Glenn E. Cahn, PhD PLLC
3205 Randall Parkway, #112
Wilmington, NC 28403
910 332 4134

www.ILMpsychtesting.com

# Understanding Psychological Test Scores

## Who and what are being compared?

There are two ways to measure people:

- Against each other. This is something everyone is familiar with, in that any competition involves such a perspective. Who is smartest, fastest, richest, tallest…? Any such competition ranks people from having the most to the least of some attribute. To take education as an example, there are class valedictorians, being the kid in school who is seen to be the brightest student. And most kids are said to be 'average' such as how tall they are for any given age, or how well they can read, write or do arithmetic. Schools and governments typically want every child to meet a certain standard of educational proficiency, which we can also call 'average' ability.
- Against oneself. This type of measurement reflects how much one improves over personal performance of the past. Professional or amateur athletes often measure themselves in such a manner. Someone may have always run the 100 yard dash in 12.5 seconds, and if the individual can now do it in 12.0 seconds, improvement has occurred. There are others who may be faster, but that is not the point. Rather, individuals simply want to know that they have improved themselves in some skill, and can take pride and satisfaction in such personal progress. One example of not caring about ranking individuals against each other are the Special Olympics, where every child who competes can be a different type of 'winner.'

Both types of measurement are valuable. Each has its place in better understanding skills and attributes. Neither form of measurement is inherently better or worse than the other. The important issue is to use each at the right time to understand a person's strengths and weaknesses.

Much of psychological testing does involve the first type of measurement, and it will be the focus of the rest of this article.

## Inches & Meters

We can measure something, such as the distance between two points, by numerous 'yardsticks'. Consider:

- inches, feet and miles
- millimeters, meters and kilometers
- furlongs
- fathoms

- nautical miles
- light years

All of the above systems of measurement are legitimate, and are used at times by different people and needs. Each has strengths and weaknesses. For instance, we would not want to measure the distance to the nearest star in inches. Nor would it make sense to measure the depth of some body of water in light years. There are other times that multiple systems can be employed equally well. For instance, at what temperature does water freeze: 0º or 32º? The correct answer is "Both!" The Celsius and Fahrenheit scales use different numbers to measure the freezing point of water.

The principles being outlined here are that sometimes which 'yardstick' is used may depend on the context. And also that it is possible to convert any of these types of measurement in to another, such as between miles and kilometers.

There are different ways to measure psychological principles. Most people are not familiar with them. But like those illustrated above, such psychological measurements each have strengths and weaknesses, and they can be converted back and forth into each other.

What's Average?

One of the most common questions I get from people I work with is "Am I normal?" Most say it with a bit of humor along with a tinge of anxiety. This paper is not going to address 'normal' but change the focus a bit and talk about 'average.' Sometimes these two terms may be used interchangeably. Whether they are the same will not be answered here. We are simply going to look at the measurement of some skill being 'below average,' 'average' or 'above average.'

None of us is perfect. Nothing that we ever do is perfect. But, we can still look at whether we have a sufficient amount of some skill to function well in our daily lives. Let's look at some common areas that are of interest to psychology and many people: IQ, and education, and use them as illustrations for better understanding test scores in general.

IQ is a familiar concept to most people. It is a way to measure how well people can function on an intellectual level, such as being able to learn in school, or deal with living on a daily basis. What kind of measurement scales are used to understand IQ?

One system is called 'standard scores.' It is a label, much like 'fathoms' or 'furlongs'. How it came to be called that is not important, and why the numerical system it employs developed is also unimportant. So, let's just look at how the system is set up.

'Average' is set at 100. Small units are '1' and larger units ('standard deviations') are set at 15. What this means is that an 'average' person has an IQ of 100. Someone may have an IQ of 98 - but there are no less than whole scores, like 98.6. We can also say that someone with an IQ of 85 is as many points lower than average as someone with an IQ of 115 is higher than average (15 points, in each case).

A second principle to understand about 'average' in psychology is that it is not always a single number. Although an IQ of 100 is 'exactly average' more commonly we talk about 'being in the average range.' For instance, the average weight for an American adult female might be 120 pounds. But, if a woman weighs 119 or 123 pounds is that 'bad'? The 'average range' for an adult female might be 110 – 130 pounds (if we ignore height as an issue) – and any woman who falls in that range is of an average weight.

