Apply principles of test review to a data set.



Item Analysis Made Easy ANCC Disclosures

Conflict of Interest – This educational activity’s planners and the presenters have indicated they have no bias or conflict of interest.

Successful completion - Criteria for judging successful completion of the webinar are attendance at the entire activity and submission of a completed evaluation tool.

CE - This workshop has been awarded 1 contact hour by NurseTim, Inc. NurseTim, Inc. is accredited as a provider of continuing nursing education by the American Nurses Credentialing Center’s Commission on Accreditation.

For more information, contact help@nursetim.com.



conference



onsite



online

Item Analysis Made Easy

Presenters

Tim Bristol, PhD, RN, CNE, ANEF, FAADN

Karin J. Sherrill, RN, MSN, CNE, CHSE, ANEF, FAADN



Striving for the perfect test...

Two Approaches to Analysis

Approaches

- **Content validity**
 - Peer review
 - Student test review
- **Statistical analysis**
 - Item difficulty
 - Item discrimination

Why?

- One does not evaluate the other
- Avoidance of emotional decisions
- Builds a stronger test that is fair to all students



Two Approaches to Analysis

Content Validity

- Are there typos or unfamiliar terms?
- Is the content appropriate to the level of student?
- Is the response to one item found in another?
- Is there one best response?

Item Analysis

- What cognitive level is the item?
- Do the discriminators perform well?
- What do the testers say about the item?

Polling Question #1

How do you best assess the content validity of your questions?

1. Feedback from students.
2. Feedback from peers.
3. Feedback from students and peers.
4. Review by myself.
5. I don't consistently do this.

National Survey of Test and Exam Creation Practices of Nursing Faculty (2017)

Which item analysis statistics do you use for your test items?

**33.66% of Respondents use
Content Validity**

Let's Try One

A client is returning from the recovery room following abdominal surgery. Upon the client's arrival to the surgical unit, which parameter would be the initial focus of the nurse's assessment?

1. Pulse oximetry reading
 2. Vital signs
 3. Pain in the incision
 4. Status of the dressing
- Does the item make sense?
 - Are there typos or unfamiliar terms?
 - Is the content appropriate?
 - Is the response to one item found in another?
 - Is there one best response?

Let's Try Another

The nurse decides to administer tablets of acetaminophen instead of the intramuscular (I.M.) meperidine she has been giving to one of her orthopedic patients. Why would she do this?

1. The patient asks for it.
 2. The patient has a fever.
 3. The pain is minimal.
 4. He has poor liver function.
- Does the item make sense?
 - Are there typos or unfamiliar terms?
 - Is the content appropriate?
 - Is the response to one item found in another?
 - Is there one best response?

Last One

The nurse determines that an elderly patient is experiencing orthostatic hypotension. After assessing the blood pressure while the patient is supine, the next nursing intervention to promote patient safety for ambulation will be?

1. Ensure that the patient is well hydrated prior to ambulation.
2. Obtain a heart rate (HR), respiratory rate (RR), and temperature (T).
3. Assess the blood pressure 1-3 minutes after the patient has changed position.
4. Dangle the feet over the bedside for 10 minutes before ambulation.

