

IQ tests and their validity in the 21st-century — Observatory of Educational Innovation

Sofía García-Bullé

We live in a culture that sees intelligence as part of a person's value. Smart people get into the best schools, obtain the best jobs, receive the highest salaries, are the leaders, the examples everyone else should follow. If you are the smartest person of a designated group, it is normal that people follow you and that a lot of doors open for you, figuratively. Being smart is a crucial component of success, but... do we measure it correctly?

IQ tests are a fundamental tool to measure student intellectual capacity. We consider them mathematically and scientifically precise, but are they really that? Or are we purposely blinding ourselves to the cultural contexts that power the biases behind the very concept of the intelligence quotient and the tests we use to measure it?

Cultural bias and the use of IQ tests as a political weapon

To understand the weak point of intelligence tests, we need to take into consideration the time in which they were created. At the beginning of the 20th century, an evolutionist philosophy dominated science and humanities. This way of thinking influenced even the way people constructed and shared knowledge.

When the first tests came out, the idea was not to use the test results for educational innovation to create new teaching strategies that could work to improve different ways of learning. The objective was to secure the survival of the fittest; the fittest were the smarter ones; the smarter ones got higher scores. But were the people who fared better the most intelligent, or were the high-scorers people with an advantage because the tests were culturally designed for them?

Back then, cognitive science supported the evolutionary theories that paved the way for eugenics. If Binet, Simon, and Terman took into consideration only the psychological profiles, mindsets, and circumstances of people of their same race, class and even gender to map how intellectual capacity works, can we say that their test is impartial and accurate?

Is it ok for us to ignore the historical proof that these tests justified racial discrimination and educational gentrification? The notion of people of diverse races having different levels of mental capacity was prevalent at the time, even among the scientific community. This misguided belief impacted the way these tests were created and applied.

Terman himself thought that racial minorities like Native Americans, African Americans, and Latin Americans had less capacity to understand abstract ideas than white people. He also believed that they made up for it in resilience, which from his point of view, made them hard workers and good at following orders.

These arguments are proven false and are indefensible nowadays. But back then they dictated who would receive the best education and who would be better trained to pass the IQ tests.

This self-fulfilling prophecy was never about education; it was about maintaining the social order. And that purpose succeeded; people saw IQ test results as an indisputable scientific truth, not a consequence of cultural and social inequity.

What do intelligence tests measure anyway?

Intelligence is defined as the capacity to understand concepts or ideas and to solve problems. Under this definition, the Stanford-Binet scale meets its objective. It assesses people's skills to understand questions and to resolve the situations included on the test. The mechanism of the exam is not under discussion, but after 100 years, the way we devise its content should be analyzed and updated.

To Antonio Andrés Pueyo, Lecturer in the Psychology Department at the University of Barcelona, intelligence is a [very complex concept](#), especially in a time in which we make machines with artificial intelligence. According to him, there are aspects of intelligence that do not fit the mechanical vision from which the first IQ tests were created.

"There are tests that evaluate different types of intelligence that combine with IQ. This is what happens with the Weschler scale. Tests can also be built to assess a single skill, like the Raven test," explains Pueyo.

All of these tests have their strengths and shortcomings. In the case of the Bidet- Stanford Test, it still has the right components to measure how apt is a student to navigate standardized learning but not how to apply that knowledge creatively or to think outside the box.

Those who got the highest scores when the test was first created were followed by Terman throughout their careers. Those students went on to get into the best universities and, subsequently, the highest paying jobs, but very few of them went beyond their expected social roles. They did not reinvent the wheel; instead, they mastered spinning along with it.

The test was not perfect. It detected talents like those of Ancel Keys, Norris Bradbury, and Shelley Smith. But it also [failed to discover the potential of Luis Álvarez](#), a student who was rejected by Terman for followup because he fell short of the IQ score cutoff; yet Álvarez went on to win the Nobel Prize for Physics in 1968.

The need to rethink what we know about intelligence and the skills that matter

The challenges that people faced at the beginning of the twentieth century are not the same that we deal with today. They did not have to worry about digital literacy, and similarly, we do not need to know all of our close friends' phone numbers by memory as they did.

If the problems we solve now are different, the skills and intellectual aptitudes required to resolve them should differ too. We live in an era of automation; machines are being improved with artificial intelligence to do jobs previously performed by humans.

This automation is not limited only to mechanical jobs but also more tasks that require analysis are being automated; for example, the profiling of information of people online to use the data to direct ads to them that align with and impact their consumer habits.

In 2018, [an AI system analyzed the most awarded commercials](#) in the last 15 years and used

the information to write the script for a Lexus ad. The results were impressive for a machine, and it leads us to question what jobs humans can do that machines cannot in a not-so-distant future.

Some intellectual activities can be mechanical and therefore performed by a machine, such as the compilation, data analysis, and pattern identifications that the AI system utilized to write the script for the Lexus ad. However, the creative thinking and artistic sensibility to direct the ad could only be executed by a human, in this case, the director Kevin Macdonald.

The demands of the modern labor market require that intelligence tests must register and measure more than just mechanical and intellectual skills. They also need to consider everything related to soft skills, now rebranded as “[power skills](#).” We must learn how to measure and nurture creative thinking, social intelligence, emotional intelligence, and other qualities beyond the existing educational standards.

We must abandon the idea of measuring a person's worth by just one IQ, or seven types of intelligence measurements, or even fifteen or thirty. What we need in these tests is that they assess the diverse ways in which people approach knowledge and problem-solving and how these can be taught. Otherwise, intelligence tests will continue to be a tool for proponents of social dominance instead of the educational resource that would improve the acquisition of knowledge for a better world.