caveon™

History and Future of Selected-Response Items: Workshop

David Foster

Created October, 2018

Workshop Order of Events

- History of MC in Testing
- Institutionalization of MC
- Liabilities of MC (have always existed)
- Changes have occurred and more are needed
- Need to break the institutionalization
- Everyone is needed in this effort
- Two more changes: SmartItems and DOMC
- Producing a SmartItem Exercise

Definition of MC and SR

Multiple-Choice (MC) is defined in this session as how it is used today almost ubiquitously:

✓Single correct answer

✓4-5 options

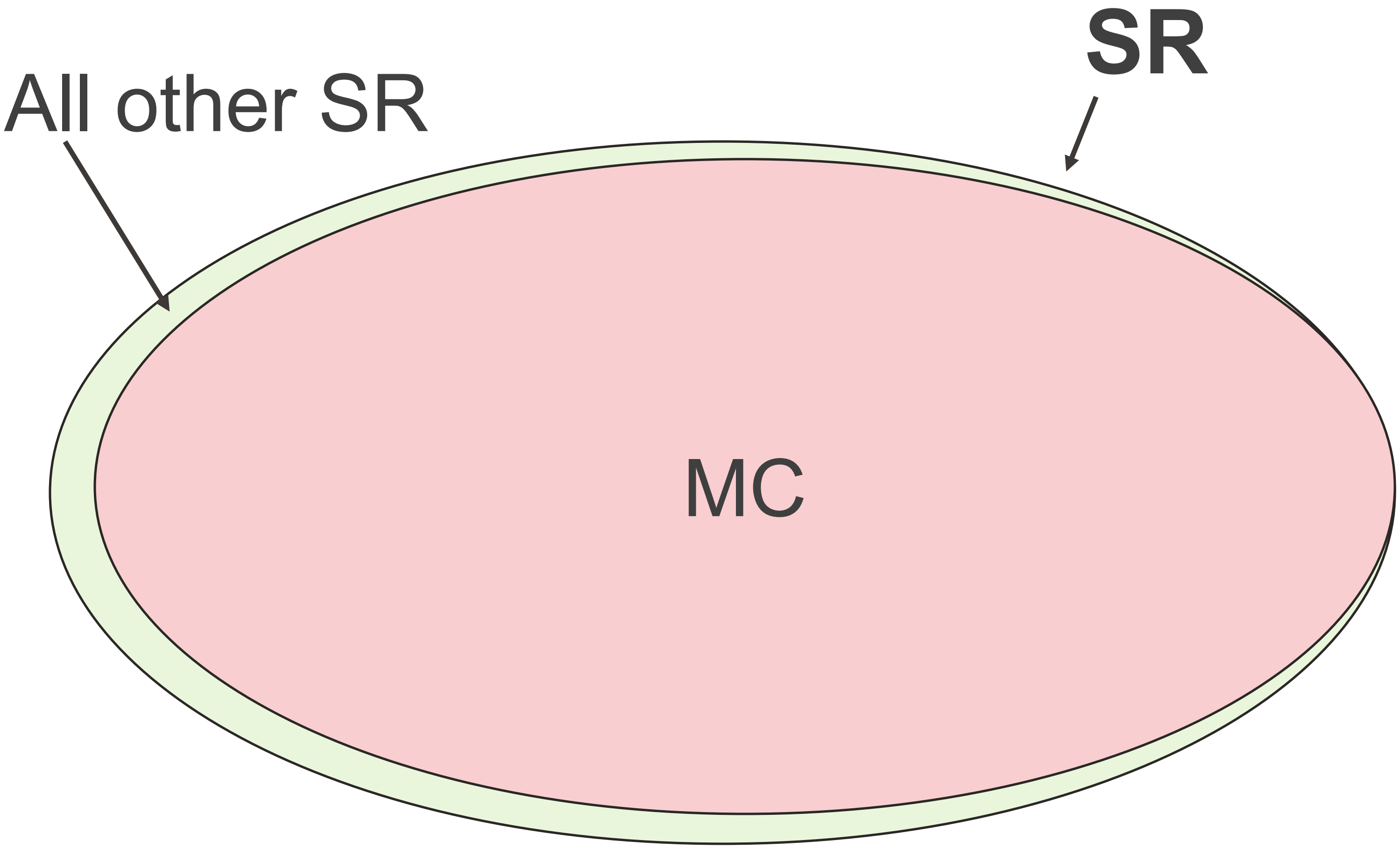
✓Unchanging content, except for random ordering of options for some programs

Selected-Response (SR)

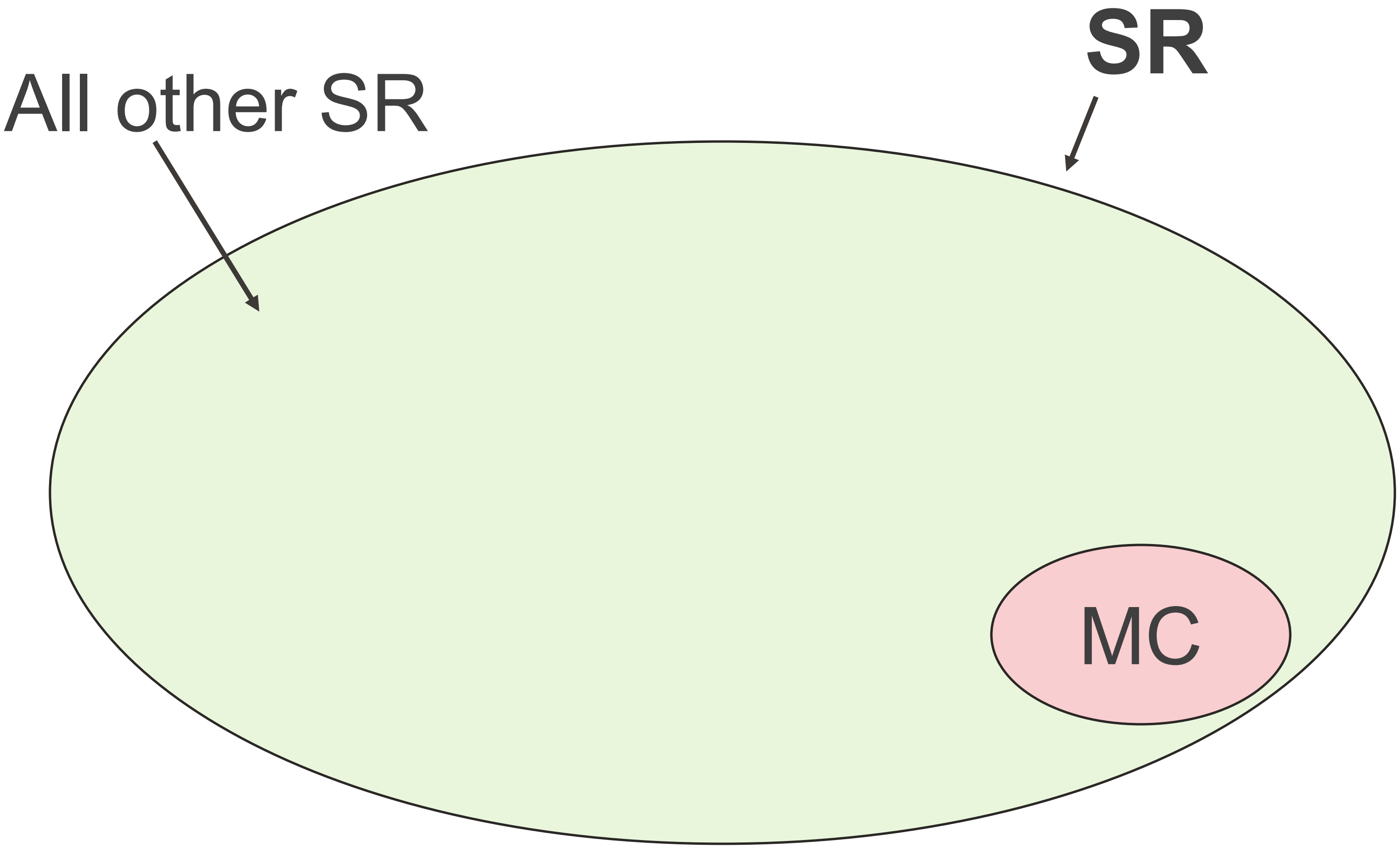
✓The test development process items where all of the content is built-in. Test takes simply need to select answers that they see.

✓Examples include MC and many other variants.

Venn Diagram: MC and SR, Ubiquity of MC versus all other SR



Venn Diagram: MC and SR, Potential Value of MC versus all other SR



Goals of the Workshop

- ✓ Impart the lessons I've learned about SR items over 36 years
- ✓ Put MC in the context of SR alternatives
- ✓ Introduce you to a future beyond MC, likely without MC
- ✓ Discuss the benefits of this future
- ✓ Discuss the best ways to bring it about

Lessons I’ve Learned about MC

- MC isn’t sacred and is has become an unnecessary liability during the past 30 years.
- SR remains a solid design choice for great measurement, better security, convenience, and easy scoring.
- Good SR designs solve problems; Bad SR designs cause problems.
- New computer technologies, getting better all the time, enable great item designs.
- Commitment to existing technology systems and development processes, ones that are reluctant to change, keeps MC, unfortunately, in a dominant role.
- Test takers quickly adapt to new SR items.
- Psychometrics for any SR are usually the same.