Quantitative vs. Qualitative

The same principle applies in psychology. The average range of skills, for something like IQ, is set between 85-115. A person may be on the low side of average (closer to 85) or the high side (nearer 115). But, all the scores in that range are still considered average. As people, we often think that 'more is better' or 'bigger is sexier.' For instance, taller people often get more status. Someone with a higher IQ may be given more acceptance or deference by society. Is this justified?

Would an American female who is 5'4" tall necessarily be a better person than another who is 5'3", and worse than a woman who is 5'5"? Or, to make it more personal, would a woman's life be better, happier, more successful if she were 5'5" tall vs. 5'4"? Having greater height may be more sexy. But, given the complexity of life, along with concepts like success or happiness, height has little if anything to do with whether most people feel good about themselves.

The same is true for psychological skills like IQ. Having an IQ of 100 vs. 98 or 103 will have no real bearing on how well someone lives their life. Higher scores can give us bragging rights. But ultimately numbers are just numbers. How we think about their meaning, on a qualitative level is far more personal. For instance, some people may feel they are financially poor if they have 'only $1 million in the bank.' Others would be incredibly happy to have that amount of money. Similarly, what it means to have a higher or lower IQ score is personally determined. I have seen parents who were thrilled when I told them that their child had an IQ of 75 – because they had previously heard the score was 60. Others have burst in to tears in my office when I said their IQ was 120 – because they had friends who purportedly had IQ's of 170.

The point being made here is that psychological test scores only offer quantitative measurement, and not qualitative. How you feel about a score is your personal choice. For psychologists, such test scores are merely comparing how close or far from a numerical average a person falls.

Ease of Understanding

Every American learns inches, feet and miles as a system of measuring length. We all use such measurements on a daily basis, and we are very comfortable in understanding their meaning. The metric system (milliliters and kilometers) is not nearly as well known in the U.S., and many people in this country are far less comfortable with it. But, metrics are far easier to work with in that it uses factors of ten, which makes calculations simple and quick.

So, in looking at units of measurement one needs to remember that what is easily understood may not always be the same as what is simple to work with. (Quick! How many inches are there in a mile? Now, how many millimeters are there in a kilometer?)

Let's look at two different units of measurement often employed in psychology: standard scores and percentiles. We have already spoken about how the 'standard score' system is designed, with an average of 100, and whole units of 1, and standard deviations of 15 being employed.

The advantage of standard scores is that whole numbers are easier to work with over fractions or decimals. The unit of 1 is also always the same, so that a score of '99' is as much less than 100 as '101' is above it. The disadvantage of standard scores is that most people outside of fields like psychology are not familiar with them. Standard scores are sort of like the metric system in the U.S. – easier to work with mathematically, but we are not taught the system, and so it seems foreign and strange to us.

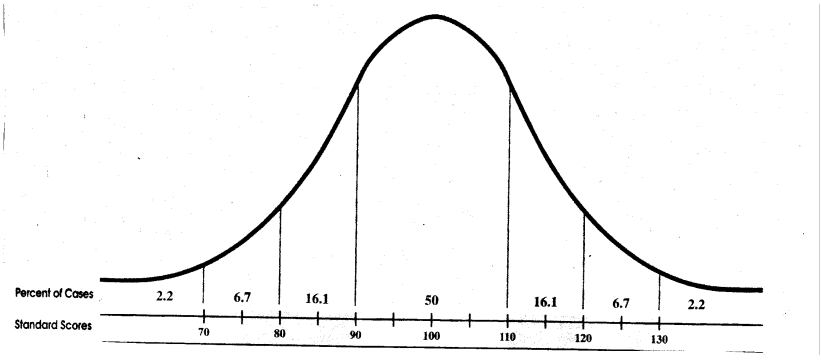
Percentiles offer different strengths and weaknesses. A percentile is a way of saying 'How many lie above this point vs. below it?' That is, if someone is at the 50th percentile, which is exactly average, an equal number of people have more or less of some attribute such as IQ, height or yearly income. So,

- if you are at the 90th percentile, what does it mean? The answer: 10% of people have more of something than you, while 90% have less of it.
- If a person is at the 1st percentile, 99% of people have more of the attribute, and only 1% has less.

