caveon™

Testwiseness and Cheating: Two Major Scourges of Testing

David Foster
Caveon

Created October, 2018

Context

Psychometrics is often called the science of educational and psychological measurement.

Let's talk about Science for a few minutes.

#1 Responsibility of Scientists

To make sure that there are no confounding variables influencing the results of a scientific experiment, so clear conclusions can be drawn.

What are confounding variables?

What does confounding mean in this experimental context?

An Example of an Experiment

A researcher compared a new teaching method using students at a school. A teacher of 4th graders used the new method. A teacher of 3rd graders did not.

An Example of an Experiment

Test scores were used to evaluate the method, which showed that the 4th graders had higher test scores, a difference that was statistically significant.

The researcher concluded that the new method was effective.

Questions

What is the one main confounding variable for this study?

What could the researcher have done to control this confounding variable?

The Same Concept Applies to Testing Professionals

Given a raw test score of 49 out of 60 for an examinee on a test, what might be confounding variables that have nothing to do with the skill being measured that may have affected the score?

#1 Responsibility of All Testing Professionals

To protect the clarity, and therefore the value, of every test score.

Which means, to control sources that confound the interpretation of scores.

What are some confounding variables for test scores?

Prepare yourself! The next slide may shock you.

Some of the Confounding Variables Affecting a Test Score

1-8 Situation and Test Variables

1. Speededness
2. Distractions
3. Room Temperature/Humidity
4. Screen Resolution
5. Font Size
6. Computer Processing Speed
7. Internet Interruptions
8. Paper test or Computerized Test differences

9-22 Test Taker Variables

9. Cheating
10. Testwiseness
11. Anxiety
12. Language
13. Reading Ability
14. Distractability
15. Illness
16. Fatigue
17. Culture
18. Gender
19. Race/Ethnicity
20. Age
21. Physical, Emotional, Social or Mental Disability
22. Boredom and Disinterest

Discussion: What can be done about them?

Exercise:

Pick 1 at random

Brainstorm a bit here

Some Raise Scores and Some Lower Scores

1-8 Situation and Test Variables

1. Speededness
2. Distractions
3. Room Temperature/Humidity
4. Screen Resolution
5. Font Size
6. Computer Processing Speed
7. Internet Interruptions
8. Paper test or Computerized Test

9-22 Test Taker Variables

9. Cheating
10. Testwiseness
11. Anxiety
12. Language
13. Reading Ability
14. Distractability
15. Illness
16. Fatigue
17. Culture
18. Gender
19. Race/Ethnicity
20. Age
21. Physical, Emotional, Social or Mental Disability
22. Boredom and Disinterest

Focusing on the two that Raise Scores

One is the clearly within the subject matter of this conference: Security

The other, Testwiseness, is closely related: Get a higher score, unrelated to the skill being measured.

Cheating and Testwiseness

Two forms of cheating, one sanctioned by the testing community and one not.

Both of these fit the definition of cheating, which is “getting a higher score using an irrelevant skill.”

Prevalence

Prevalence of Cheating

Most of the time prevalence isn't known.

Caveon's Chief Scientist has seen rates of prevalence from 2% to 60%, using conservative data forensics measures. Cheaters get higher scores, non-cheaters get lower scores.

Prevalence of Testwiseness: A Form of Cheating

Are your tests made up of multiple-choice questions?

Then 100% your test scores are confounded by testwiseness, helping some test takers, and comparatively, hurting others. And you don't know which is which.

There is a 5% to 10% confounding effect on test scores.

Testwiseness: One Example

What is the second task of the Triwizard Tournament?

- Besting the Dragon
- Facing the Merpeople
- Solving the Riddle of the Golden Egg
- Navigating the Maze
- Capturing the Goblet of Fire

Testwiseness: One Example

What is the second task of the Triwizard Tournament?

- Besting the Dragon
- Facing the Merpeople
- Solving the Riddle of the Golden Egg
- Navigating the Maze
- Capturing the Goblet of Fire

4.7% Testwiseness Effect!

Testwiseness: A Second Example

How many times does Harry Potter watch the Sorting Ceremony?

- 4
- 5
- 3
- 7
- 6

Testwiseness: A Second Example

How many times does Harry Potter watch the Sorting Ceremony?

- 4
- 5
- 3
- 7
- 6

10.1% Testwiseness Effect!

Testwiseness: A Third Example

The value of money depends on

- what the government says each dollar is worth.
- the cost of producing it.
- how much each dollar will buy.
- the amount of gold the government backs each dollar with.

Testwiseness: A Third, More Blatant Example

The value of money depends on

- what the government says each dollar is worth.
- the cost of producing it.
- how much each dollar will buy.
- the amount of gold the government backs each dollar with.

???% Testwiseness Effect!

Test and Item Design

To control for confounding variables, some associated with testwiseness and cheating

Design Steps We've already taken

One type of cheating is **copying from a neighbor**. We have put in place ways to prevent it:

1. Seating spread apart
2. Screen privacy adapters
3. Carrels with barriers
4. Randomized items (Test Design)
5. Randomized options (Item Design)

This type of cheating is no longer the problem it once was.

Design Steps We've Already Taken

Another type of cheating is **sharing a key for MC questions**. We have put in place ways to prevent it:

1. Randomizing MC options so the key is never the same.
2. Creating and randomizing forms, so a memorized key for one form is useless.
3. Using different item types so that a simple answer key is less relevant.

ASIDE: Importance of Randomization

Randomization results in unpredictability.

Unpredictability is anathema to cheaters, who depend on consistency.

Removing Confounding Factors with New Item Design

The major fraud support for pre-knowledge cheating is **stealing test content**. We have put in place ways to prevent it:

1. **SmartItems** randomize test content for test takers across the breadth of an entire skill, making specific content completely unpredictable.

SmartItems

- ✓ SmartItems cover the breadth of their associated skill completely.
- ✓ SmartItems change for every test taker within the boundaries of the skill, construct, objective or competency being measured.
- ✓ A SmartItem can vary in thousands or millions of ways.
- ✓ Scientific research indicates that SmartItems are psychometrically sound.

SmartItem Demonstration

https://scorpion.caveon.com/projects/618/items/editor?item_id=40462

This SmartItem covers the entire breadth of a chapter on Research Methods in an introductory psychology textbook.

Another New Design Step

Another support for pre-knowledge cheating is **using testwiseness**. We have put in place ways to prevent it:

1. Testwiseness requires the direct comparison of options. DOMC prevents that from happening.

Several scientific experiments have establish this benefit of DOMC.

Discrete Option Multiple Choice or DOMC

DOMC present options one at a time (rather than simultaneously) until the item is answered correctly or incorrectly.

https://scorpion.caveon.com/projects/618/items/editor?item_id=35300

This is the same item content as before.

Item Designs to Consider

SmartItems do away with item theft and most forms of cheating.

DOMC removes most testwiseness effects.

Plus, they can be combined!

The old saying is “Kill two birds with one stone.”

caveon™

Thanks!