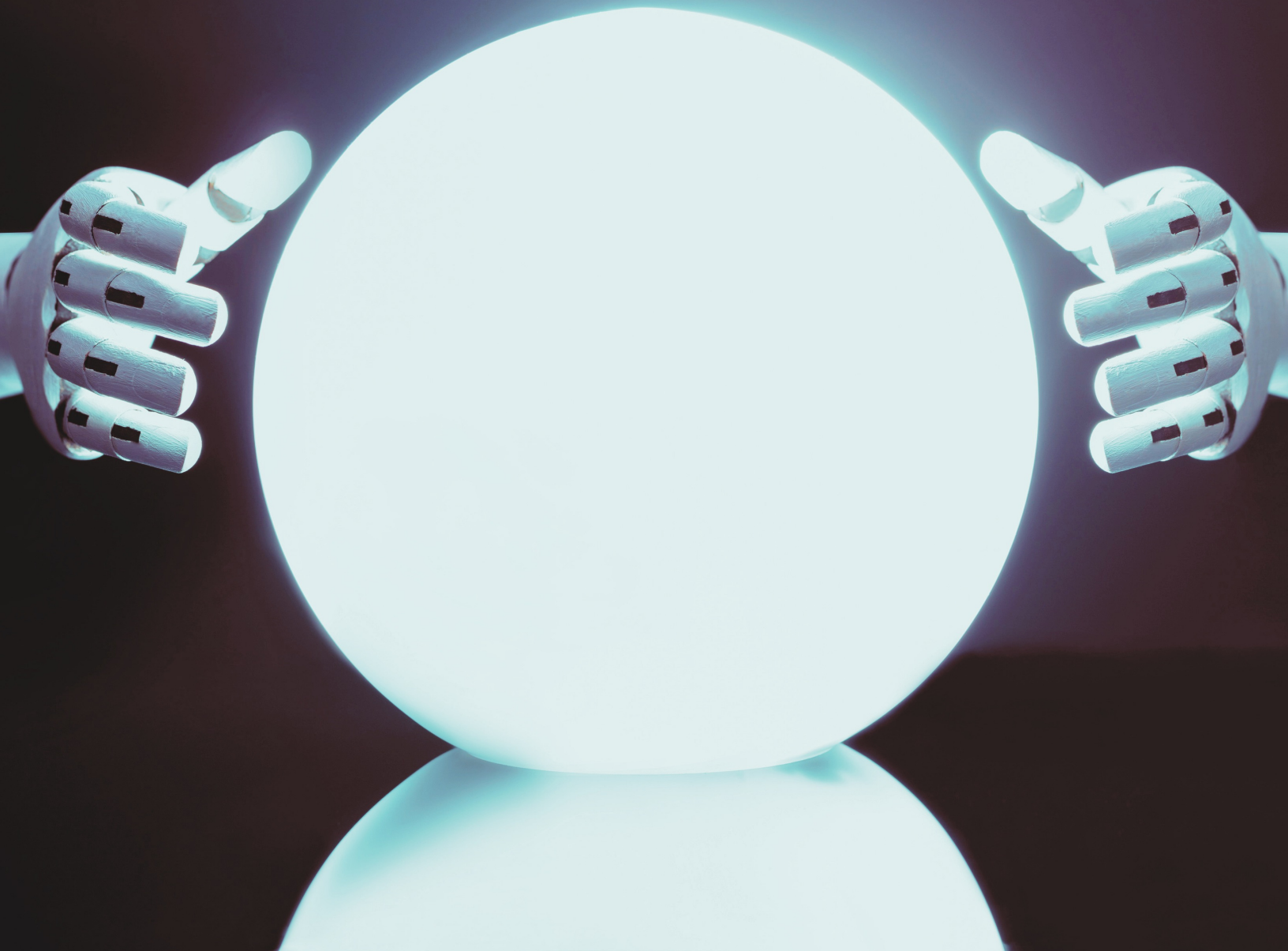# THE FUTURE
OF
MULTIPLE-CHOICE

BY DAVID FOSTER | CEO | CAVEON TEST SECURITY

APRIL 2021

INTRODUCTION

As society's needs change, technology evolves to meet them.

At the turn of the 20th century, innovation was booming. Among the most noteworthy innovations of that time are three that have significantly impacted modern everyday lives: in 1908, Henry Ford introduced the first mass-produced automobile that was affordable to most Americans; in 1915, the first transcontinental telephone call took place from New York to San Francisco. Within the testing industry, there was one invention that altered the way assessments would work for the next century. In 1915, Frederick J. Kelly introduced the **multiple-choice item**.