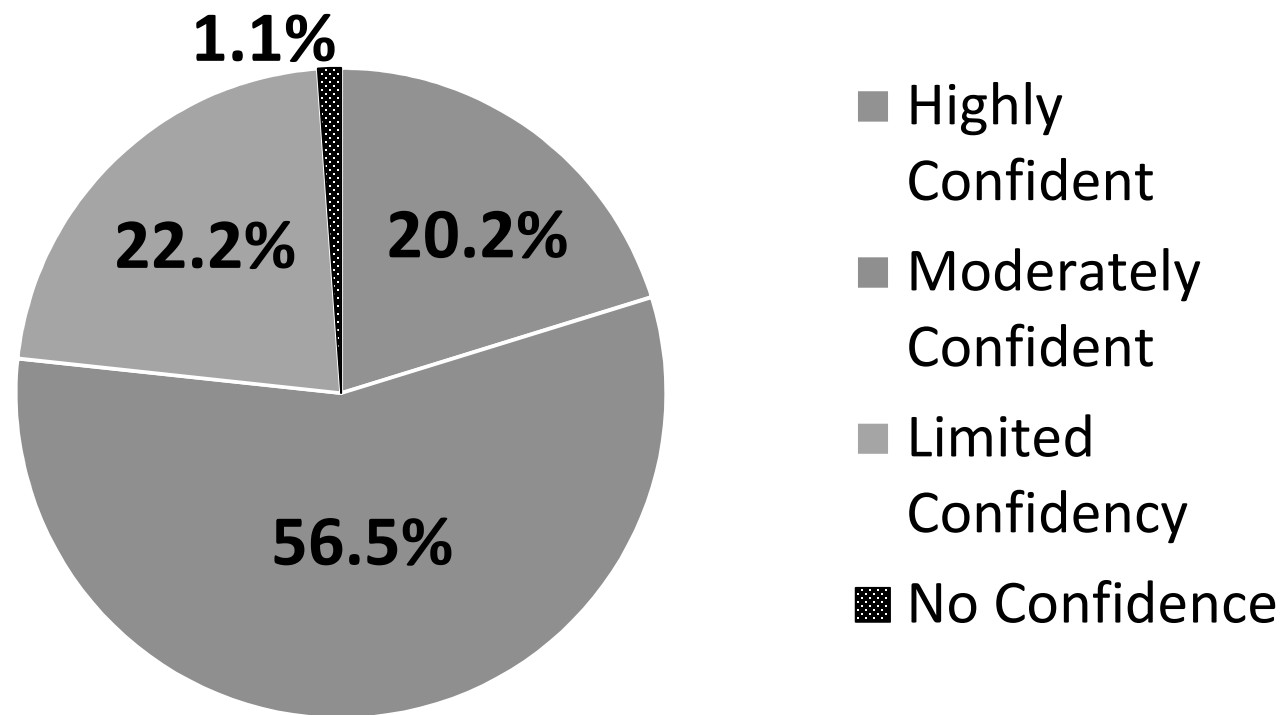
Polling Question #2

How comfortable do you feel with the process of item analysis?

1. I'm not comfortable; that's why I'm here.
2. I can do it with guidance.
3. I'm independent, but lack confidence.
4. I've got it, most the time.

National Survey of Test and Exam Creation Practices of Nursing Faculty (2017)

Which best describes your confidence in analyzing an exam and each item?



Statistical Analysis

- Statistical examination of
 - Test questions
 - Student answers
- Purpose: to assess how individual test items contribute to the overall reliability of the test.

Exam Analysis

- **Exam Difficulty Level**

- Mean
- Median
- Mode

- **Exam Reliability**

- Consistency of exam – repeatable
- Measured with KR-20

Reliability: KR - 20 Interpretation

- Range 0.0 to +1.0
- 1.0 is perfect reliability
- 0.0 has no reliability

| Reliability Coefficient | Interpretation |
|-------------------------|-----------------------|
| 0.90 and up | Excellent |
| 0.80-0.89 | Good |
| 0.70-0.79 | Adequate |
| Less than 0.70 | Limited Applicability |

***Numbers may vary based on resource used.*

Item Analysis

- **Item Difficulty**

- % of students getting the item correct
- ***p value***
 - Good = 0.30 to 0.80
 - Ideal = 0.70 to 0.80

- **Item Discrimination** –best performers get the item right and the poor performers get the item wrong

- ***point biserial (PBI or PBCC)*** - correlation between item score and test score
 - Best students get right – positive
 - Best students get wrong – negative
 - Item < 0.15 should be revised

***Numbers may vary based on resource used.*

Total Possible Points: 45.0
 Students in this group: 65
 Median Score: 36
 Mean Score: 35.2
 High Score: 42
 Low Score: 24