Overall purpose of today’s session:

To introduce you to a likely “Future of MC”

NEED: The driving force behind item change! (Or any change, for that matter)

Items change in response to needs:

1. Better measurement
2. More fairness
3. More security
4. Less expensive
5. More convenient
6. Quicker
7. Easier

These are all legitimate reasons for change!

HISTORY OF MC/SR

BEFORE MC

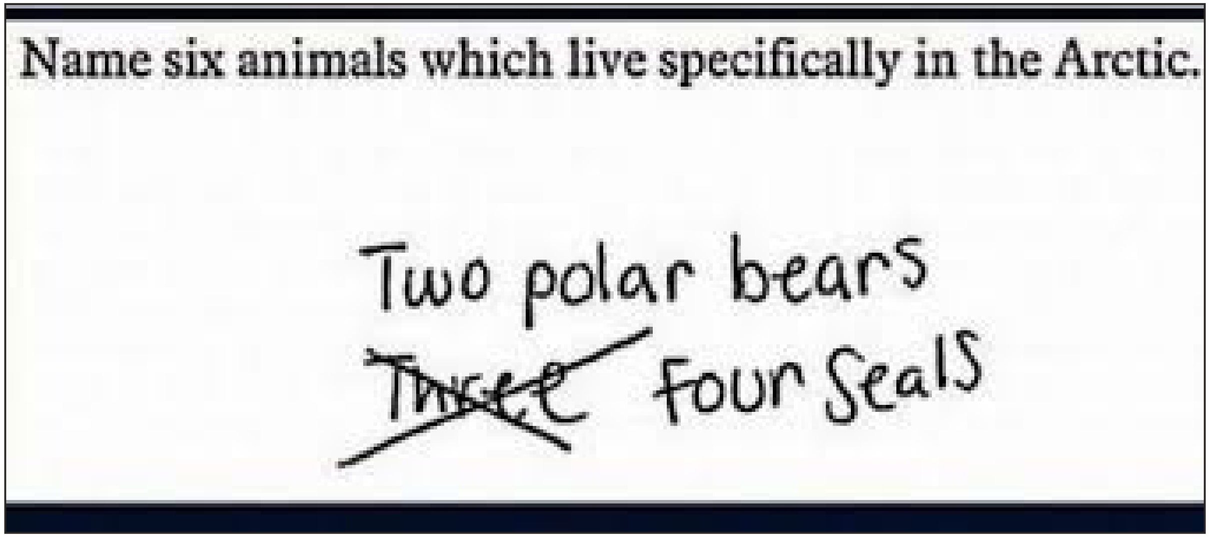
Mr 18256
SALV

Before MC was CR

Before MC individuals were asked questions orally. And they provided their answers orally. This type of assessment happens rarely today.

Also Before MC, more CR

Before MC, paper tests would require written answers.



Name six animals which live specifically in the Arctic.

Two polar bears

~~Three~~ four Seals

Before MC: Constructed Response was not working

- Too many errors
- Too slow and inefficient
- Scoring bias
- Not able to scale

A new solution was desperately needed!

Enter: A GREAT INVENTION!

Frederick James Kelly (1880-1959)

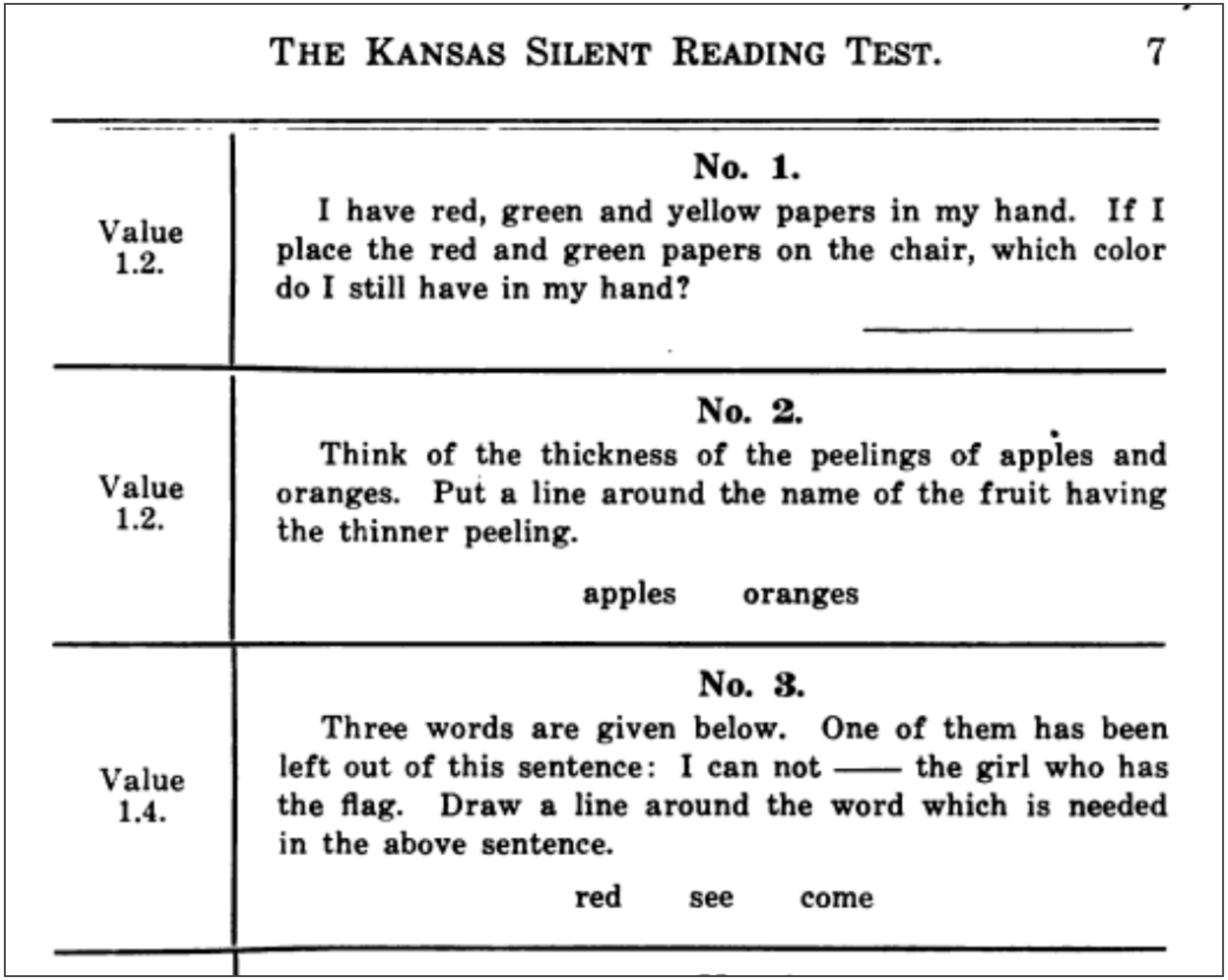
1914

Dissertation on teachers' scoring errors

1915

Kansas Silent Reading Test

1st test with MC items



THE KANSAS SILENT READING TEST. 7

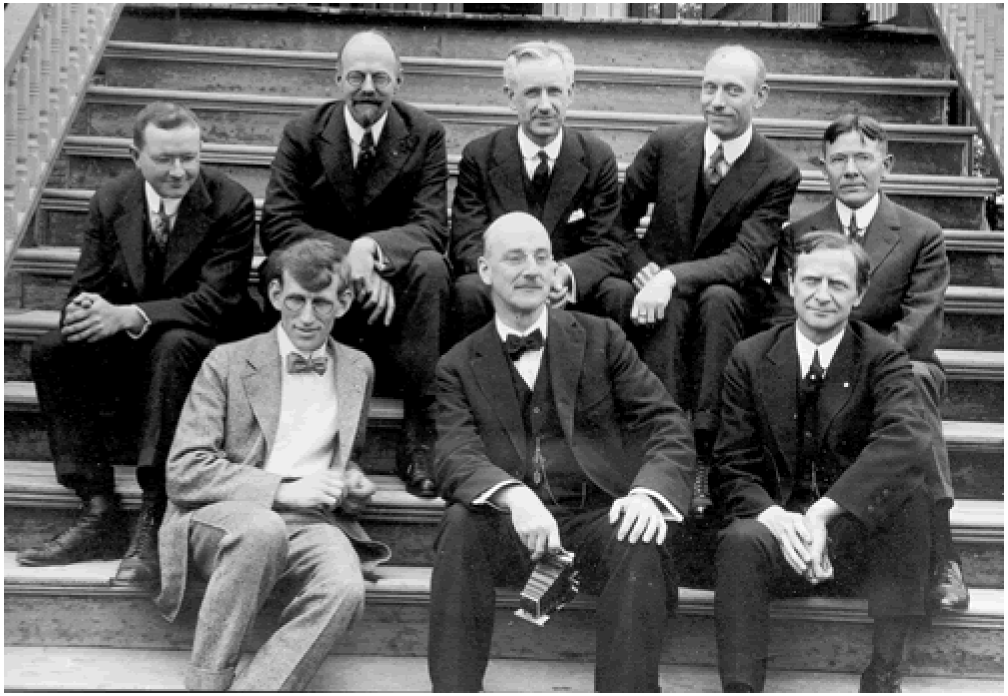
Value 1.2.	**No. 1.** I have red, green and yellow papers in my hand. If I place the red and green papers on the chair, which color do I still have in my hand? ____________
Value 1.2.	**No. 2.** Think of the thickness of the peelings of apples and oranges. Put a line around the name of the fruit having the thinner peeling. apples oranges
Value 1.4.	**No. 3.** Three words are given below. One of them has been left out of this sentence: I can not —— the girl who has the flag. Draw a line around the word which is needed in the above sentence. red see come

F. J. Kelly Invented MC…

…to standardize responding and scoring in order to reduce scoring errors and bias, and

…to reduce time and effort in test administration and scoring.

