

Civil Service System Passing Score: Using a score of 70 vs. 70%

Richard Joines, President
Management & Personnel Systems, Inc.
Walnut Creek, CA 94595
(925) 932-0203

Overview: Using an a priori civil service passing score of 70% is not in keeping with sound testing principles. The purpose of this paper is to explain the psychometric issues that are involved, along with relevant testing and regulatory concerns, so that H.R. staff who find themselves in such a situation may explain to Civil Service Commissioners or others why such a rule needs to be changed to set a score of 70 as passing, NOT a score of 70%.

Regulatory/Federal Concerns: The Uniform Guidelines on Employee Selection Procedures (1978) address the issue of establishing passing scores on tests (see General Principles, Section 5H, Cutoff Scores), as follows:

"H. Cutoff Scores. Where cutoff scores are used, they should normally be set as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force. Where applicants are ranked on the basis of properly validated selection procedures and those applicants scoring below a higher cutoff than appropriate in light of such expectations have little or no chance of being selected for employment, the higher cutoff score may be appropriate, but the degree of adverse impact should be considered."

This language focuses on two points. First, it is suggested that passing scores (which are the same as "cutoff scores") should generally be set at the level associated with minimally acceptable performance on the job.

Second, it is stated that practical issues may play a role in setting a passing score that is higher than the level associated with minimally acceptable job performance (e.g., if you have so many candidates relative to the number of positions to be filled that those below the higher cutoff have little or no chance of being hired, you may use the higher cutoff). However, in choosing a higher cutoff, the degree of adverse impact should be considered. By direct implication, the Guidelines are suggesting that the passing point should be lowered if the higher (practical) passing score under consideration significantly increases adverse impact. In other words, it would be unwise to raise the passing score to a level just above a group of minorities or women. Failing these candidates would likely be viewed as a conscious act of discrimination.

The reader should note that the federal Uniform Guidelines on Employee Selection Procedures treat passing scores as an important issue, with the language on adverse impact suggesting that passing scores should be established after the exam has been given. The language in the Uniform Guidelines is contrary to the idea that there is some a priori 70% correct standard that is always the score level associated with minimally acceptable job performance.

As the Regional Psychologist for the Western Region of the U.S. Office of Personnel Management

from 1975-1980, I conducted reviews of the major government agencies under our jurisdiction, including the state governments of California, Arizona, Nevada, and Hawaii. In conducting these reviews, I had internal guidelines to follow. Any agency using a blind 70% pass rule would have been told to change it to conform to the requirements of the Uniform Guidelines. There are additional reasons of a technical nature that make it inappropriate to try to adhere to a rule that says passing is always 70%.

For one thing, many tests are in use that simply cannot be scored on a percentage correct basis, such as oral interviews, biographical inventories, personality instruments, and assessment exercises (in-baskets, report exercises, role-plays, and group discussions, etc). It doesn't make any sense to try to fit these tests into a mold that claims a standard of 70% can be used to determine those who should pass these kinds of tests because these tests cannot be scored in such a manner.

Take the example of a typical interview rating system in which the raters use a 1 - 5 rating scale on the factors being rated. Suppose there are five evaluation factors (e.g., oral communications, interpersonal relations, job knowledge, etc). This would mean there is a total of 25 points on the test as a whole. In keeping with standard professional practice, the scale might look something like this:

1	2	3	4	5
Poorly Qualified	Minimally Qualified	Qualified	Very Qualified	Outstanding

Candidates who are rated "minimally qualified" are at the "2" level on the rating scale, and if this is their score on each of the five factors, they total score is 10 points. This is 40% of the points possible. Consider, however, what happens if we change the scale as shown below:

0	1	2	3	4
Poorly Qualified	Minimally Qualified	Qualified	Very Qualified	Outstanding

Using this rating scale, a total of 20 points is possible (5 factors x 4 = 20). The person who is rated as "minimally qualified" on each of the five factors would have a total score of 5 points; thus, these candidates would score only 25% of the points possible.

Clearly, something is wrong here because in both instances the candidates were rated one point above the lowest point on the scale, but the "percentage correct" changed from 40% on the first scale to 25% on the second scale. You should be asking yourself, "What's the trick?"

The answer is that neither of these scales represents measurement on what is known as a ratio scale. A ratio scale is one that has an "absolute" zero. Absolute zero, in psychometric terms, is the point at which "none" of the quality or property being measured exists.

Everyone in the United States is familiar with the Fahrenheit scale. Most of us have some direct experience with a temperature of zero degrees Fahrenheit. However, it is important to know that

the Fahrenheit scale is not a ratio scale, and zero degrees Fahrenheit doesn't really mean absolute zero. At zero degrees Fahrenheit, there is still warmth. Zero is warmer than 10 degrees below zero. Because the Fahrenheit scale does not have an absolute zero point, it is not a ratio scale, and thus, we cannot say that 30 degrees is twice as warm as 15 degrees. We simply can't interpret ratios in this manner (i.e., $30/15=2$ but this cannot be interpreted to mean twice as much heat). In order to make such statements, your scale would have to be a "ratio" scale that has an absolute zero.

On the Fahrenheit scale, absolute zero isn't reached until you get to -459 degrees. That is the point at which there is no heat. On the Celsius scale, you reach true zero at -273 degrees Celsius. On the Kelvin scale, however, zero is absolute zero and the Kelvin scale is a ratio scale. Please see the attached web site explanation about temperatures and absolute zero. The Fahrenheit and Celsius scales are actually "interval" scales. The concept of measurement on an interval scale will soon be explained.