Teddy Roosevelt is attributed with saying: "The more we know about the past, the better prepared we are for the future." In accordance with this wisdom, I am going to take you on a journey into the past in hopes we can gain insight into where both test security, and the testing industry as a whole, is headed. In this white paper, I will discuss these three notable inventions of the 20th century and how they have adapted to meet society's changing needs over the years. I will then discuss the multiple-choice item, its evolution, and the need our society has for an innovation that addresses the technology of the 21st century. Finally, I will introduce the SmartItem, explain how it works, describe its benefits, and illustrate the ways SmartItem technology addresses the changing needs of test takers and test programs alike.
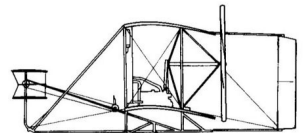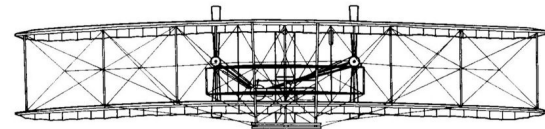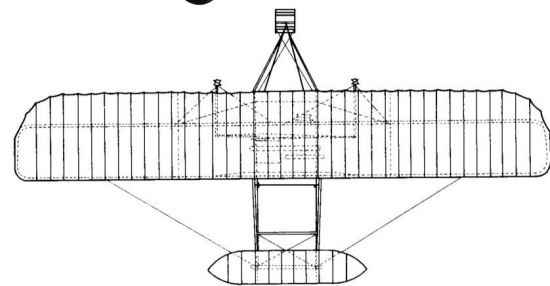
NOTABLE INVENTIONS OF THE 20TH CENTURY

THREE STORIES OF HOW INNOVATIVE TECHNOLOGY MET SOCIETY'S NEEDS

Plato said, "Necessity is the mother of invention." The telephone and the automobile were both invented near the turn of the twentieth century. Each of these innovations was created to meet a gap in society's needs, and each evolved over the years as society's needs shifted.

The Evolution of the Telephone

In 1876, Alexander Graham Bell and his assistant, Thomas Watson, held the first telephone call in history. Paramount to the communicative needs of that time, the telephone replaced the expensive and far less capable telegraph. From there, the first permanent outdoor telephone wire was completed, subscribers began to be designated by numbers instead of their names, and eventually, in 1927, the first transatlantic service from New York to London became operational.

Since then, the telephone has **evolved dramatically** to suit society's needs. Communication was the initial need, but today, the phone isn't only a way to communicate; it's a way to do many other daily tasks: reading, keeping track of our calendars, searching the internet, and more.
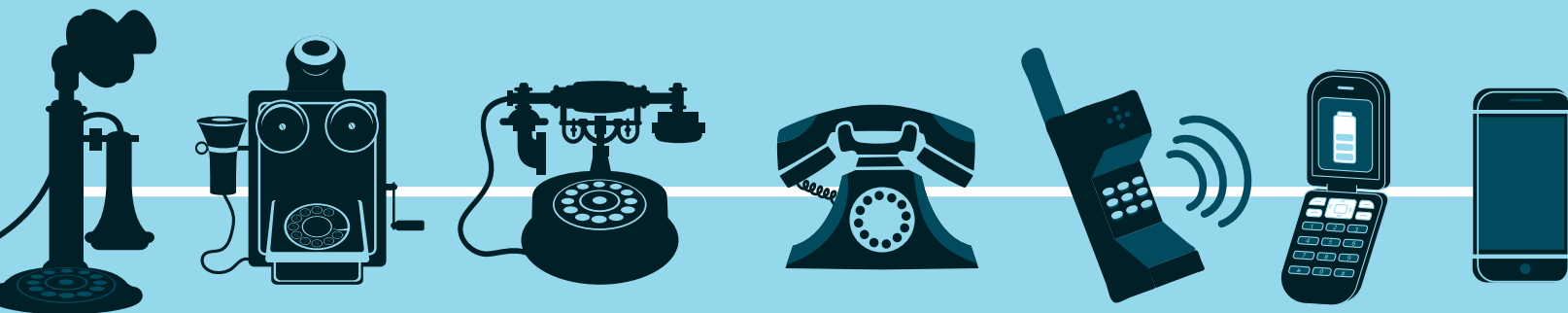
The Evolution of the Automobile

In 1908, Henry Ford introduced the first <u>mass-produced automobile that was affordable</u> to most Americans: The Model T. The need for the Model T was transportation, but there were already other cars before this version. What Henry Ford really invented was not necessarily the car, but a way to produce the car quickly and inexpensively so that people everywhere could purchase automobiles for themselves.