Step 1:
 Analyze
 the Exam
 as a whole

| No. | Correct Group Responses | | | Point Biserial | Correct Answer | Response Frequencies - * indicate | | | | |
|-----|-------------------------|-----------|-----------|-------------------|-------------------|-----------------------------------|-----|-----|-----|-----|
| | Total | Upper 27% | Lower 27% | | | A | B | C | D | E |
| 1 | 60.00% | 60.00% | 60.00% | -0.04 | B+C+D+E | 0 | *19 | *17 | *17 | *18 |
| 2 | 50.00% | 60.00% | 20.00% | 0.37 | C | 1 | 9 | *10 | 0 | 0 |
| 3 | 70.00% | 100.00% | 40.00% | 0.30 | B | 0 | *14 | 0 | 6 | 0 |
| 4 | 60.00% | 100.00% | 20.00% | 0.65 | B | 6 | *12 | 0 | 2 | 0 |
| 5 | 75.00% | 100.00% | 100.00% | -0.01 | D | 1 | 3 | 1 | *15 | 0 |
| 6 | 60.00% | 100.00% | 20.00% | 0.58 | B | 0 | *12 | 4 | 4 | 0 |
| 7 | 100.00% | 100.00% | 100.00% | 0.00 | C | 0 | 0 | *20 | 0 | 0 |
| 8 | 35.00% | 60.00% | 0.00% | 0.45 | A+D+E | *17 | 3 | 6 | *19 | *17 |
| 9 | 75.00% | 100.00% | 60.00% | 0.47 | A | *16 | 0 | 5 | 1 | 1 |
| 10 | 95.00% | 100.00% | 80.00% | 0.33 | D | 0 | 0 | 1 | *19 | 0 |
| 11 | 85.00% | 100.00% | 80.00% | 0.20 | D | 1 | 0 | 2 | *17 | 0 |
| 12 | 85.00% | 100.00% | 80.00% | 0.09 | D | 0 | 2 | 1 | *17 | 0 |
| 13 | 65.00% | 100.00% | 40.00% | 0.60 | B | 2 | *13 | 2 | 3 | 0 |
| 14 | 60.00% | 100.00% | 60.00% | 0.45 | D | 4 | 0 | 4 | *12 | 0 |
| 15 | 65.00% | 100.00% | 0.00% | 0.73 | B | 2 | *13 | 1 | 4 | 0 |

Total Possible Points: 45.0

Students in this group: 65

Median Score: 36

Mean Score: 35.2

High Score: 42

Low Score: 24

Step 2:
Analyze the
Item
Difficulty &
Identify the
p-values
< 30% or
> 80%

| No. | Correct Group Responses | | | Point Biserial | Correct Answer | Response Frequencies - * indicate | | | | |
|-----|-------------------------|-----------|-----------|----------------|----------------|-----------------------------------|-----|-----|-----|-----|
| | Total | Upper 27% | Lower 27% | | | A | B | C | D | E |
| 1 | 60.00% | 60.00% | 60.00% | -0.04 | B+C+D+E | 0 | *19 | *17 | *17 | *18 |
| 2 | 50.00% | 60.00% | 20.00% | 0.37 | C | 1 | 9 | *10 | 0 | 0 |
| 3 | 70.00% | 100.00% | 40.00% | 0.30 | B | 0 | *14 | 0 | 6 | 0 |
| 4 | 60.00% | 100.00% | 20.00% | 0.65 | B | 6 | *12 | 0 | 2 | 0 |
| 5 | 75.00% | 100.00% | 100.00% | -0.01 | D | 1 | 3 | 1 | *15 | 0 |
| 6 | 60.00% | 100.00% | 20.00% | 0.58 | B | 0 | *12 | 4 | 4 | 0 |
| 7 | 100.00% | 100.00% | 100.00% | 0.00 | C | 0 | 0 | *20 | 0 | 0 |
| 8 | 35.00% | 60.00% | 0.00% | 0.45 | A+D+E | *17 | 3 | 6 | *19 | *17 |
| 9 | 75.00% | 100.00% | 60.00% | 0.47 | A | *16 | 0 | 5 | 1 | 1 |
| 10 | 95.00% | 100.00% | 80.00% | 0.33 | D | 0 | 0 | 1 | *19 | 0 |
| 11 | 85.00% | 100.00% | 80.00% | 0.20 | D | 1 | 0 | 2 | *17 | 0 |
| 12 | 85.00% | 100.00% | 80.00% | 0.09 | D | 0 | 2 | 1 | *17 | 0 |
| 13 | 65.00% | 100.00% | 40.00% | 0.60 | B | 2 | *13 | 2 | 3 | 0 |
| 14 | 60.00% | 100.00% | 60.00% | 0.45 | D | 4 | 0 | 4 | *12 | 0 |
| 15 | 65.00% | 100.00% | 0.00% | 0.73 | B | 2 | *13 | 1 | 4 | 0 |