Again, there is no qualitative judgment being made here when percentiles are offered. It is merely a rank ordering of who quantitatively has more, an average amount, or less of something.

The disadvantages of percentiles are that they can employ less than whole numbers, and the unit of measurement is not constant. Look at the following table.

| Standard Score | Percentile |
|---|---|
| 55 | 0.3 |
| 65 | 1 |
| 75 | 5 |
| 85 | 16 |
| 95 | 37 |
| **100** | **50** |
| 105 | 63 |
| 115 | 84 |
| 125 | 95 |
| 135 | 99 |
| 145 | 99.7 |



Compare the following sets of numbers:

- 55 and 65
- 75 and 85
- 85 and 95
- 95 and 105

In each case, there are 10 standard score points between these sets of numbers. But, look at how many percentile points there are. From 55 to 65 there are only 0.7 units. From 95 to 105 there are 26 units. Consequently, you can not add or subtract percentile scores the way you can the 'standard' ones. Consequently, measuring change over time for any one person is far harder when percentiles are employed. For instance, if a child is reading at the 0.3rd percentile initially, and with tutoring rises to the 1st percentile, most people would not be impressed. Rising from the 37th to the 63 percentile over time sounds far better. But, the amount of progress is actually the same – 10 standard score points.

The above table is also color coded for a reason. Look at the similar colors, such as the two blue rows, or the two red ones. Now look at the graph to the right of the table, which is often referred to as the 'bell curve' (given its shape) or 'normal curve' which is another name for it.

Many human attributes fit in to this 'normal distribution.' The color coding of the table shows the principle. For instance,

- an IQ score of 85 means that about 16% of people have lower scores and 84% have higher.
- an IQ of 115 means that about 16% of people have higher scores, and 84% are lower.

The graph also illustrates the 'average range' (between standard scores of 85 and 115) which encompasses about 68% of people. Plus, the graph shows that one can go between a standard score and a percentile very easily. (Tables exist that convert each standard score into a percentile, for greater ease of use.)

How much better or worse?

Some parents have asked me to in effect translate a quantitative standard score, say 105, in to a qualitative one ('Is that 'high'?) Or a different way of asking the same question might be 'If tutoring a child raises their standard score by 5 points, is that worthwhile or valuable?'

There are a couple of ways to answer these questions. One already has been mentioned, which is that 'value' or 'how high/good is this score?' is personal, and something that only you can decide for yourself.

A second way to answer the question is to briefly look at a term, 'standard deviation.' This refers to how far scores vary from the mathematical average. With standard scores, the standard deviation is 15 points. Generally, one standard deviation is a noticeable amount. For instance, let's say that the average American adult female is roughly 5'4" tall. And, let's also say that the

standard deviation for women is three inches. So, if you look at women who are 5'1" (a standard deviation shorter), 5'4" (exactly average), and 5'7" tall (a standard deviation taller), the difference in their heights will be obvious.

How easily can you notice the difference between women who are say, 5'3½", 5'4" and 5'4½"? Such small fractions of a standard deviation might be noticeable, but they are not as easy to see or care about as full standard deviation units.

One of the more common examples of minimal changes being given too much weight by society is with SAT scores for high school students. The average score for SAT's was set at 500, with standard deviations of 100. Over the years, media have reported that over some period of time SAT scores rose or fell by say, 5 points. They then exclaimed 'how great!' or 'how horrible!' this was. Five SAT points is $1/20^{th}$ of a standard deviation. Such a change is roughly equal to getting excited over women's height changing by about a seventh of an inch. 'Women used to be 5'4", and now they are 5'4 and one-seventh of an inch tall!' That's not much of a difference.

What fraction of a standard deviation is significant? No precise point can be set, and it is more a personal preference. But, keep the above example in mind when you start to care about small fractions of a standard deviation.

A third way to look at 'how far from average?' something falls is by changing again the yardstick we use, and employ another that might offer greater clarity. For instance, there are other test score yardsticks that are called 'grade equivalents' or 'age equivalents.' Like everything else, they have pros and cons. Their main benefit is that they are fairly easy to understand. Their main downfall is that they are difficult to work with on an arithmetic level, and people can misunderstand their meaning if they are not first explained.