Like the telephone, the automobile has evolved dramatically to meet society's changing needs. Transportation was the initial need, sure. But then society's needs evolved to include affordability, then safety, and now luxury features. There are are even videos of a <u>Tesla in space</u>—actual video footage of a car orbiting the Earth. I'm not sure we're going to drive automobiles in space, but this contrast does show how things have changed. Unquestionably, cars are still used for transportation. But they're used to take you to your destination safely, quickly, and enjoyably. Things have changed quite a bit from the hand-cranked Model T in 1908.
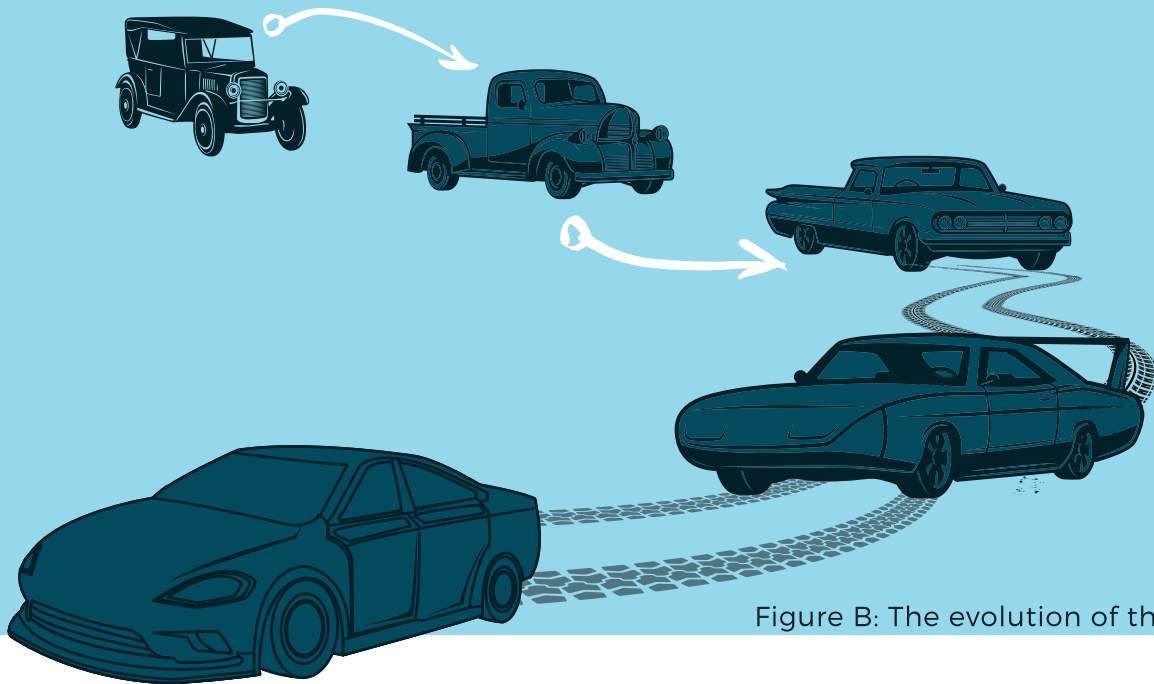
Figure B: The evolution of the automobile

These inventions, and dozens of others from this same time period, were early versions of the same technology we rely on today. In their wildest dreams, the inventors could not have seen how their inventions would evolve over time to fit society's needs. I'm sure Alexander Graham Bell could never have imagined that his telephone would evolve from a wall-mounted box to a handheld computer. Meanwhile, the Model T has gone through dozens of adaptations and has morphed into self-driving and all-electric cars that are on the road today. Over time, society's needs changed, and these forms of technology adapted to meet them. The result is that nowadays, we cannot imagine life without them.

THE EVOLUTION OF THE MULTIPLE-CHOICE ITEM

There is one invention from the early 20th century that is often overlooked. Honestly, it is rarely even thought of as an invention at all, even among those of us in the testing field who use it every day. Like the others, it was created to solve an important societal need, and its impact has been profound. I'm talking about the multiple-choice item.

Like telephones and automobiles, the multiple-choice item is still in use over a century after it was first invented. But while the wall-telephone has evolved into a smartphone capable of instant communication with any corner of the globe, the multiple-choice item has remained the exact same as the day it was first invented. As society changes, so should technology. It is a process we are familiar with, even expect—especially in our current technology-based world. So why is this not the case with the multiple-choice item?