Close behind Kelly: The Vineland Group



The Vineland Group charged with assessing army recruits in 1915

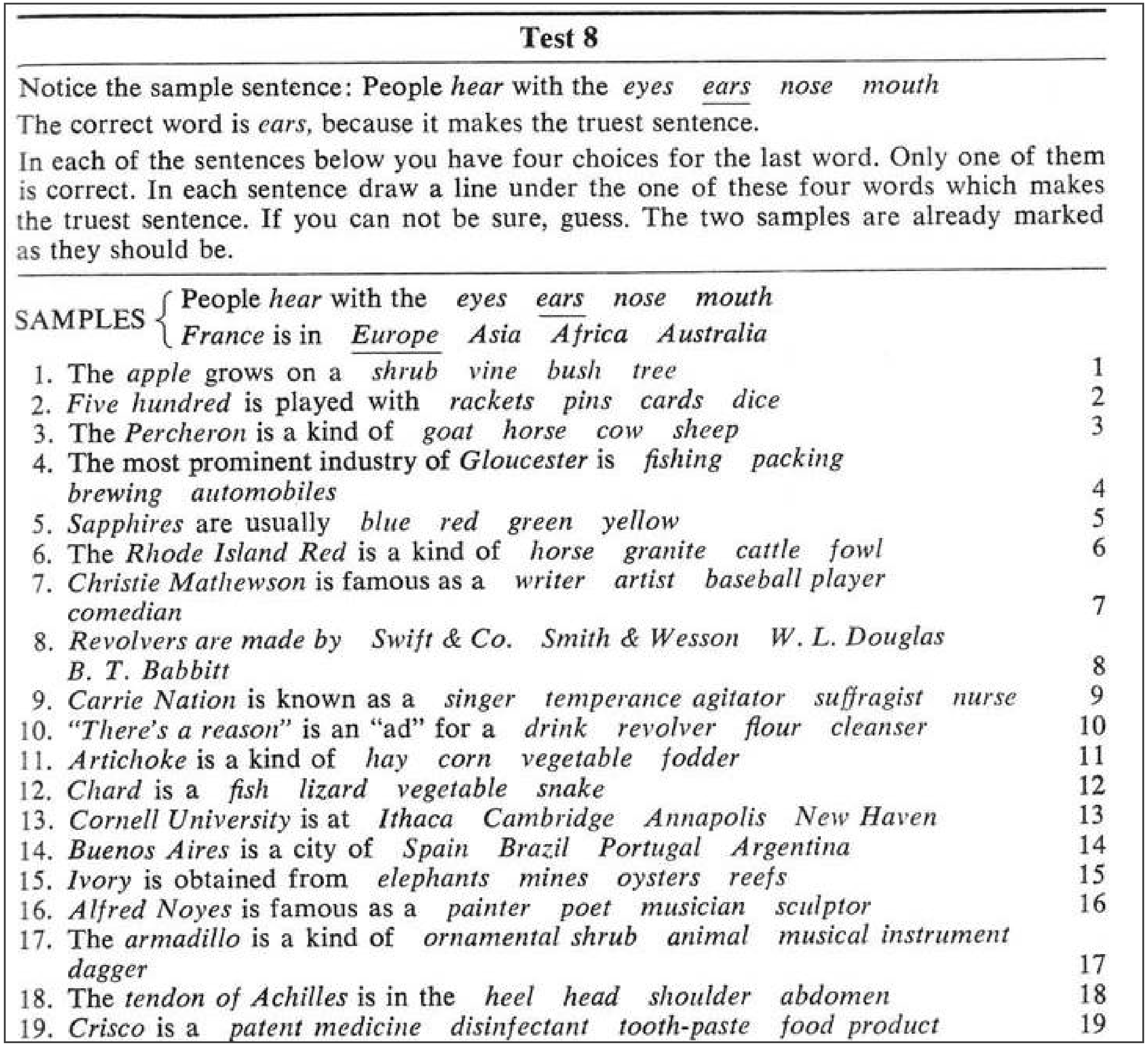
Robert Yerkes (1876-1956)

Louis Terman (1877-1956)

Arthur Otis (1886-1963)

…and others

2 Years later, 1917, the Army Alpha



Test 8

Notice the sample sentence: People *hear* with the *eyes* *ears* *nose* *mouth*

The correct word is *ears,* because it makes the truest sentence.

In each of the sentences below you have four choices for the last word. Only one of them is correct. In each sentence draw a line under the one of these four words which makes the truest sentence. If you can not be sure, guess. The two samples are already marked as they should be.

SAMPLES { People *hear* with the *eyes* *ears* *nose* *mouth*
France is in *Europe* *Asia* *Africa* *Australia*

1. The *apple* grows on a *shrub* *vine* *bush* *tree* 1
2. *Five hundred* is played with *rackets* *pins* *cards* *dice* 2
3. The *Percheron* is a kind of *goat* *horse* *cow* *sheep* 3
4. The most prominent industry of *Gloucester* is *fishing* *packing* *brewing* *automobiles* 4
5. *Sapphires* are usually *blue* *red* *green* *yellow* 5
6. The *Rhode Island Red* is a kind of *horse* *granite* *cattle* *fowl* 6
7. *Christie Mathewson* is famous as a *writer* *artist* *baseball player* *comedian* 7
8. *Revolvers are made by* *Swift & Co.* *Smith & Wesson* *W. L. Douglas* *B. T. Babbitt* 8
9. *Carrie Nation* is known as a *singer* *temperance agitator* *suffragist* *nurse* 9
10. *"There's a reason"* is an "ad" for a *drink* *revolver* *flour* *cleanser* 10
11. *Artichoke* is a kind of *hay* *corn* *vegetable* *fodder* 11
12. *Chard* is a *fish* *lizard* *vegetable* *snake* 12
13. *Cornell University* is at *Ithaca* *Cambridge* *Annapolis* *New Haven* 13
14. *Buenos Aires* is a city of *Spain* *Brazil* *Portugal* *Argentina* 14
15. *Ivory* is obtained from *elephants* *mines* *oysters* *reefs* 15
16. *Alfred Noyes* is famous as a *painter* *poet* *musician* *sculptor* 16
17. The *armadillo* is a kind of *ornamental shrub* *animal* *musical instrument* *dagger* 17
18. The *tendon of Achilles* is in the *heel* *head* *shoulder* *abdomen* 18
19. *Crisco* is a *patent medicine* *disinfectant* *tooth-paste* *food product* 19

The Army Needed…

…to administer and score correctly millions of tests ASAP!

There was a worldwide war going on.

Point: Kelly Invented MC, but it was actually a initial variety of SR

In other words, what Kelly really invented was an item format where the choices were created in advance, laid out in front of the test takers, who’s job it was to choose one of them.