Let's say a child is given academic achievement tests, measuring skills like reading, writing, or arithmetic. What is the best way to report these scores to a parent or teacher, to show if a child is doing ok, very well, or needs extra help? Let's look at some numbers, which might illustrate say, reading ability for kids in $1^{st}$ grade.

| Raw Score | Standard Score | Percentile | Grade Equivalent | Age Equivalent |
|---|---|---|---|---|
| 30 | 76 | 5 | K.1 | 5.4 |
| 50 | 100 | 50 | 1.1 | 6.4 |
| 70 | 116 | 86 | 1.9 | 7.4 |
| 90 | 136 | 99 | 3.1 | 8.8 |
| 110 | 156 | 99.9+ | 6.8 | 12.0 |

What the headings in the table mean:

- Raw score: how many items were answered correctly. If a test has say, 150 questions on it, this would be how many were gotten right.
- Standard score: as already noted, this is a scale that is created, where 100 is exactly average. So a child who correctly answer 70 items, earning a standard score of 116 is just over 1 standard deviation above average. (Remember: a standard deviation in this system is 15 points.)

- Percentile: how this child ranks against other kids in the 1st grade.
- Grade equivalent: this does not mean that a child is capable of doing the work of some particular grade level. For instance, if a child in 1st grade earns a raw score of 110 on this test, the implication is not that 2nd through 5th grades can be skipped, and the kid can instantly go in to 6th grade. Grade equivalent (with tenths of a grade shown after the decimal point) means that the raw score is equal to that of a child at such a level. So, the average child in 1st grade is getting 50 items right on this particular test. The average child near the end of 6th grade is getting 110 items right on the test. If your child correctly answers 110 items on the test, then their score is equal to that of an average kid at the end of 6th grade. But, it is wrong to assume that your child is capable of doing 6th grade work.
- Age equivalent: the same principle applies as for Grade Equivalent, only the number after the decimal point is expressing months. So, a raw score of 110 on this test is equal to what the average 12 year old child earns.

Which of these measuring units is the easiest to understand? It really depends on you, and what your preference is. Grade Equivalents (GE) are commonly used in schools, although as noted above, they may be misunderstood – and do not mean 'doing the work of' a particular grade. Grade Equivalents do give some indication of how far behind or ahead a child is in some skill. For instance, saying that a child is at the 5th percentile for reading in 1st grade tells you they are near the bottom of the class. But, a GE of K.1 shows that the child is working at the same level as someone a full year behind. Given how much learning occurs in that single year, it might be a better way to represent that there is a serious lag in reading skill, and perhaps the child should be held back and repeat 1st grade. The 5th percentile, or GE of K.1 both represent the same score. It is up to you as a parent or teacher to choose which is more easily understood or can be used in a meaningful manner.

When are 'age equivalents' best employed? I use them sometimes when kids have stayed behind in a grade for some reason. For instance, kids who are mentally retarded typically are not in regular classes, or if they are, may not be promoted every year. So, in a situation like mental retardation, academic performance is often reported as 'this child, who is age 'x' is like a 6 year old.' But, one also often hears 'this child, age 'x' has a 1st grade ability.' Both are acceptable, and again, it is a matter of personal preference as to which is used or more easily understood and applied.

How are test scores employed?

A person has undergone psychological testing for some reason, and the measure has been scored. Now what? At some point there has to be an interpretation as to what the score means, and what if anything should be done about it. A common example of this problem arises with car owners who take their vehicle back to the dealer because of some complaint, such as it making a noise or lacking power. And the service writer says 'They all do that, it's normal.'

There are two possibilities for that 'it's normal' response: accurate or inaccurate. In psychology, when some issue which occurs commonly is perceived by a professional as being a problem, it's often referred to as 'over-pathologizing.' As an example, there is currently considerable controversy as to whether kids who are physically active have attention deficit disorder (ADHD)

and need treatment such as being put on medication like Ritalin – or are they simply 'being kids?'

But there is the flip side of this possibility. Consider an adult who has a blood pressure of 130/85. Is that 'high' and in need of treatment? Medical doctors might say that it is 'borderline high' and hold off on prescribing medication. But, what happens if ten years down the road research shows that a b/p of 130/85 leads to 'x'% increased risk of stroke or heart attacks? How big does 'x' have to be before medical science says that medication is now warranted? And who determines that magical tip point? Your doctor? The government? The pharmaceutical companies that manufacture blood pressure medications?