The History of the Multiple-Choice Item

In 1914, Frederick J. Kelly wrote his dissertation from Columbia University in New York on scoring errors made by teachers as they graded students' tests. In this dissertation, Kelly documented a wide variety of errors. Some of the errors were from simple carelessness and the difficult process of grading these tests. Some errors were purposefully made by teachers in order to give certain students a better chance of being admitted into a university. The tests in the early 1900s were made up of (what we would now call) "constructed-response" items. On a test, students would be asked a question, either orally or on paper, and would construct a response in the same manner. Teachers and their assistants would then score the responses. Kelly analyzed this process and determined that it resulted in a large number of errors.

As a result of this experience, Kelly was determined to create a "standardized" method of testing that reduced the number of errors, both accidental and biased. The result was the multiple-choice question type, first introduced by Kelly in 1915 on the Kansas Silent Reading Test.

TEST I, FOR GRADES 3, 4 AND 5

Directions for Giving the Tests

After telling the children not to open the papers, ask the children on the front seats to distribute the papers, placing one upon the desk of each pupil in the class. Have each child fill in the blank space at the top of this page   Then make clear the following:

Instructions to be Read by Teacher and Pupils Together

This little five-minute game is given to see how quickly and accurately pupils can read silently.   To show what sort of game it is, let us read this:

Below are given the names of four animals.   Draw a line around the name of each animal that is useful on the farm:

cow   tiger   rat   wolf

This exercise tells us to draw a line around the word, cow.   No other answer is right   Even if a line is drawn *under* the word cow, the exercise is wrong, and counts nothing.   The game consists of a lot of just such exercises, so it is wise to study each exercise carefully enough to be sure that you know exactly what you are asked to do. The number of exercises which you can finish thus in five minutes will make your score, so do them as fast as you can, being sure to do them right.   Stop at once when time is called.   Do not open the papers until told, so that all may begin at the same time.

The teacher should then be sure that each pupil has a good pencil or pen.   Note the minute and second by the watch, and say, *begin*.

Allow exactly five minutes

Answer no questions of the pupils which arise from not understanding what to do with any given exercise.

When time is up say *stop* and then collect the papers at once.

No. 1.   Value 1.2

I have red, green and yellow papers in my hand.   If I place red and green papers on the chair, which color do I still have in my hand?

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

The multiple-choice question was an invention, and like the phone and the automobile, this innovation was created because society needed it. Society needed a fair way to score tests (without the errors Kelly saw in his dissertation research). To achieve fairer scoring methods, it was clear to Kelly that more standardized methods of testing were needed, and his multiple-choice question helped achieve this goal. An additional benefit of the multiple-choice question was that it reduced the amount of time needed for test administration and scoring.
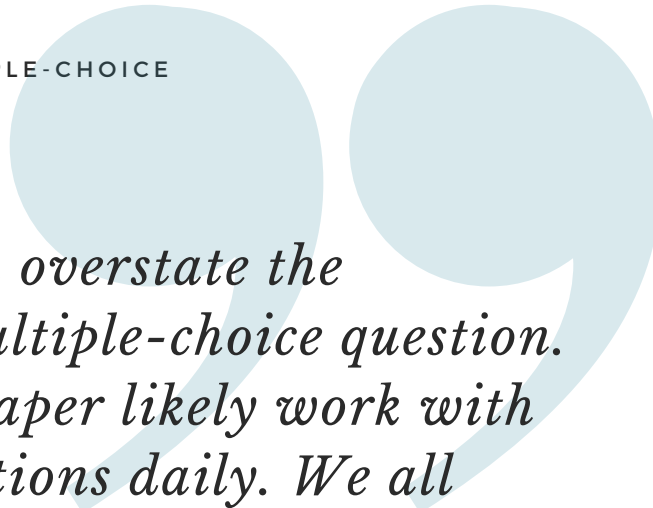
Kelly's influence did not end with the Kansas Silent Reading Test, however. When the United States first entered World War I, a group of individuals called the Vineland Group (led by Robert Yerkes of Yale University) were commissioned by the U. S. Army to create a test that would assess and classify the millions of new army recruits. The goal was to quickly and effectively sort each recruit into the military position that would best match their aptitudes, abilities, and interests. One member of the Vineland Group, Arthur Otis, was familiar with Kelly's work and suggested using the multiple-choice question type on the test. It only took two weeks before the entire group agreed that the Army's recruitment exams should be entirely based on the new multiple-choice item. In 1917, the Army Alpha multiple-choice test officially went into use.

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

"I don't think we can overstate the importance of the multiple-choice question. Those reading this paper likely work with multiple-choice questions daily. We all have a debt to F.J Kelly for coming up with this innovative item type. It may have been inevitable, but it's nice to have brilliant people pushing these concepts."