Unintended Consequences

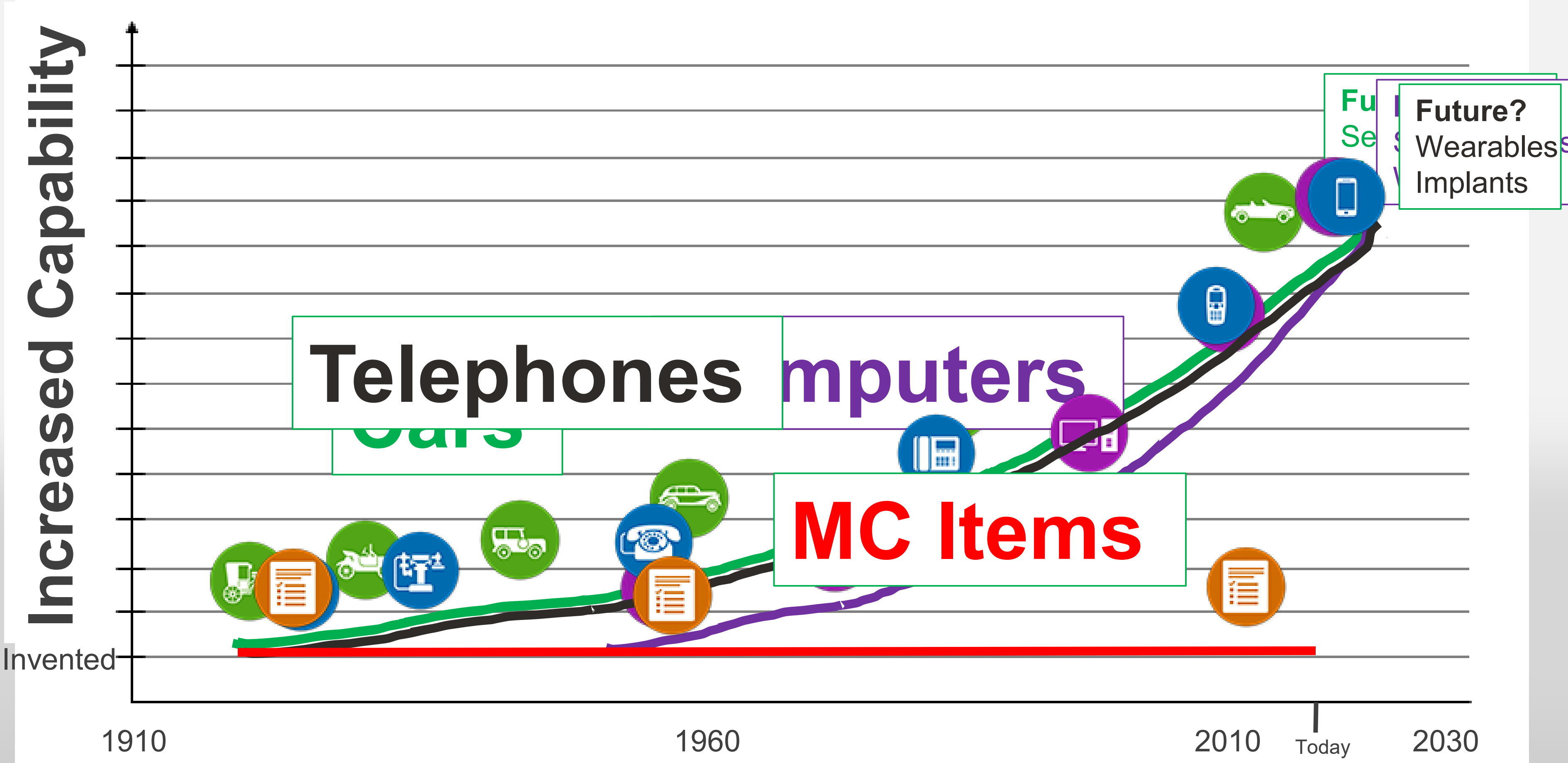
The Good News! These two early uses of MC questions established MC forever as a valuable testing tool.

The Bad News! These uses institutionalized MC, discouraging innovation for decades.

Let's look at other inventions in the early 1900's.

Let’s contract MC with other inventions of the early 20th Century.

Advance of Inventions Through Innovation



Innovation means that earlier versions are replaced by a better way.

History teaches us…

…MC, the foundation of all of our tests, has not experienced innovative change in over 103 years!

The Institutionalization of MC
How did it happen?

Systems support it.

Processes support it.

“Best practices” support it.

Fear of a legal challenge when something new is tried.

Topic of change is often ignored.

“We have a handle on this”. Well we should, after 100 years.

still there is much unknown about MC

and really, we don’t have a handle on it

New technologies like AIG support it.

Scanning systems and bubble sheets.

Scoring methods.

IRT and other analysis methods are built for it.

Critics are discredited or ignored

Quote about MC by Historian Samelson in 1987

"Would F. J. Kelly, were he still alive, be happy to see the permanent institutionalization of his invention? Or would he be horrified to find that 70 years of sophisticated analysis techniques, computerization, and research have not produced any new breakthroughs or even significant improvements of this rather primitive, if ingenious, pre-World War I technique, which is still the basic vehicle for many important decisions about individuals?" (p. 124)

NEEDS Then and Now

Then: large scale testing was impossible, scoring was unreliable, scoring was biased

⬇

MC

Now: New Problems

- **Theft of tests and cheating**
- **Limitations on measurement**
- **Unfairness**
- **Expensive to create and administer tests**
- **Public dissatisfaction**

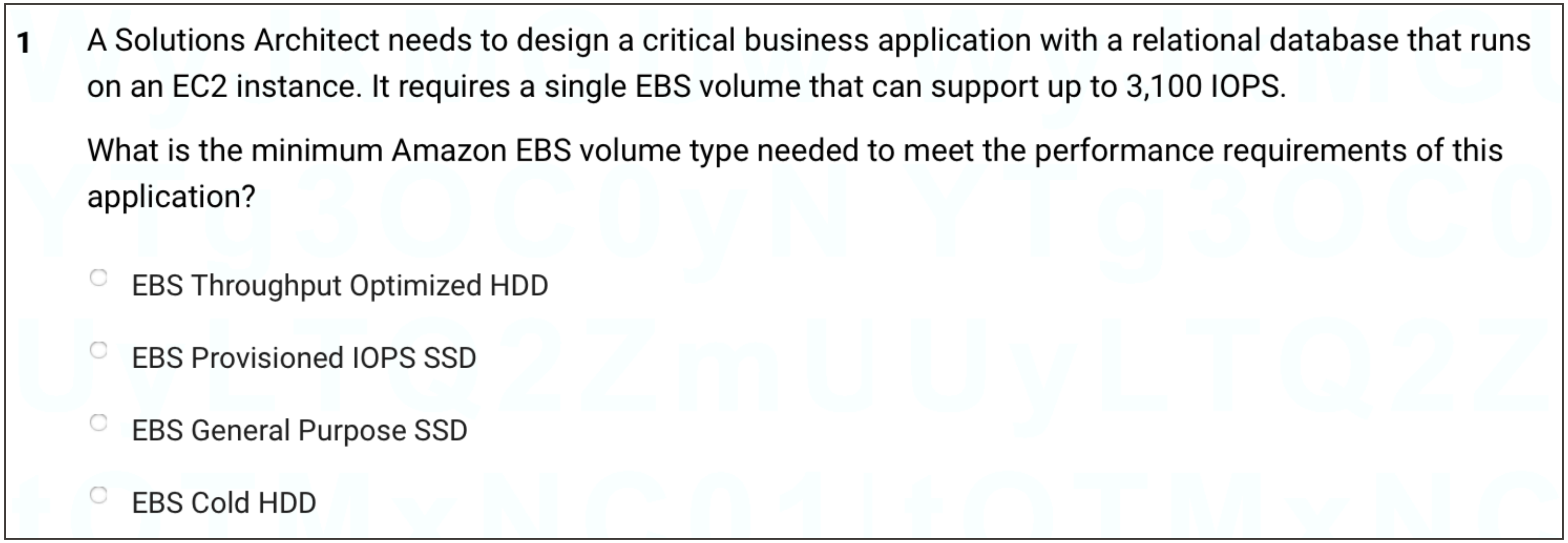
Discussion

Criticisms of MC Today

Selected Response Variants

#1 SR Variant: MC with Randomized Options (introduced 1980's)

Example:



1 A Solutions Architect needs to design a critical business application with a relational database that runs on an EC2 instance. It requires a single EBS volume that can support up to 3,100 IOPS.

What is the minimum Amazon EBS volume type needed to meet the performance requirements of this application?

○ EBS Throughput Optimized HDD

○ EBS Provisioned IOPS SSD

○ EBS General Purpose SSD

○ EBS Cold HDD

Comments: Makes some forms of copying/cheating very difficult. Like all MC, susceptible to testwiseness.

#2 SR Variant: Negatively Worded MC (introduced early 20th century)

Example:

Which is **NOT** a breed of herding dogs?

A. Belgian Tervuren

B. Border Collie

C. German Shepherd

D. Great Pyrenees

Comments: Used because systems could only accommodate a single correct option. Causes confusion and errors, especially in some populations.

Advice: AVOID! Good replacement is available.

#3 SR Variant: Multiple-Correct Selected Response, Prompted

(Introduced early 1990’s)

Example:

Which are 3 breeds of herding dogs? (Choose three.)

A. Belgian Turvuren
B. Border Collie
C. German Shepherd
D. Great Pyrenees

Comment: Removes the errors associated with Negatively worded MC and common alternative, SATA.

Advice: Use instead of Select All That Apply (SATA).

#4 SR Variant: True/False or Two-Choice MC (introduced early 20th century)

Example:

The Eiffel Tower is in Marseilles.

A. True

B. False

Comments: Easy to guess. Difficult to create. Easy to steal, like all MC.

Advice: Avoid because there are better SR alternatives.

#5 SR Variant: K-Type MC

(introduced middle 20th century)

Example:

1. A student suffers an injured ankle while running to first base in a softball game. The teacher examines the indicated area. The symptoms are typical of a sprained ankle, although the injury may in fact be more severe. Which of the following steps should be included in the first aid administered to the student?

