Evaluating the use of ‘none of the above’ in multiple choice testing

Matt Pachai
McMaster University
Acknowledgements

- Dr. Joe Kim
- Dr. David DiBattista
- Yvonne Chen
- The Pedagogy Research Lab
Outline

1) The goal of multiple choice (MC)
2) None of the above (NOTA) in MC
3) The present experiment
4) Future directions and implications
Goals of Testing

• What are your goals in testing students?
  – Assessment?
  – Discrimination?
  – Learning?
MC Guidelines

• Haladyna and Downing (1989a) examined 46 textbook passages on MC

• Produced 43 recommendations for a “good” question
Sample Guidelines

- Use Positives, not Negatives, in the Stem
- Avoid None of the Above
- Avoid complex (Type K) questions
Which of the following would not increase obedience in the Milgram experiment?

i. Moving the experimenter to another room
ii. Moving the experiment to a run down building
iii. Dressing the experimenter in dirty clothes
iv. Moving the learner closer to the teacher

a) i and ii
b) ii and iii
c) i, ii, and iii
d) iii and iv
e) None of the above
Empirical Support

• Only half of these recommendations were empirically examined

• A clear need for rigorous examination remains

Haladyna and Downing, 1989b
Measurement Tools

- How do we examine our test’s ability to achieve our goals?
  - **Difficulty**: Percent Correct
  - **Discrimination**: Point-biserial correlation
  - **Learning**: Retention
Performance

• A simple way to measure knowledge at two levels

• **Students:**
  – How many questions did each student answer correctly?

• **Concepts:**
  – What percentage of students got a particular question correct?
Point-Biserial Correlation

- A measure of a question’s ability to discriminate between students

- What is the correlation between the answers for a particular question and each student's final score?
Point-Biserial Correlation

<table>
<thead>
<tr>
<th>Grade Category</th>
<th>Options</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>% A</td>
<td>A</td>
<td>0</td>
<td>0</td>
<td>90</td>
</tr>
<tr>
<td>% B</td>
<td>B</td>
<td>5</td>
<td>2</td>
<td>83</td>
</tr>
<tr>
<td>% C</td>
<td>C*</td>
<td>5</td>
<td>1</td>
<td>66</td>
</tr>
<tr>
<td>% D</td>
<td>D</td>
<td>23</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>% F</td>
<td>F</td>
<td>32</td>
<td>7</td>
<td>37</td>
</tr>
</tbody>
</table>

Point-biserial correlation = 0.32
Cognitive psychologists have extensively studied retention of material.

**Basic Paradigm:**
- Session 1: teach a concept
- Session 2: test retention after a delay
Numerous studies suggest testing improves learning.

Carpenter et al., 2008; Roediger and Karpicke (2006)
The Impact of Flaws

• Flawed questions are more difficult (Downing, 2005)
• Test flaws may hurt high achieving students more than low (Tarrant and Ware, 2008)
Specific Flaws

• Previous studies classify flawed questions based on a large number of guidelines
• Hard to decipher which specific flaws have which specific effects
The Case of NOTA

- In a recent review, 48% of textbook authors agreed that NOTA should be avoided (Haladyna et al., 2002)
Empirical Evidence

• The few studies examining NOTA have produced mixed results

• NOTA may:
  – increase difficulty and discrimination
  – not change difficulty and discrimination
  – increase difficulty but not discrimination
Mixed Messages

• “When NOTA is correct... it rewards examinees with serious knowledge deficiencies or misinformation” ... “Any stem or option format that reduces an item’s ability to distinguish between candidates with full and misinformation should not be used” (Gross, 1994)
• “NOTA should remain an option in the item-writer’s toolbox, as long as its use is appropriately considered. However, given the complexity of its effects, NOTA should generally be avoided by novice item writers.” (Haladyna et al., 2002)
General Questions

• What effect does NOTA have on:
  – Assessment?
  – Discrimination?
  – Learning? (not addressed today)
Our Study