The U.S. Army used multiple-choice items to assess recruits more quickly and with fewer scoring errors than ever before, which certainly helped the war effort. After this, the use of this item type spread, and it was eventually institutionalized in both the education sector and the workplace, in the United States and abroad.

The Ubiquity of the Multiple-Choice Question

Since its inception, multiple-choice has remained the dominant item type. It is used in virtually every country in the world and in all varieties of tests, whether they are for education or the workplace, paper-based or computerized, high-stakes or low-stakes. Its prevalence is truly astonishing.

However, the multiple-choice item type has had only minor adjustments since it was first invented, and these adjustments have not had much effect. The multiple-choice item used today is not much different from what F.J. Kelly invented more than 100 years ago.
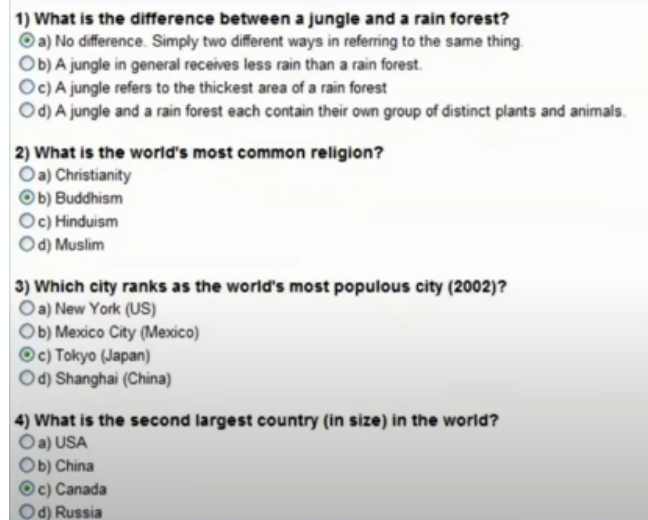
Like multiple-choice, the telephone and automobile discussed previously are still in use nearly a century after they were first invented. However, while multiple-choice remains unchanged, consider how phones and automobiles have altered in the past ten decades. The wall-telephone has evolved into a smartphone capable of instant communication with any corner of the globe, and the Model T has gone through dozens of adaptations and has morphed into the self-driving and all-electric cars that are on the road today.

As society's needs changed, these inventions have quickly evolved to meet them. It is a process we are familiar with, and even expect, especially in our current technology-based world.

21st-Century Testing Needs

At the beginning of the 20th century, testing needed to do three things:

1 Assess large numbers of people in a relatively short period of time

2 Reduce or eliminate scoring errors

3 Reduce the time and effort for test administration

Kelly's invention of the multiple-choice question addressed (either in full or in part) these needs. But more than one century later, society's needs have changed. Here are just a few of society's needs in testing today:

1 Improve the security of exams

2 Improve the fairness of our exams

3 Reduce costs of test development

4 Reduce costs of test administration

5 Improve the convenience of testing

6 Reflect depth and breadth of content knowledge rather than encouraging "teaching to the test"

Multiple-choice, in its original form, does not offer solutions to these problems. In fact, it may actually be contributing to some of them.

The question then becomes: "**How can multiple-choice evolve to help address the current needs in testing?**"

The Non-Evolution of the Multiple-Choice Item

Despite the pervasiveness of the multiple-choice question, very little has changed since 1915. The dominant item type today in tests is the single-correct, four-choice multiple-choice item. While the content may have changed, the question looks and behaves exactly the same. The question we must ask ourselves is: Why is this? Why hasn't the multiple-choice question evolved similarly to the phone and automobile?

The historian, Samelson (1987), wondered the same thing in his written report. After describing the history of the multiple-choice item, he questioned:

"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?"

Samelson asked that question in 1987. Yet here we are, decades later, and most testing programs are still working with that same multiple-choice question.

THE SMARTITEM™

An Evolution to the Multiple-Choice Question

In 2018, Caveon introduced the SmartItem, a new version of the traditional multiple-choice item type that significantly changes the look and operation of selected-response items. The SmartItem was invented primarily to help with test security threats but has other significant benefits as well. It is an evolution and shows promise as a replacement of the multiple-choice item. You can access more in-depth information about the SmartItem by reading various infographics and booklets, research studies, and eBooks written on the subject, or by visiting the Caveon website, but I will cover the basics here.

How is a SmartItem different from other items?

- **A SmartItem covers (or is able to cover) the entire skill** as described in a competency statement, learning or assessment objective, or educational standard.

