



Are Your Multiple-Choice Tests “FIT”?

Using the Fairness of Items Tool (FIT) as a Component of the Test Development Process

Nikole Anderson Hicks, PhD, RNC, CNE

Background

- Multiple-choice (MC) examinations are common
- Conclusions based on MC examinations have high stakes consequences
- Developing well-constructed test items is difficult & time consuming
- Nurse educators lack adequate preparation & sufficient time
- Use of textbook test bank items in examinations is common
- Violations of item-writing guidelines in nursing examinations are common

Theoretical Foundations

- High quality test items are necessary for reliable, valid, discriminating, and unbiased assessments of student learning.
- Item quality is improved through item-writing procedures, obtaining pretest reliability data, and using post-administration analysis data to guide revision.
- Test quality is improved through adequate planning of assessments & developing a test blueprint.

Framework for Quality Assessment

Every effective assessment must meet the following criteria:

- **Valid** – measures what it is designed to measure
- **Reliable** – consistently measures what it is designed to measure
- **Discriminating** – distinguishes between the more knowledgeable & less knowledgeable students
- **Practical** – useful & practical for its purposes
- **Unbiased** – fair to examinees & contains items that students of equal ability are equally likely to answer correctly



Review of Literature

- Interventions that improve item writing:
 - Faculty education & practice
 - Using pre-established guidelines
 - Peer review process
 - Revising items through linguistic modification



Methodology

- Phase 1 – Developing the Fairness of Items Tool (FIT)
- Phase 2 – Expert Review & Validation
- Phase 3 – Use of the FIT by Nursing Faculty

This research study received approval from the Institutional Review Boards at University of Northern Colorado and the University of Cincinnati.

Phase 2 Results

Validity Index	Review 1	Review 2
S-CVI	.90	.988
S-CVI/UA	.63	.97
Face Validity	1.0	.92
Proportion Relevant	Expert 1 = .93	Expert 1 = 1.0
	Expert 2 = .98	Expert 2 = 1.0
	Expert 3 = .73	Expert 3 = .97
	Expert 4 = .90	Expert 4 = .97
	Expert 5 = .98	
ACP	.90	.99

Notes:

S-CVI Scale item content validity index

S-CVI/UA Universal calculation method for the scale
item content validity index

ACP Average congruency percentage

Fairness of Items Tool (FIT)

Evaluate the Stem

1. Use a question format.
2. Eliminate extraneous words (e.g., *of the following*).
3. Present a single, clearly defined question with the problem in the stem.
4. Avoid negatively phrased questions, double negatives, and the use of *except*.
5. Use active verbs and present tense.
6. Write questions at the application or above cognitive level.
7. Write questions that require multilogical thinking (require knowledge of more than one fact/concept).
8. Make sure content is current.
9. Avoid testing student opinions (e.g., use *nurse* instead of *you* as the subject).
10. Test important content and avoid trivia.

Evaluate the Options

11. Make sure options are similar grammatically and in length and amount of detail.
12. Avoid none-of-the-above and all-of-the-above. Use three options instead.
13. Avoid negatively phrased options.
14. Avoid repeating material in the options – move repetitive words to the stem.
15. Avoid repeating words in the stem and correct option.
16. Avoid overlapping options.
17. Eliminate multiple-multiples.
18. Make sure all distracters are plausible.
19. If the stem asks what should be done first or which action is best, all options must be correct with only one option being the first or best.
20. Make sure there is only one correct answer.
21. Write options that require a high level of discrimination to select the correct answer.

Linguistic/Structural Bias

22. Use a parsimonious style and short simple sentences.
23. Use correct grammar, punctuation, capitalization, and spelling.
24. Use precise terms (avoid frequently, appropriate).
25. Avoid absolute terms (always, never, all).
26. Use straight-forward, uncomplicated language. Test nursing content, not vocabulary or reading.
27. Write items that can be comprehended on the first reading. Avoid tricky or misleading items.
28. Ensure that items are independent of each other.
29. Be specific and clear with directions.
30. Use consistent spacing, question numbering/lettering, page numbering. Make sure options appear on the same page as the question.