   I. Elevate the injured leg
   II. Apply ice to the injured area
   III. Apply direct pressure to the site of the injury

   a. I only
   b. II only
   c. I and II only
   d. I and III only

Comment: Complicated and error prone. Unnecessary today.

Advice: Easily avoided!

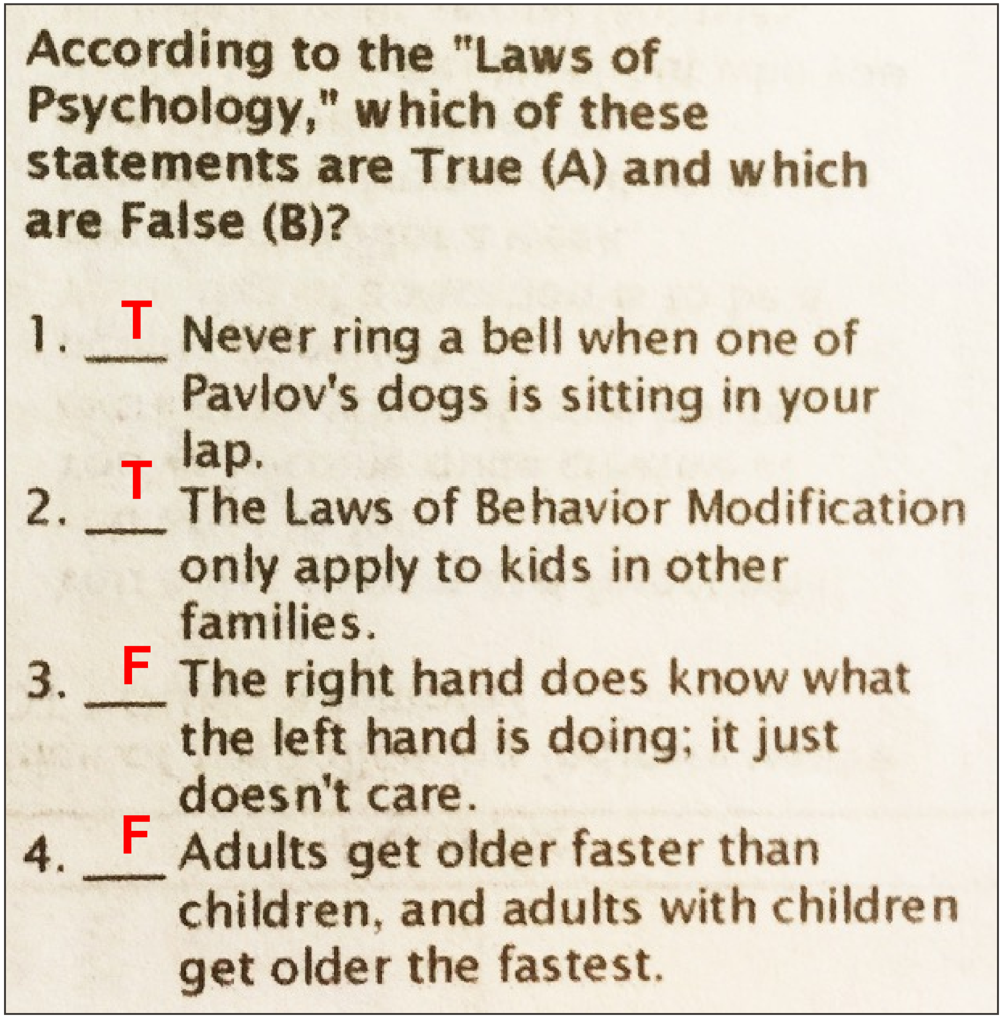
#6 SR Variant: Multiple True/False

(introduced middle 20th century)

Example:

Comments: Nothing more than many T/F questions strung together. Simply several MCs.

Advice: Easy to avoid.



According to the "Laws of Psychology," which of these statements are True (A) and which are False (B)?

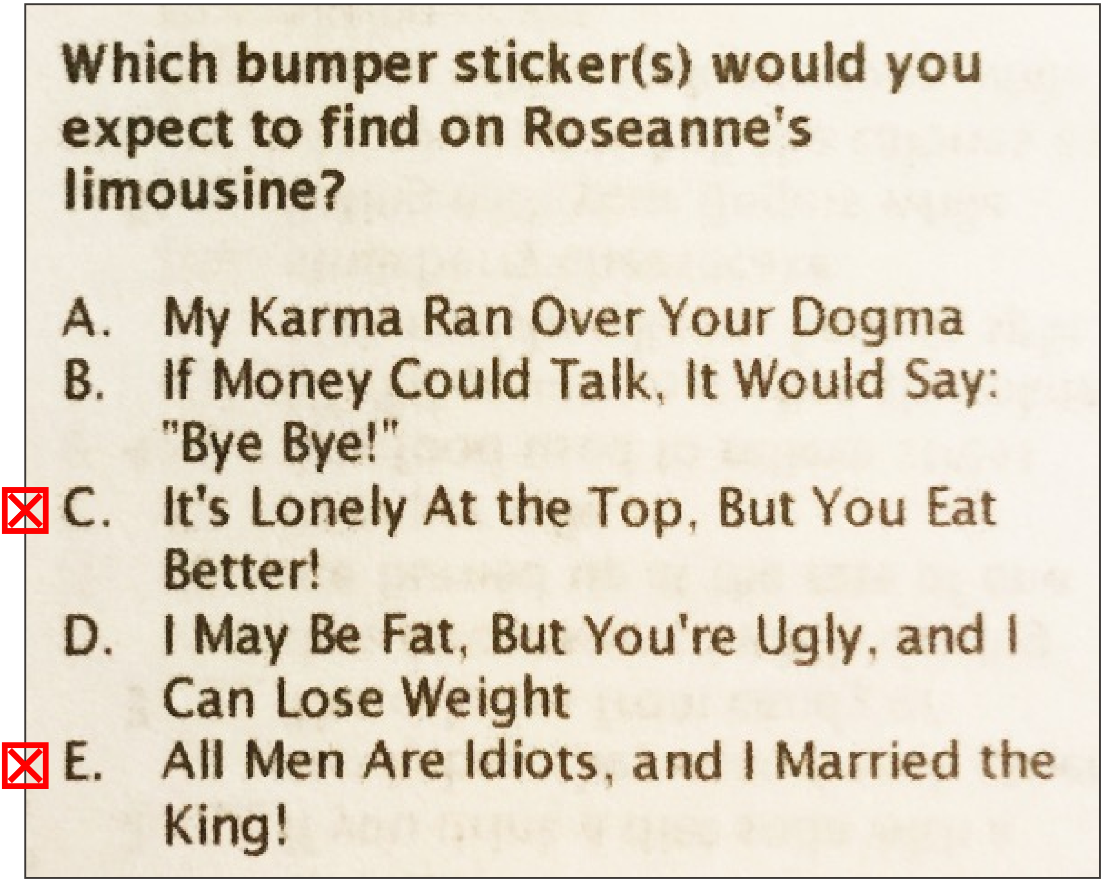
1. T Never ring a bell when one of Pavlov's dogs is sitting in your lap.
2. T The Laws of Behavior Modification only apply to kids in other families.
3. F The right hand does know what the left hand is doing; it just doesn't care.
4. F Adults get older faster than children, and adults with children get older the fastest.

#7 SR Variant: Select All That Apply or SATA (introduced 1990s)

Example:

Comments: Unnecessarily difficult. Actually Causes errors, especially in high-performing test takers!

Advice: Definitely AVOID! Replace with prompted variety (#3 above).



Which bumper sticker(s) would you expect to find on Roseanne's limousine?