  - Take this objective as an example: "The student can add or subtract 2-digit numbers." The SmartItem for this objective will be built to use all 2-digit numbers from 10 to 99, and both operations of addition and subtraction. One item covers the entire skill set.

- By design, **the SmartItem will present a different version of the item each time it is given to a test taker**. By definition, each item version is congruent with the objective or standard or competency.

- **A SmartItem renders the items on-the-fly** as part of the item presentation to the test taker.

  - Just prior to the test taker seeing the item, a version of the item is rendered. These were not created in advance and reviewed. They were simply rendered on-the-fly, and they would be one of hundreds, perhaps thousands, or even millions of versions that would come from one SmartItem.

- Each item version rendered from a SmartItem cannot be, and **does not need to be, field tested** prior to its use.

No doubt, those four characteristics make you wonder whether we're crazy, or whether this could be possible. Clearly, a SmartItem is not a type of Automated Item Generation (AIG)—AIG creates item versions but does so with the purpose of expanding an item bank with static items that will then be reviewed and perhaps field tested.

HOW IT WORKS

Behind the Scenes of the SmartItem

Determining the Skill

Since the SmartItem covers the entire breadth and depth of a skill, creating a SmartItem starts with just that: the skill. A SmartItem focuses on the skill more than a traditional item. The image below shows a development screen for a selected-response item for the Common Core State Standard CCSS.MATH.CONTENT.3.OA.A.4.

This standard reads:

> *"Determine the unknown whole number in a multiplication or division equation relating three whole numbers. For example, determine the unknown number that makes the equation true in each of the equations 8 × ? = 48, 5 = _ ÷ 3, 6 × 6 = ?"*
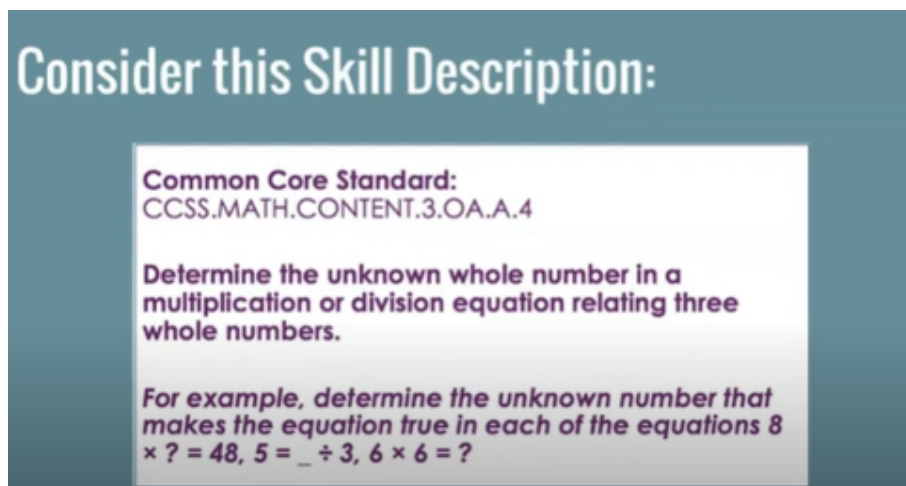

Notice how it covers both multiplication and division, the same standard. As you can see in the question, the unknown number could be in different places in the equation. With the unknown number being presented in different areas of this question, the SmartItem should likely present versions of the item where the unknown number changes the position before and after the equal sign. Each of these points should be considered when creating the SmartItem. The process of creating a SmartItem starts by truly understanding the skill being tested and designing a question that can cover it entirely.

Creating Item Variations

Now that you have a feeling for the skill description behind this item, let's look behind the scenes of this particular SmartItem. You can create a SmartItem using a graphical user interface, code, or a plethora of response options. Then, you beef it up with various content to cover the entire domain.