Total Possible Points: 45.0

Students in this group: 65

Median Score: 36

Mean Score: 35.2

High Score: 42

Low Score: 24

**Step 3: Review
the Item
Discrimination
& Identify the
PBCC
<0.15**

| No. | Correct Group Responses | | | Point Biserial | Correct Answer | Response Frequencies - * indicat | | | | |
|-----|-------------------------|-----------|-----------|-------------------|-------------------|----------------------------------|-----|-----|-----|-----|
| | Total | Upper 27% | Lower 27% | | | A | B | C | D | E |
| 1 | 60.00% | 60.00% | 60.00% | -0.04 | B+C+D+E | 0 | *19 | *17 | *17 | *18 |
| 2 | 50.00% | 60.00% | 20.00% | 0.37 | C | 1 | 9 | *10 | 0 | 0 |
| 3 | 70.00% | 100.00% | 40.00% | 0.30 | B | 0 | *14 | 0 | 6 | 0 |
| 4 | 60.00% | 100.00% | 20.00% | 0.65 | B | 6 | *12 | 0 | 2 | 0 |
| 5 | 75.00% | 100.00% | 100.00% | -0.01 | D | 1 | 3 | 1 | *15 | 0 |
| 6 | 60.00% | 100.00% | 20.00% | 0.58 | B | 0 | *12 | 4 | 4 | 0 |
| 7 | 100.00% | 100.00% | 100.00% | 0.00 | C | 0 | 0 | *20 | 0 | 0 |
| 8 | 35.00% | 60.00% | 0.00% | 0.45 | A+D+E | *17 | 3 | 6 | *19 | *17 |
| 9 | 75.00% | 100.00% | 60.00% | 0.47 | A | *16 | 0 | 5 | 1 | 1 |
| 10 | 95.00% | 100.00% | 80.00% | 0.33 | D | 0 | 0 | 1 | *19 | 0 |
| 11 | 85.00% | 100.00% | 80.00% | 0.20 | D | 1 | 0 | 2 | *17 | 0 |
| 12 | 85.00% | 100.00% | 80.00% | 0.09 | D | 0 | 2 | 1 | *17 | 0 |
| 13 | 65.00% | 100.00% | 40.00% | 0.60 | B | 2 | *13 | 2 | 3 | 0 |
| 14 | 60.00% | 100.00% | 60.00% | 0.45 | D | 4 | 0 | 4 | *12 | 0 |
| 15 | 65.00% | 100.00% | 0.00% | 0.73 | B | 2 | *13 | 1 | 4 | 0 |

Total Possible Points: 45.0

Students in this group: 65

Median Score: 36

Mean Score: 35.2

High Score: 42

Low Score: 24

Step 4:
Fix Non-
Distractors

| No. | Correct Group Responses | | | Point Biserial | Correct Answer | Response Frequencies - * indicat | | | | |
|-----|-------------------------|-----------|-----------|-------------------|-------------------|----------------------------------|-----|-----|-----|-----|
| | Total | Upper 27% | Lower 27% | | | A | B | C | D | E |
| 1 | 60.00% | 60.00% | 60.00% | -0.04 | B+C+D+E | 0 | *19 | *17 | *17 | *18 |
| 2 | 50.00% | 60.00% | 20.00% | 0.37 | C | 1 | 9 | *10 | 0 | 0 |
| 3 | 70.00% | 100.00% | 40.00% | 0.30 | B | 0 | *14 | 0 | 6 | 0 |
| 4 | 60.00% | 100.00% | 20.00% | 0.65 | B | 6 | *12 | 0 | 2 | 0 |
| 5 | 75.00% | 100.00% | 100.00% | -0.01 | D | 1 | 3 | 1 | *15 | 0 |
| 6 | 60.00% | 100.00% | 20.00% | 0.58 | B | 0 | *12 | 4 | 4 | 0 |
| 7 | 100.00% | 100.00% | 100.00% | 0.00 | C | 0 | 0 | *20 | 0 | 0 |
| 8 | 35.00% | 60.00% | 0.00% | 0.45 | A+D+E | *17 | 3 | 6 | *19 | *17 |
| 9 | 75.00% | 100.00% | 60.00% | 0.47 | A | *16 | 0 | 5 | 1 | 1 |
| 10 | 95.00% | 100.00% | 80.00% | 0.33 | D | 0 | 0 | 1 | *19 | 0 |
| 11 | 85.00% | 100.00% | 80.00% | 0.20 | D | 1 | 0 | 2 | *17 | 0 |
| 12 | 85.00% | 100.00% | 80.00% | 0.09 | D | 0 | 2 | 1 | *17 | 0 |
| 13 | 65.00% | 100.00% | 40.00% | 0.60 | B | 2 | *13 | 2 | 3 | 0 |
| 14 | 60.00% | 100.00% | 60.00% | 0.45 | D | 4 | 0 | 4 | *12 | 0 |
| 15 | 65.00% | 100.00% | 0.00% | 0.73 | B | 2 | *13 | 1 | 4 | 0 |