A. My Karma Ran Over Your Dogma

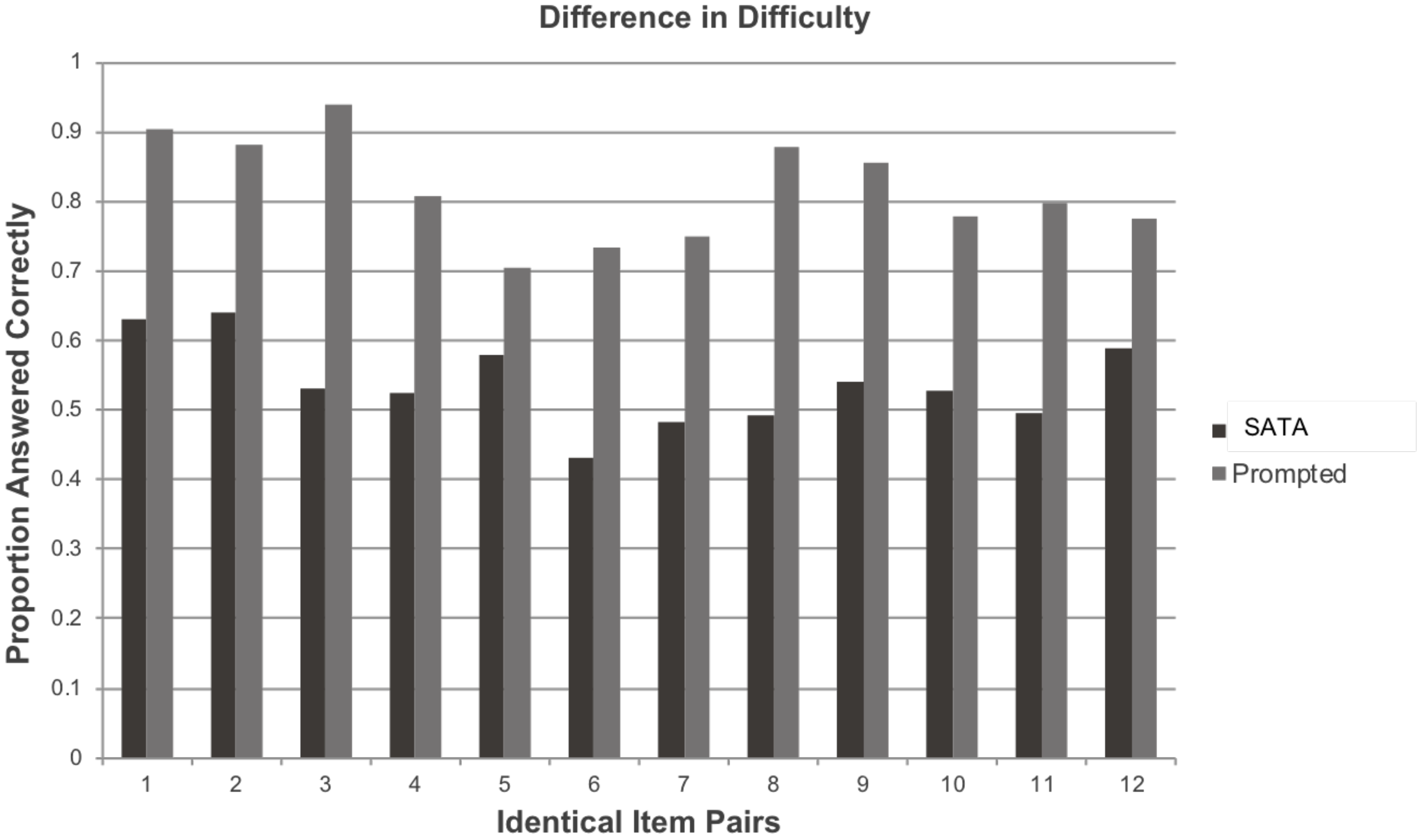
B. If Money Could Talk, It Would Say: "Bye Bye!"

☒ C. It's Lonely At the Top, But You Eat Better!

D. I May Be Fat, But You're Ugly, and I Can Lose Weight

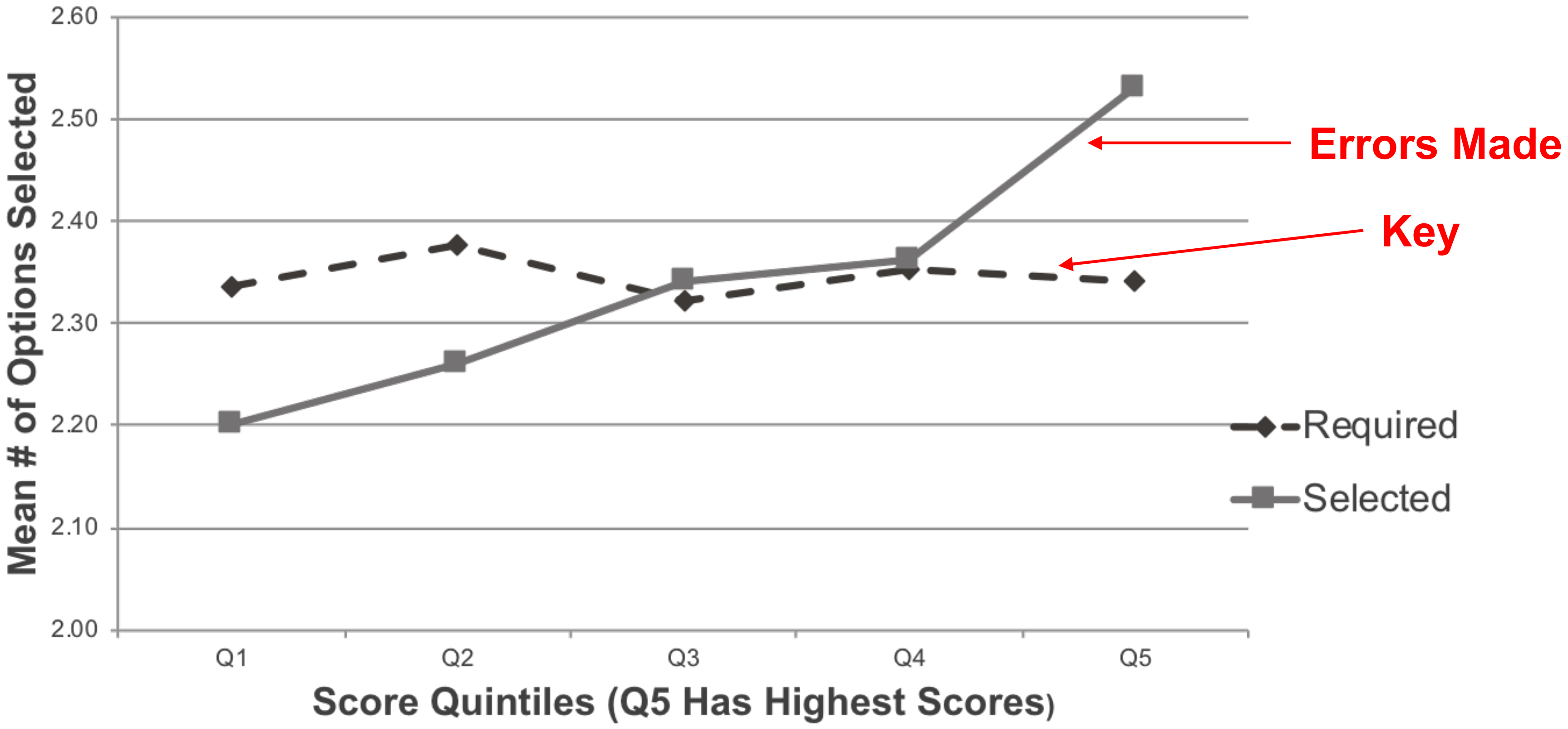
☒ E. All Men Are Idiots, and I Married the King!

Major Finding: Difficulty Difference between SATA and Prompted



SATA # of Responses Made to Items

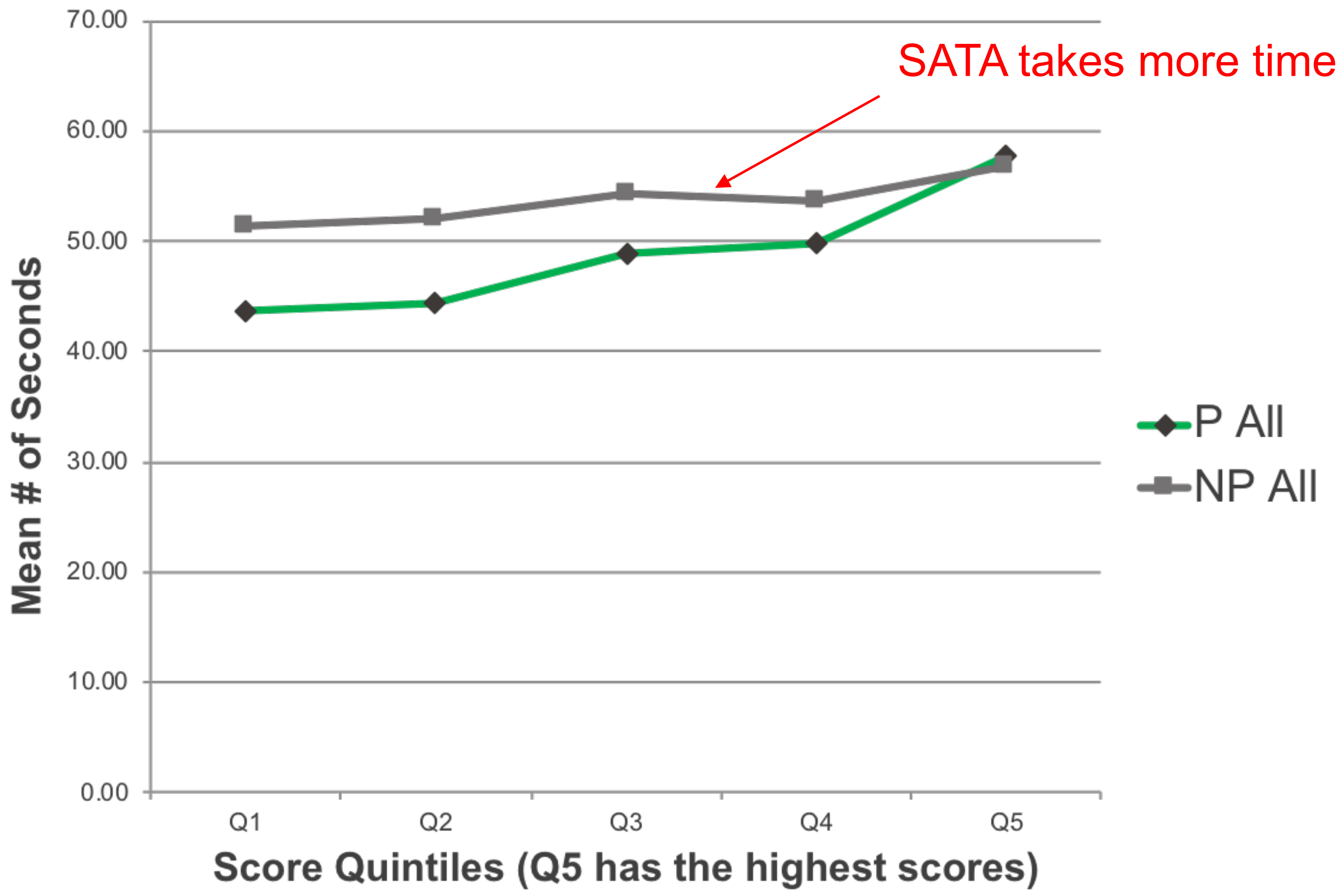
Mean Number of Selections for Incorrect SATA Items by Quintile



