



## How IQ Tests are Scored

Lyndsay T Wilson 1.3K reads

*Administering an IQ test is seldom straightforward. Psychologists and psychometrists first need to choose a test to use, then decide if any supplemental test are needed or whether to omit a subtest entirely. The answers then need to be scored and interpreted according to normed tables designed specifically for that test.*

In the past, when IQ tests were exclusively given to children, IQ was a ratio of “mental age” to chronological age. Here, mental age is the child’s degree of performance on a range of tasks and chronological age is how old they are in years and months.

For example, a 9-year-old who performed the same as the average 11-year-old would earn the score  $(11/9) \times 100 = 122$ . Note that this would be higher than the score the child would obtain if they performed as well as other 9-year-olds in their group:  $(9/9) \times 100 = 100$ . In this way, “average” performance is always a score of 100, regardless of age or of the actual numerical test score earned.

IQ test originator Alfred Binet suggested that this result then be used as a general diagnostic – if a child scored only 70, for example, (i.e. their mental age matched children of a much lower chronological age) then they would benefit from remedial education pitched at that level.

The above formula works because in general, children grow more intelligent with each year of life. However, their intellectual development usually slows at around 16 – 18, making this method unsuitable for adults. To obtain meaningful scores for adults, more effort has to be made to determine what an “average” score for any group really is.

Adult IQ scores are normative and not absolute, which means they have value relative to the population they were normed to, i.e. people of the same group. Because intelligence is so heavily dependent on culture and context, it’s important to remember that no score in itself denotes intellectual ability. Rather, IQ is a statistical measure designed to rank an individual against group performance. Though IQ tests have been rightly criticized for how well they truly accommodate variation in race, class, culture and gender, almost all IQ tests in use today have been extensively refined over time and tested on substantial groups of people from all over the world.

When psychologists design IQ tests, they begin by collecting a sample of people and administering a potential test to obtain “raw scores.” These raw scores are then processed using appropriate statistics – typically a raw mean and standard deviation are obtained, which

are converted to z-scores, which can then be converted to IQ scores, and compared. IQ scores have a normal distribution (they fall on a bell curve), a mean of 100 and a standard deviation of 15.

On a standard normal distribution, a standard deviation is a measure of spread. A standard deviation of 15 points means that 68.2% of the group will fall within one standard deviation of the mean, either above or below. Furthermore, 95.4% will fall within two standard deviations, and 99.6% of all scores will fall within three standard deviations.

Another way to talk about IQ scores is to say what proportion of the norm group falls *at or below* that score. For example, a score of 120 on a normal distribution is nearly in the 91<sup>st</sup> percentile. This means that almost 91% of all scores fall at or below the score of 120.

As you can see, it's incredibly rare to achieve a score that falls four standard deviations or more from the mean (only 0.2% of a group will achieve a score above 145 *or* below 55). Though many people like to boast about high IQ scores, consider that a score of just 120 means they can outperform 90% of their peers – no easy feat.

Besides being rare, such extreme scores are in any case difficult to measure; validity and reliability are less certain if, for example, the test taker has severe attentional or literacy issues. If a test-taker is on the other extreme, the test questions themselves may not be sensitive or challenging enough.

With high-score test takers, the obvious question becomes, who is qualified to devise the questions? Marilyn vos Savant was a well-known child prodigy and purported “smartest person in the world” with the *Guinness Book of World Records* title to prove it. She worked as a writer answering tricky reader questions and mind-bending puzzles in her “Ask Marilyn” newspaper column.

There she was posed a now-classic probability puzzle called the Monty Hall Problem, to which she gave her answer. Her proposed solution, however, was so unpopular that she received thousands of hate-filled letters from academics, mathematicians and logicians around the world who claimed she was wrong – and an idiot for spreading misinformation.

However, decades later, Marilyn was proved correct and her critics were forced to eat their words. While an unusual case, this does show that with very high-caliber and unintuitive problems (like the Monty Hall problem), groups of even intelligent or expert people can be mistaken about what counts as an intelligent answer. Interestingly, Marilyn vos Savant believed that IQ tests weren't particularly valid.

# EXPLORABLE

## Quiz Time!



Quiz:  
Psychology 101 Part 2



Quiz:  
Psychology 101 Part 2



Quiz:  
Flags in Europe

[See all quizzes](#) ⇒

## The Flynn Effect

IQ scores typically fall on a standard normal distribution with 100 set as the mean, and 15 as the standard deviation. Naturally, the standardizing process is repeated every so often on new sample groups – people who are usually younger/born later. However, since around the 1930s, researchers have noticed that when more recent norm groups take the older tests, their group mean is significantly higher than 100.

This increase appears to hold for both crystallized and fluid intelligence, and has been observed in many countries spanning more than 100 years. Named the Flynn Effect after one of its main theorists, James R. Flynn (and coined by the authors of the controversial book *The Bell Curve*, in fact) this apparent increase in intelligence over time has been attributed to many possible causes.

Some theorists have shown that the apparent increase is down to the more extreme scores clustering more tightly around the mean, or that only the lowest scores have gradually shifted upward over time with increase in global education. Others have suggested that the increase is just an artifact of people becoming more adept at very particular cognitive skills and not necessarily gaining intelligence.

“Test familiarity” may be behind some of the increase in more developed countries, while overall improvements in health and nutrition may be responsible in those populations where such factors were actively suppressing potential IQ expression. Whatever the cause, however, and whether the increase continues or eventually slows, any test taken today will have been correctly normed and standardized to reflect the modern population’s abilities.

---

**Source URL:** <https://explorable.com/iq-test-scoring>