Total Possible Points: 45.0

Students in this group: 65

Median Score: 36

Mean Score: 35.2

High Score: 42

Low Score: 24

Step 5:
Look at
Individual
Items and
Rewrite Them

| No. | Correct Group Responses | | | Point Biserial | Correct Answer | Response Frequencies - * indicat | | | | |
|-----|-------------------------|-----------|-----------|-------------------|-------------------|----------------------------------|-----|-----|-----|-----|
| | Total | Upper 27% | Lower 27% | | | A | B | C | D | E |
| 1 | 60.00% | 60.00% | 60.00% | -0.04 | B+C+D+E | 0 | *19 | *17 | *17 | *18 |
| 2 | 50.00% | 60.00% | 20.00% | 0.37 | C | 1 | 9 | *10 | 0 | 0 |
| 3 | 70.00% | 100.00% | 40.00% | 0.30 | B | 0 | *14 | 0 | 6 | 0 |
| 4 | 60.00% | 100.00% | 20.00% | 0.65 | B | 6 | *12 | 0 | 2 | 0 |
| 5 | 75.00% | 100.00% | 100.00% | -0.01 | D | 1 | 3 | 1 | *15 | 0 |
| 6 | 60.00% | 100.00% | 20.00% | 0.58 | B | 0 | *12 | 4 | 4 | 0 |
| 7 | 100.00% | 100.00% | 100.00% | 0.00 | C | 0 | 0 | *20 | 0 | 0 |
| 8 | 35.00% | 60.00% | 0.00% | 0.45 | A+D+E | *17 | 3 | 6 | *19 | *17 |
| 9 | 75.00% | 100.00% | 60.00% | 0.47 | A | *16 | 0 | 5 | 1 | 1 |
| 10 | 95.00% | 100.00% | 80.00% | 0.33 | D | 0 | 0 | 1 | *19 | 0 |
| 11 | 85.00% | 100.00% | 80.00% | 0.20 | D | 1 | 0 | 2 | *17 | 0 |
| 12 | 85.00% | 100.00% | 80.00% | 0.09 | D | 0 | 2 | 1 | *17 | 0 |
| 13 | 65.00% | 100.00% | 40.00% | 0.60 | B | 2 | *13 | 2 | 3 | 0 |
| 14 | 60.00% | 100.00% | 60.00% | 0.45 | D | 4 | 0 | 4 | *12 | 0 |
| 15 | 65.00% | 100.00% | 0.00% | 0.73 | B | 2 | *13 | 1 | 4 | 0 |

Example #1

| Option | % Right (p Value) | PBI |
|----------|-------------------|-------------|
| 1 | 0.10 | -0.28 |
| 2 | 0.05 | -0.12 |
| 3 | 0.30 | -0.36 |
| 4 | 0.55 | 0.62 |

- N = 52
- 55% picked the correct response.
- All distractors had some hits.

Example #1

| Option | % Right (p Value) | PBI |
|----------|-------------------|-------------|
| 1 | 0.10 | -0.28 |
| 2 | 0.05 | -0.12 |
| 3 | 0.30 | -0.36 |
| 4 | 0.55 | 0.62 |

- N = 52
- High achievers chose the correct response more than low achievers.
- We know this because all distractors had negative PBI.

Example #2

| Option | % Right | PBI |
|----------|-------------|-------------|
| 1 | 0.35 | 0.48 |
| 2 | 0.10 | -0.24 |
| 3 | 0.15 | -0.04 |
| 4 | 0.40 | -0.29 |

- N = 20
- 35% picked the correct response & 40% picked a incorrect response.
- All distractors had some hits.