Errors Made

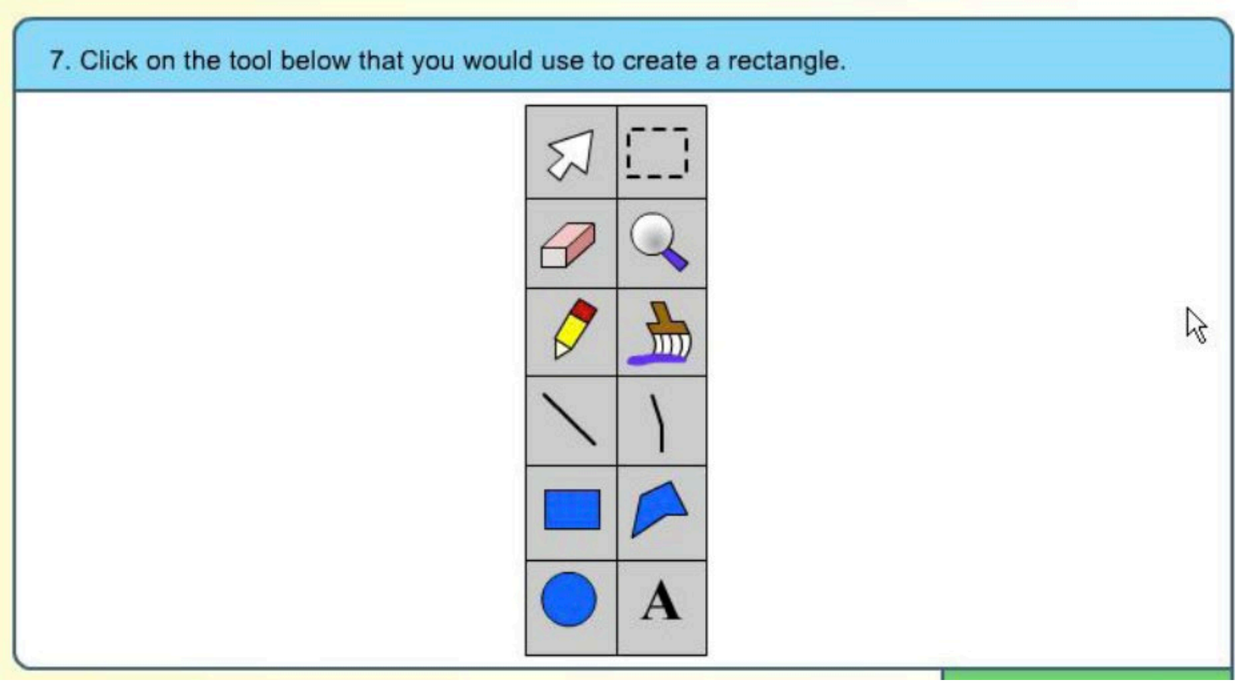
Key

Timing: Prompted vs. Non-Prompted



#8 SR Alternative: Hot-Spot MC (introduced 1990's)

Example:



7. Click on the tool below that you would use to create a rectangle.

From Proftesting.com

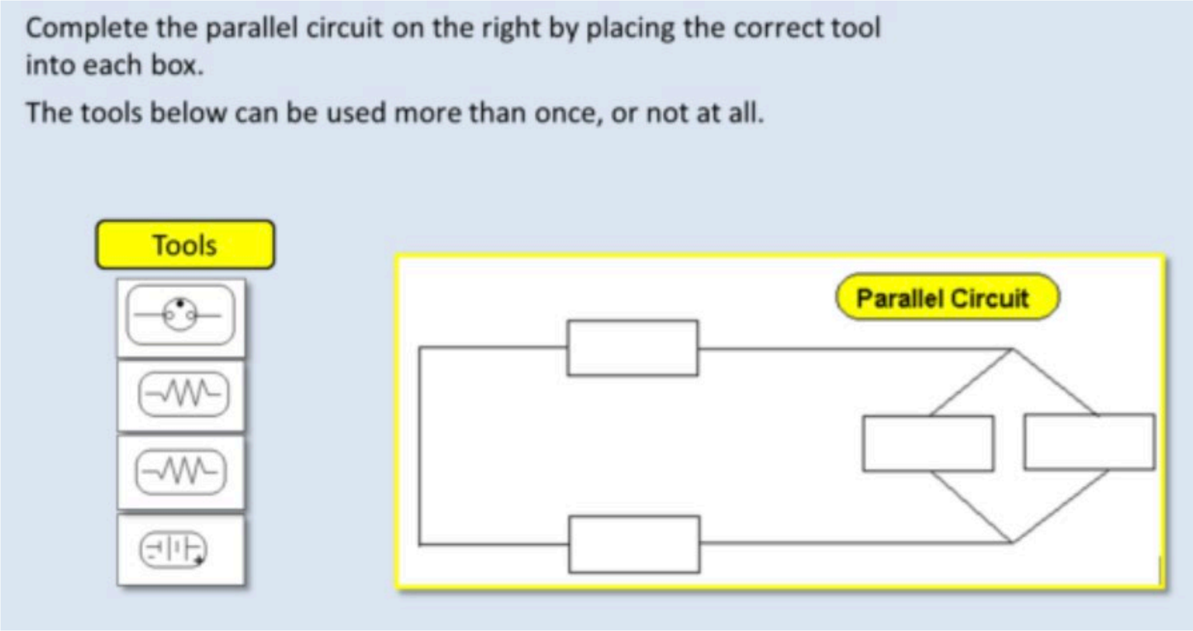
Comments: MC with a different response process. Useful for the evaluation of some skills.

Advice: Use when needed.

#9 SR Alternative: Drag-and-Drop

(introduced 1990's)

Example:



Complete the parallel circuit on the right by placing the correct tool into each box.

The tools below can be used more than once, or not at all.

From Proftesting.com

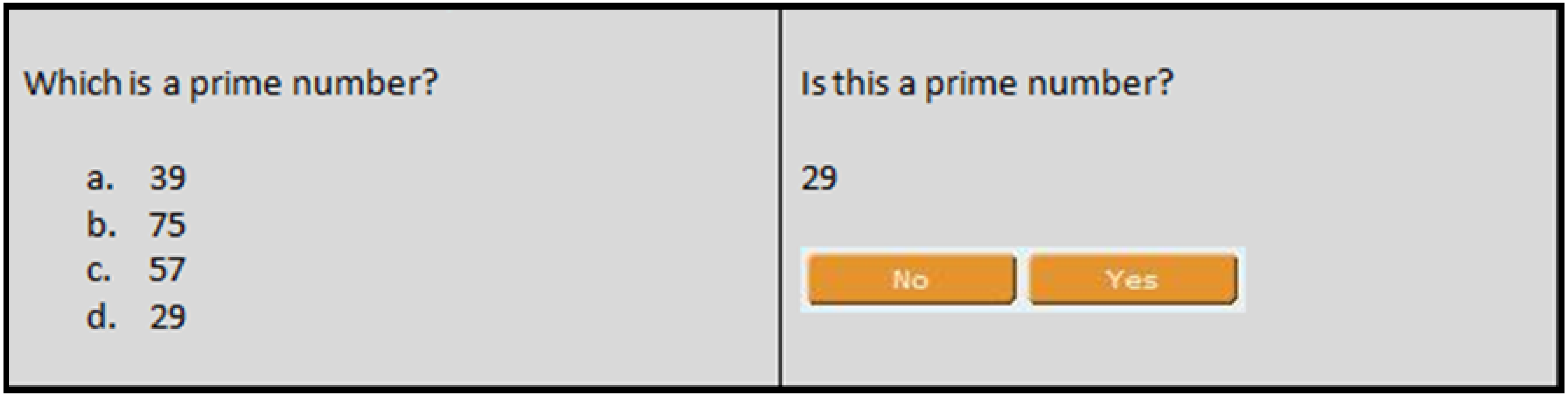
Comments: Examinees drag objects to destinations using available tools. Useful for the evaluation of some skills.