• We examined NOTA on two of our Introductory Psychology examinations (approx 3000 students/year)

• Advantages of our population:
  – A large class
  – Highly motivated students
  – Topical questions, basic and applied
Test Design

• Five versions of each test were produced
• Each test contained 5 experimental questions, randomly distributed
Each test version had one question in each of the following conditions:

- No NOTA (control)
- NOTA as key
- NOTA replacing distractor #1
- NOTA replacing distractor #2
- NOTA replacing distractor #3
## Summary of Design

<table>
<thead>
<tr>
<th>FORM 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Normal</td>
</tr>
<tr>
<td>Q2</td>
<td>NOTA D3</td>
</tr>
<tr>
<td>Q3</td>
<td>NOTA D2</td>
</tr>
<tr>
<td>Q4</td>
<td>NOTA D1</td>
</tr>
<tr>
<td>Q5</td>
<td>NOTA key</td>
</tr>
</tbody>
</table>
Harlow's studies of infant monkeys raised with surrogate mothers indicated that infants became attached to the surrogate mother:

a) from which **food** was most often delivered.

b) **that provided the most contact comfort**.

c) that was present when **danger** was presented.

d) that was present for the greatest amount of **time**.
Harlow's studies of infant monkeys raised with surrogate mothers indicated that infants became attached to the surrogate mother:

a) from which food was most often delivered.
b) that was present when danger was presented.
c) that was present for the greatest amount of time.

d) None of the above
Harlow's studies of infant monkeys raised with surrogate mothers indicated that infants became attached to the surrogate mother:

a) that provided the most contact comfort.

b) that was present when danger was presented.

c) that was present for the greatest amount of time.

d) None of the above
Harlow's studies of infant monkeys raised with surrogate mothers indicated that infants became attached to the surrogate mother:

a) from which food was most often delivered.

b) that provided the most contact comfort.

c) that was present for the greatest amount of time.

d) None of the above
Harlow's studies of infant monkeys raised with surrogate mothers indicated that infants became attached to the surrogate mother:

a) from which **food** was most often delivered.

b) **that provided the most contact comfort.**

c) that was present when **danger** was presented.

d) None of the above
Recoding Distractors

• Distractors were recoded as either high frequency, middle frequency, or low frequency selections

• Harlow's studies of infant monkeys raised with surrogate mothers indicated that infants became attached to the surrogate mother:
  a) from which food was most often delivered. (HF: 19%)
  b) that provided the most contact comfort.
  c) that was present when danger was presented. (LF: 4%)
  d) that was present for the greatest amount of time. (MF: 17%)
Analysis

• Independent Variable: Condition
  – Normal
  – NOTA-Key
  – NOTA-HF
  – NOTA-MF
  – NOTA-LF

• Dependent Variables
  – Performance (% correct)
  – Discrimination (point-biserial correlation)
Performance

* = p < 0.001
Discrimination

Point Biserial Correlation

Normal | NOTA-Key | NOTA-HF | NOTA-MF | NOTA-LF

p > 0.05
Implications

• What effect does NOTA have on:
  – **Assessment:**
    • **Key:** Increased difficulty
    • **Distractor:** Less effective than a good distractor
  – **Discrimination:** No effect
  – **Learning:** Negative testing effect?  
    (Odegard and Koen, 2007)
Future Directions

• When NOTA is the correct answer, do the students selecting it know the truth?
  — Fill in the correct response for a bonus
• Understanding the specific effects of writing “errors” is highly important

• Test writers should be thoughtful in question writing
  – Questions should be matched to the goals of the test
Evaluating the use of ‘none of the above’ in multiple choice testing

Questions?
References


• Odegard, T. N., & Koen, J. D. (2007). "None of the above" as a correct and incorrect alternative on a multiple-choice test: Implications for the testing effect. *Memory, 15*(8), 873-885.