The same principle applies to psychological test scores. With extreme scores (like, a b/p of 200/150) the answer becomes obvious. But, with more subtle scores problems arise. A common one is that people generally look out for their own best interest. Using the above example, pharmaceutical companies like making money on drug sales. Governments or insurance companies like to pay fewer bills, and keep more money for themselves rather than reimbursing the public. So who is the 'right' arbitrator as to whether a person should be put on b/p medication?

The point being made here is that you should always consider if the interpretation of a test score, such as in the recommendations that follow, reflect a bias in the professional's favor. If it does, that by itself is not automatic proof that the recommendation is wrong. In our example, high blood pressure may well be present, and it may be best to treat it with medication. Being too cynical can be hazardous to our health. But, the presence of what may be professional bias should raise some cautionary flags in your mind, and cause you to further investigate the issue.

Who wants to be average?

As I already have mentioned, how we react to a test score, on a qualitative level, is a personal choice. But 'being average' is generally not very sexy or appealing. Most people like to think of themselves as being better than average in some attribute, whether it is intelligence, talent for doing something, good looks, or honesty.

So, what is 'average' and is it good or bad to be such?

For many psychological attributes, such as those that fall in to the 'normal curve of distribution' being average is defined as the score that roughly 68% earn. If you want this translated so it might be more easily understood, it is all the people who fall between the $16^{th}$ and $84^{th}$ percentiles. Anyone in that range is technically 'average.'

However, 'average' is a statistic. Let's take a theoretical example to illustrate why 'being average' can be a complicated issue to understand.

Let's say that all $1^{st}$ grade kids in this country are taught how to read using the old-fashioned method, including your oldest child. When their reading ability is measured with a particular test like the one discussed above, the average $1^{st}$ grader gets 50 out of 150 items right. Consult the

chart: a raw score of 50 equals a standard score of 100, the 50th percentile, and a Grade Equivalent of 1.1. All these numbers mean the same thing: exactly average.

Now let's say that some new and improved method is invented that boosts reading scores for 1st graders. You have a second child who is taught this new way, as are all other children in the country. With the new method most kids in 1st grade get 70 of 150 items right. Your second child gets a score of 70. Which of your two kids is smarter? Technically, both are 'average.'

How can this be? There is a popular radio show where the standing joke is 'Every child in the town is above average.' The humor is that if all kids are 'above average' then automatically they must be defined as being 'average.' A more common example of this problem – of how 'average' changes when norms are updated, is found with what people should weigh according to the table of a particular life insurance company.

Let's say that according to the official height/weight table of the life insurance company, women of a particular height can weigh between 110-130 pounds and be considered 'average.' Over time, medical research learns that a few extra pounds of weight dramatically boosts the risk of diseases like heart attacks, strokes and diabetes. And new research on women who are too skinny, perhaps from anorexia, it is learned that there is also a dramatic rise in disease and death.

So, one day the life insurance company revises its tables, and now says that women should weigh between 115-125 pounds. If a woman used to weigh 110 or 130 pounds and was healthy, and the next day the new norms come out, is she suddenly 'not so healthy'? That's right. Her weight is the same, but understanding what is 'average' or desirable has changed. This principle, of updating norms, occurs regularly in psychology, but it is done with less fanfare than something like the height/weight tables that medical doctors use.

There is nothing good or bad about being average. Numbers are just numbers. Average simply refers to where most people are at with a particular attribute. Most people by definition are average. What a person does with some skill, such as 'average intelligence' is not dictated so much by the number but personal choice. For instance, I have seen many individuals who were mentally retarded, with below average intelligence – who were happy, well-adjusted, and successful in the lives they led. I also have seen people with 'rocket scientist' IQ's who have committed suicide, or landed in prison for criminal activity.

Numbers only tell us quantity. They can be employed to a good end, such as suggesting that a student needs extra help catching up in a weak subject area. Or, that someone is intellectually gifted and should be challenged with more advanced work. What they mean to us on a qualitative level is something far more personal, and can only be determined by our own choices.