Advice: Use when needed.

#10 SR Alternative: Discrete Option Multiple Choice (Introduced 2009)

Example:

Multiple Choice | **DOMC**



Which is a prime number?

a. 39
b. 75
c. 57
d. 29

Is this a prime number?

29

Comments: Scientific research indicates DOMC is more secure, removes testwiseness, causes fewer errors, takes less time to answer, and more.

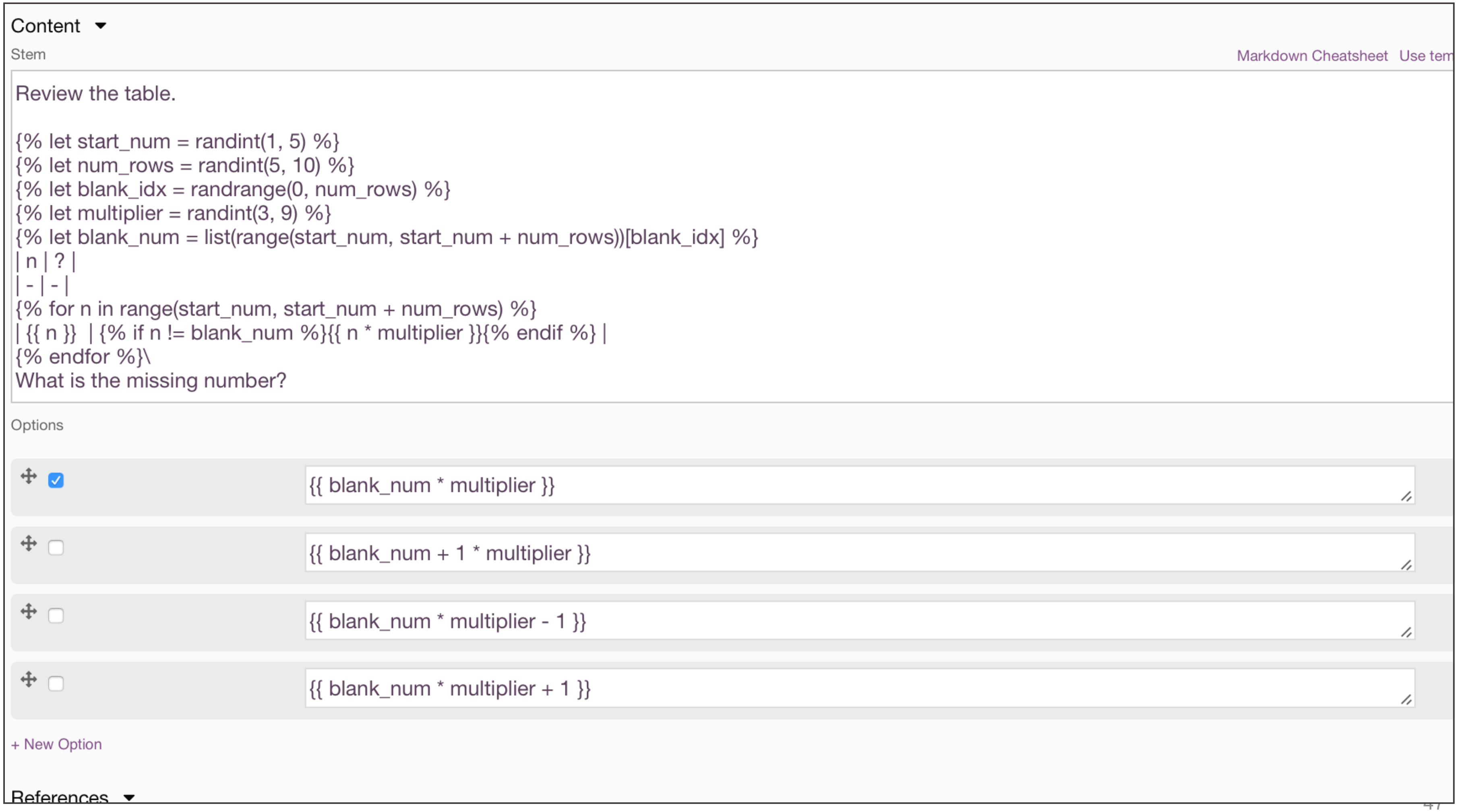
Advice: Switch to DOMC as soon as possible.

#11 SR Variant: SmartItems (Introduced 2018)

1. SmartItems cover the breadth of its associated skill completely.
2. SmartItems change for every test taker within the boundaries of the skill, construct, objective or competency being measured.
3. A SmartItem can vary in thousands or millions of ways.
4. Works with any SR variant format.
5. Scientific research indicates that SmartItems are psychometrically sound.
6. Research indicates that its benefits, as incredible as they are, are real.

#11 SR Variant: SmartItems

Behind the scenes of an Example:



Content ▾

Stem

```
Review the table.

{% let start_num = randint(1, 5) %}
{% let num_rows = randint(5, 10) %}
{% let blank_idx = randrange(0, num_rows) %}
{% let multiplier = randint(3, 9) %}
{% let blank_num = list(range(start_num, start_num + num_rows))[blank_idx] %}
| n | ? |
| - | - |
{% for n in range(start_num, start_num + num_rows) %}
| {{ n }}  | {% if n != blank_num %}{{ n * multiplier }}{% endif %} |
{% endfor %}\
What is the missing number?
```

Options

☑ {{ blank_num * multiplier }}

□ {{ blank_num + 1 * multiplier }}

□ {{ blank_num * multiplier - 1 }}

□ {{ blank_num * multiplier + 1 }}

+ New Option

References ▾

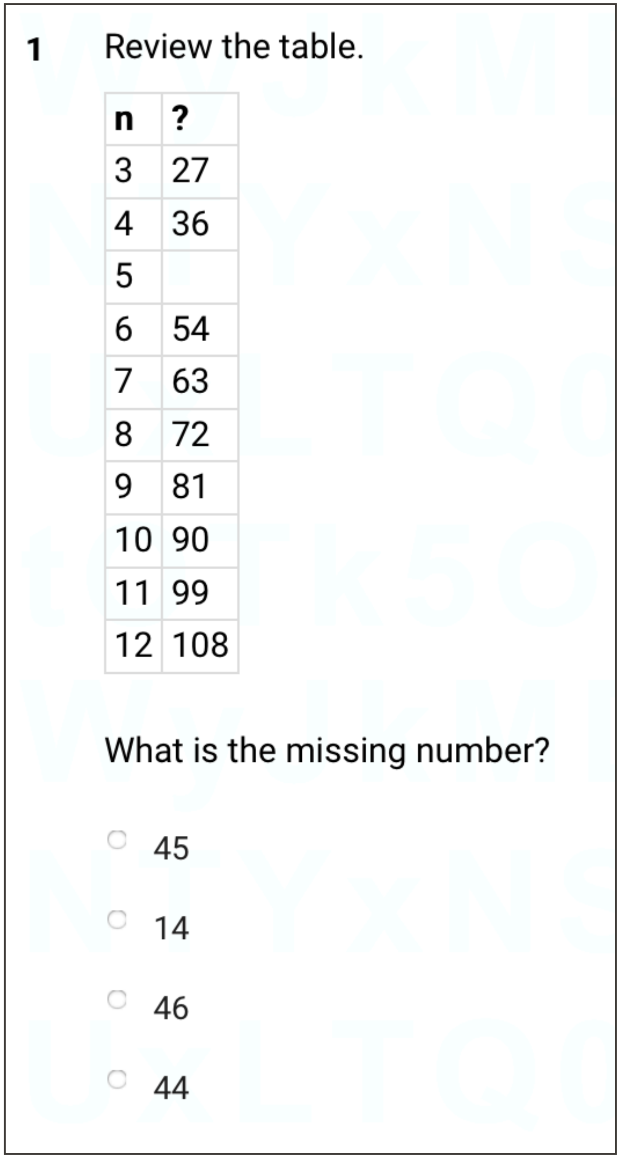
SmartItems: A SR Alternative

What an examinee sees: Each examinee sees something different.

Comments: SR variant with the great benefits.

- Stops theft completely
- Stops almost all cheating
- Indestructible; saves most development costs each year
- Motivates the proper kind of learning and preparation

Advice: Investigate how it works and how it can help you.



1 Review the table.

n	?
3	27
4	36
5	
6	54
7	63
8	72
9	81
10	90
11	99
12	108

What is the missing number?

○ 45

○ 14

○ 46

○ 44

Conclusions

Selected Response is still critically needed.

MC needs to be replaced or used sparingly.

SmartItems and DOMC should be considered, even if today's systems don't support them. Systems must change.

Even newer item designs should be invented.