Cultural Bias

31. Avoid dominant culture (literature, music, movies, sports, foods) unless essential to safe, effective nursing practice.
32. Eliminate all names.
33. Eliminate all slang.
34. Use terminology from textbook, notes, and common words (home vs. abode).
35. Eliminate humor.
36. Avoid stereotyping and over-representation of cultural groups.
37. Use gender-specific language only when necessary to test nursing content.
38. Present the person first, not the diagnosis.

Phase 3 Results

Equivalence – Independence of Scores

- 1,190 values tested ($p < .05$)
- 95.5% demonstrated independence ($n = 1,136$)

Stability – Split-half Reliability

- KR-20 = .799 ($\alpha = .05$)

Equivalence – Interrater Agreement

Interpretation of Agreement	<i>n</i> (%)
Perfect Agreement (.9-1.0)	240 (46.8%)
Excellent Agreement (.8-.89)	100 (19.5%)
Very Good Agreement (.7-.79)	76 (14.8%)
Good Agreement (.6-.69)	47 (9.2%)
Fair Agreement (.5-.59)	48 (9.4%)
Poor Agreement (Below .5)	2 (0.4%)

Phase 3 Results

Construct Validity – Known Groups Comparison

Level	Known Biased Item(s)	Known Fair Item(s)
Guideline	2.7 +/- 2.4	0.92 +/- 1.4
Dimension (ST)	.29 +/- .41	.1 +/- .13
Test Item	7.84 +/- 4.8	2.76 +/- 2.84

Internal Consistency Reliability – Cronbach's alpha

Test Item	α	<i>n</i>
B-18	.737	67
B-13	.73	70
B-1	.706	66
B-11	.694	80
B-35	.651	65
F-10	.598	67

Recommendations

- Use a systematic process for test development that incorporates the FIT for developing high quality MC test items.
- Use the FIT to develop item banks of quality MC test items to save time in test development.
- Incorporate the FIT in faculty development. Repetition & practice will lead to improvement in MC test items.
- Develop a test review process with faculty peer review using the FIT.
- Three-option items should be implemented as a standard alternative in nursing education.

Implications

- Improving the quality of MC test items used in nursing examinations has the potential to **improve student success** and better prepare all nursing students for licensure and certification examinations.
- Indirectly, the FIT has the potential to increase the **quality, quantity, and diversity of nurses** joining the workforce. These improvements in student success also have a positive impact on nursing program accreditation rates and ability to recruit high quality students.
- Improving student success **benefits faculty** with improved evaluations and less time devoted to remediating students who are performing poorly on examinations containing biased test items.

Acknowledgements

- This research study used Research Electronic Database Capture (REDCap), a secure, web-based survey tool and database, supported by Center for Clinical and Translational Science and Training grant UL1-RR026314
- Thank you to the following individuals who contributed to the development of this research study:
 - Dr. Janice Hayes, Dr. Faye Hummel, Dr. Vicki Wilson, and Dr. Lisa Rue ~ University of Northern Colorado
 - Dr. Jun Ying ~ University of Cincinnati
 - Dr. Kathleen LaSala ~ University of South Carolina

References

- Morrison, S., Nibert, A., & Flick, J. (2006). *Critical thinking and test item writing* (2nd ed.). Houston, TX: Health Education Systems, Inc.
- Quinn, F. M. (2000). *The principles and practice of nurse education* (4th ed.). Cheltenham, United Kingdom: Stanley Thornes.
- Schueneman, J. D. (1984). A theoretical framework for the exploration of causes and effects of bias in testing. *Educational Psychologist*, 19(4), 219-225.
- Tarrant, M., & Ware, J. (2012). A framework for improving the quality of multiple-choice assessments. *Nurse Educator*, 37(3), 98-104.
doi:10.1097/NNE.obo13e31825041do