1   **Create the Item**
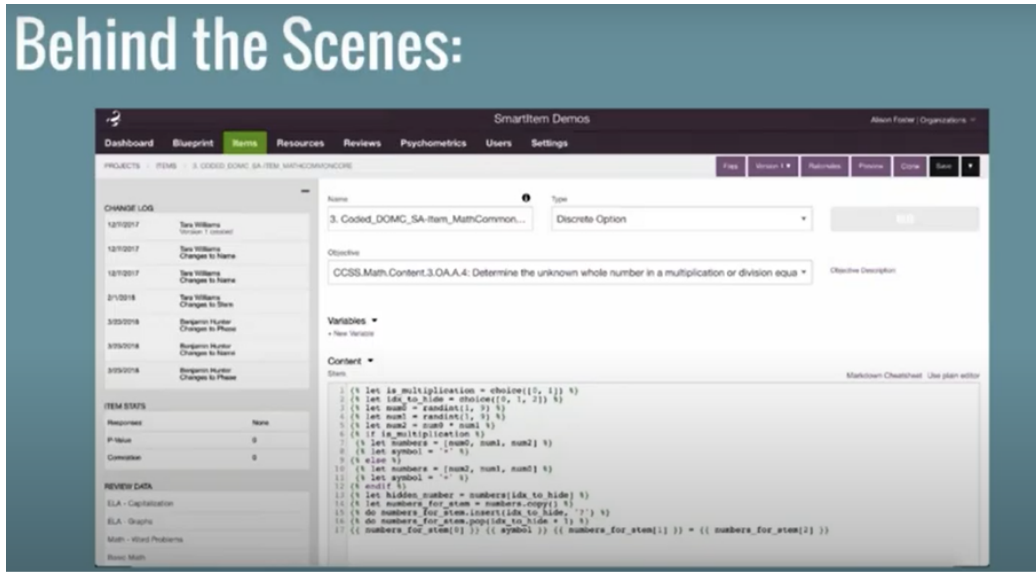
There are several ways to create a SmartItem. You can use code, you can use a Graphical User Interface (GUI), or you can write a plethora of items. The above graphic displays the item authoring tool, <u>Scorpion</u>™, which supports creating SmartItems. While coding necessitates a special skill set, using a GUI and/or extensive options to create a SmartItem can be done by any item writer.

Returning to our third-grade common core question, this SmartItem was created using 17 lines of code (seen in the image below), but could also have been created using Scorpion's GUI interface.

If you want to run this item a few times to see how variations are produced, <u>please click here</u>.

You will see one SmartItem repeated five times. Pay attention to the changing numbers, position of the equal sign, and location of the unknown number.

```
Stem                                                    Use template code editor Use plain editor
 1  {% let is_multiplication = choice([0, 1]) %}
 2  {% let idx_to_hide = choice([0, 1, 2]) %}
 3  {% let num0 = randint(1, 9) %}
 4  {% let num1 = randint(1, 9) %}
 5  {% let num2 = num0 * num1 %}
 6  {% if is_multiplication %}
 7    {% let numbers = [num0, num1, num2] %}
 8    {% let symbol = '×' %}
 9  {% else %}
10    {% let numbers = [num2, num1, num0] %}
11    {% let symbol = '÷' %}
12  {% endif %}
13  {% let hidden_number = numbers[idx_to_hide] %}
14  {% let numbers_for_stem = numbers.copy() %}
15  {% do numbers_for_stem.insert(idx_to_hide, '?') %}
16  {% do numbers_for_stem.pop(idx_to_hide + 1) %}
17  {{ numbers_for_stem[0] }} {{ symbol }} {{ numbers_for_stem[1] }} = {{ numbers_for_stem[2] }}
```

2 **Using Response Options**

This same third-grade item can then be "beefed up" using a different response format, in this case the Discrete Option Multiple Choice™. (Learn more about DOMC™ in this white paper.) The image below shows a sample DOMC item with a single option showing. Using DOMC improves this item by enhancing security and improving fairness by decreasing the impact of test taking skills. In addition, using DOMC increases the total potential pool of questions rendered by this SmartItem to an astonishing 1,317,226 item variations!

If you want to run this item a few times to see how variations are produced, please click here.

You will see one SmartItem repeated five times. Pay attention to the changing numbers, position of the equal sign, and location of the unknown number.

A sample DOMC item with one option showing.

3 **Content**

Remember, in order to cover an entire objective, a SmartItem will need to include the appropriate amount of content. For example, if the objective requires a student to differentiate between mammals and non-mammals, the SmartItem would incorporate the names and characteristics (live birth, hair, etc.) of hundreds of common mammals and perhaps even a greater number of obvious non-mammals to adequately cover that particular content domain. If the objective requires the student to identify the amendments to the U. S. Constitution that protected civil liberties and civil rights, then the SmartItem would include all the relevant amendments.

Regardless of whether a large or a small amount of content is assumed in the objective, it is important that the SmartItem covers all of it as part of the item design.

In summary, a SmartItem begins with creating a skill around the objective. Next, a SmartItem creates item variations using either a Graphical user interface, code, or many response options. Finally, the SmartItem is filled with content to cover the entire domain. In the end, you have a SmartItem that covers an entire objective and can render in millions of ways.