Example #2

| Option | % Right | PBI |
|----------|-------------|-------------|
| 1 | 0.35 | 0.48 |
| 2 | 0.10 | -0.24 |
| 3 | 0.15 | -0.04 |
| 4 | 0.40 | -0.29 |

- N = 20
- High achievers picked the correct response more than low achievers.
- All distractors had negative PBI.

What happened here?

| Option | % Right | PBI |
|----------|-------------|-------------|
| 1 | 0.75 | 0.36 |
| 2 | 0.00 | - |
| 3 | 0.25 | - 0.36 |
| 4 | 0.00 | - |

- N = 28

What happened here?

| Option | % Right | PBI |
|----------|-------------|---------------|
| 1 | 0.25 | 0.26 |
| 2 | 0.10 | -0.05 |
| 3 | 0.65 | - 0.16 |
| 4 | 0.00 | - |

- N = 72
- Versus N = 12

Black: Invalid question, adjust per department policy and rewrite. **Dark Gray:** Needs revision: furthest from Ideal.

Light Gray: Consider revision – approaching ideal. **White:** Ideal for discrimination & level of difficulty

| | Point Biserial > 0.30 | Point Biserial 0.20 – 0.29 | Point Biserial 0.15 – 0.19 | Point Biserial 0.10 – 0.14 | Point Biserial < 0.09 | Point Biserial Negative | Total |
|---|--------------------------|-------------------------------|-------------------------------|-------------------------------|--------------------------|----------------------------|-------|
| P Value 0-30% | | | | | | | |
| P Value: Very Difficult: 31-50% | | | | | | | |
| P Value: Difficult: 51-62% | | | | | | | |
| P Value: Average: 63-80% | | | | | | | |
| P Value: Easy: 81-95% | | | | | | | |
| P Value: Very Easy if not predesignated as Mastery: 96-100% | | | | | | | |
| Total Number of Items | | | | | | | |

Analysis Considerations

- Distractors
 - Make all plausible (right and wrong answers).
 - At least 1 student should select every distractor.
- Consider the number of test takers
- Consider the history of the specific item over time
- Consider the variables impacting the current testers
- Should everyone achieve 100% - ever?
 - In academia, not desirable and lacks grade differentiation
 - In practice, is desirable to demonstrate competency

What is a Mastery Item?



What if Stats are Bad?

- **Action #1**: Give credit for more than one answer choice. (this is the preferred action).
 - Response: Many student's scores will rise.
- **Action #2**: Nullify the test item by giving credit for all choices.
 - Response: Total # of items remain same; lower performers scores will go up, better performers will stay the same.
- **Action #3**: Delete the test item from the exam and recalculate with one less total item.
 - Total # of items drops, poorer students scores will go up, stronger students scores will go down.



Putting it all together...

The client's vital signs are temperature (oral) 101°F (38.5°C), heart rate 80 beats per minute, respiratory rate 16 breaths per minute, and blood pressure 128/80 mmHg. Which action should the nurse take next?

1. Ask the patient if he has had a lot to drink in the last few minutes.
2. Notify the primary care provider of the patient's temperature.
3. Ask the patient if he is feeling chilled.
4. Take the temperature by a different route.



The Stats

| Option | % Right | PB |
|----------|-------------|-------------|
| 1 | 0.25 | 0.48 |
| 2 | 0.25 | -0.24 |
| 3 | 0.15 | -0.04 |
| 4 | 0.35 | -0.29 |

- N=40
- 25% (stronger students) selected correct response (1), despite the typo.
- Not many chose (3); most selected (4). Determine why and consider adjustments.

Reliability

Increases with

- More testers
- More items
- Variable Item performance
- Fewer content topics
- Consistent testing environment

Decreases with

- Fewer testers
- Fewer items
- Easy Items
- More content topics
- Distracting testing environment



Reliability & Validity

A test can be reliable without being valid

YET

A test cannot be valid unless it is reliable.

Good stats happen without content validity

YET

Content validity doesn't matter without good stats.



Practice Makes Perfect

Developing valid and reliable tests requires

- Practice
- Support through peer review and statistical software
- Consistency in application of policies and standards
- Integrity and work
- Practice!

Tim Bristol: tim@nursetim.com

Karin J. Sherrill: ksherrill@nursetim.com