THE BENEFITS OF THE SMARTITEM

How the SmartItem can help your testing program

The SmartItem characteristics described on page 10 combine to provide some startling benefits that address testing's 21st-century needs. Here are the benefits of SmartItem technology:

Security (Theft):

This may sound exaggerated or overstated, but it is not: **SmartItems make stealing test questions and sharing them a useless exercise**. With so many renderings, examinees can no longer cheat by sharing test content, by buying questions and answers, by recording an entire test and sharing it with a friend, etc.

Stealing and sharing questions provides no useful information to other test takers who might try to use that pre-knowledge to cheat.

Security (Cheating):

The most damaging form of cheating seen today depends on someone else successfully stealing the question content beforehand. If harvesting is rendered useless by SmartItems, then cheating by having pre-knowledge of the test content will not be effective either. Other types of cheating are made more difficult as well.

There are two types of cheating the SmartItem does not impact: proxy test taking and having a third party look over your shoulder and divulge the answers. Fortunately, both of these methods of cheating can be caught by proctors or by systems that ensure a test taker's identity.

Cost Savings:

Regardless of the size of the testing program, item creation and re-creation is quite an expensive undertaking (it can cost upwards of $2,500 for large testing programs to re-create a single item.) However, with the SmartItem, continual item development and item replacement are no longer needed. Since the SmartItem cannot be compromised, it does not need to be revised or maintained throughout its lifespan. Thus, the efforts we put forth today to replace items for whatever reason are no longer needed.

While it may cost more to initially create the SmartItem (they certainly take more time to create than a typical item—though it would undoubtedly take less time for a subject matter expert to create one SmartItem than it would to write five traditional items), they save substantial costs in time, effort, and money when it comes to re-writing and re-developing tests.

Learning and Preparation:

The act of "**teaching to the test"** is an unfortunate byproduct of current testing policies and procedures. Luckily, teaching to the test is no longer encouraged when a test utilizes SmartItem technology. The only way an examinee can prepare for an exam composed of SmartItems is to become competent across the entire set of skills the test measures. While this is a different way than we're used to learning and preparing to take tests, the result is deeper and better teaching, training, and learning.

Fairness:

Tests today are unfair for two main reasons. The first is access to pre-knowledge, whether that be through an exposed item or because the individual has taken the test before. The second is test-taking skills, which is when an examinee is able to "game the test" because they have experience or training in taking that type of exam. Individuals with test-taking skills are reported to have a 5-10% higher score than those without. Most recently, when a major IT organization converted their test questions to versions of the SmartItem, pass rates in India dropped from 88% to 8%, indicating that almost everyone in that region was using braindump material and cheating using pre-knowledge on the exam—which the SmartItem effectively stopped dead in its tracks.

Because a test compiled of SmartItems will reduce the impact of test taking skills, the test will be more fair to all test takers. Additionally, if DOMC items are utilized on the exam, the impact of test taking skills is further removed. SmartItems level the playing field by removing test-taking skills and provide a fairer system for those who can't access specialized training or don't want to cheat using stolen test content.

Convenience:

Because many of the security threats are neutralized, tests containing SmartItems can be given in circumstances that would previously have been considered "risky." Almost the entire reason for going to a testing center, or to a gymnasium on Saturday mornings, is for security reasons. However, with SmartItem technology, this is no longer required. For example, a SmartItem-based test can be securely given in a home and monitored by online proctors. As long as the proctors (online or on-site) can authenticate the test taker and prevent coaching and proxy testing, the degree of risk associated with other threats is greatly curtailed, and it is quite possible that the test will have equal or better security than in its in-person counterpart.

These are six benefits of utilizing the SmartItem, and I'm sure that the list is not complete and that more benefits will appear as we begin incorporating and refining SmartItem technology. In the meantime, these are significant enough benefits that any testing program should consider using SmartItems as part of their exams.

As you read the above benefits, I hope you were able to imagine what your world would look like if you were able to confidently deploy exams with SmartItem technology—if there would be less headache and more cash available to you, and if you were able to worry less about different issues and enjoy the confidence in decisions you would be able to make based on test scores you can trust. I genuinely hope you feel that way and see that value.

CONCLUSION

The multiple-choice question was state-of-the-art when it was first invented by Frederick Kelly in 1915. It met the needs of the time and has served us well over the decades since. However, while this item type has increased in popularity and scope over the past century, it has not evolved to meet the current needs of society such as security, fairness, learning and preparation, convenience, and cost savings. Unlike technologies such as the phone and automobile, which have continually adapted and improved, the multiple-choice item has remained frozen in its original iteration. It is time for us to embrace innovation and technology and bring the multiple-choice question into the 21st century where it can evolve to address our current needs and proactively tackle the problems of the future. Entering the SmartItem.