An example of a common "ratio" scale would be "length." We can take a ruler and measure the length of an object, and if object A is 6 inches long and object B is 3 inches long, we can say that object A is twice as long as object B. This is because there is a point on the scale that we consider to be absolute zero, i.e., the point at which we say the object has no length whatsoever.

In psychological measurement, we know that measurement at the "ratio" level is just not possible. Our tests are not, and probably will never be so precise. When we interview candidates or conduct an assessment center and give them the lowest possible score on factors such as oral communications or interpersonal skills, we are not saying that they have absolutely no oral communications ability, or absolutely no interpersonal skills. We are merely saying that they are so low on our measurement scale that they warrant the lowest possible rating.

So, how precise are psychological tests, including interviews and assessment centers and most all forms of employment tests? Typically, they are considered to be one step below ratio scales and are at the "interval" scale level of precision. The scales given previously, ranging from 1 - 5 and from 0 - 4, represent measurement on an interval scale.

In terms of scientific measurement, the following scales are possible, listed from highest to lowest in terms of degree of precision of measurement:

- Ratio (Kelvin scale: see enclosed explanation of absolute zero; or "length")
- Interval (interviews, ratings on supplemental applications, assessment tests, etc.)
- Ordinal (rank ordering people from tallest to shortest)
- Nominal (categories such as male or female)

On the 1 - 5 interval scale, 5 is bigger than 4 by "one" point; 4 is bigger than 3 by "one" point; 3 is bigger than 2 by "one" point; 2 is bigger than 1 by "one" point; AND the interval of "one" point is the same distance in each of these cases (i.e., the one point interval represents the same amount of increase; thus a candidate rated very qualified is one unit better than the qualified candidate; and the qualified candidate is one unit better than the minimally qualified candidate).

While this may sound simplistic, it is NOT a simple topic. It is treated very seriously in graduate level statistics classes. The level of measurement that we, as scientists, attain is very important to our research and the types of formulas that we use to quantify our results. It is incorrect to say that

someone who scores 4 on an interval scale is twice as qualified as someone who scores 2. Testing practitioners need to understand this point very clearly. It is a fundamental concept that any researcher must understand.

On the 1 - 5 scale, the midpoint is 3 and this point is the equivalent of a 2 on the 0 - 4 scale. The fact that $3/5 = 60\%$ and $2/4 = 50\%$ is irrelevant because it is mathematically incorrect to compute "ratios" on an interval scale. A ratio is formed when you divide one number by another, which you then typically convert to a percentage; and this can only meaningfully be done where you have a ratio scale, such as the Kelvin scale. On the Kelvin scale, you can accurately state that 30 degrees is twice as warm as 15 degrees.

It is important to understand that we NEVER attain ratio scale accuracy of measurement on interviews, ratings of supplemental applications, assessment centers, or similar processes, including written job knowledge or similar tests. The best we can do is measurement at the interval scale level, and this typically works just fine. However, it is important to understand the limitations of our measurement processes.

As a psychologist with OPM, I developed examining systems that were implemented for use in hiring federal employees. When we evaluated candidates for blue-collar jobs, we typically used the job element system and rated candidates on 5 - 6 job factors, each on a 0 - 4 rating scale, with "2" considered passing. We would "transmute" these scores to a 100 point scale with 70 as passing.

The psychologists knew how to devise the correct mathematical formulas for doing this, but instead of having our staffing specialists do this, we provided them a three ring binder with "transmutation" tables. To use these tables, the staffing specialist first had to know how many factors were rated. If five factors were rated, the staffing specialist would turn to the transmutation table for five factors. The staffing specialist would look up the candidate's raw score on the five factors, then record the associated civil service score. The table would look like this:

Number of Factors: 5

Raw Point Total	Transmuted Score
0	40
1	43
2	46
3	49
4	52
5	55
6	58
7	61
8	64
9	67
10	70
11	73
12	76
13	79
14	82

15	85
16	88
17	91
18	94
19	97
20	100

Incidentally, the formula to transmute these scores is: $Y = 3X + 40$ (where Y = Civil Service Score, and X = candidate's raw score).

By the way, we routinely transmute scores on our General Management In-Basket (GMIB) and other assessment exercises to Civil Service Scales, on a 1 - 100 basis, with 70 as passing. Just don't get the impression that this means a 70 = 70% correct. It doesn't. There is no absolute zero on the GMIB; and it is not possible to report meaningful percentage correct scores on the GMIB. And yet, the GMIB received fantastic reviews in Buros' Mental Measurements Yearbook (1995, 12th edition) which attested to the reliability and validity of the GMIB.

Beware of anyone marketing interview or assessment center types of tests who tells you that their tests can be scored in a way that is consistent with having a passing score equal to 70% correct on the test. If so, they must be using "magic" tests, because those tests don't exist in the real world.

Also, if you know someone who has to take the MMPI to determine if he is psychotic, let's hope he doesn't answer in a positive direction on 70% of the items on the schizophrenia scale, because he'll probably NEVER get out of the asylum. Note I don't say score 70% correct because there really are no "right" or "wrong" answers in an absolute sense, it's just how the answer key categorizes answers and assigns points on different scales that form the test. Our in-basket test doesn't measure schizophrenia, but we do measure factors such as leadership and managing conflict and we use formulas to weight information obtained from different elements of the test to compute these scores.

And above all, just remember that you cannot properly compute a percentage correct because measurement is not taking place on a ratio scale on these tests, and in my opinion, never will. It just isn't possible for these kinds of assessment tools because the best we can do is interval scale measurement. When someone figures out how to find "absolute zero" on an evaluation of leadership or interpersonal relations, perhaps we can re-visit this issue, but I don't believe this will ever happen.

What is absolute zero?

(Lansing State Journal, January 29, 1992)

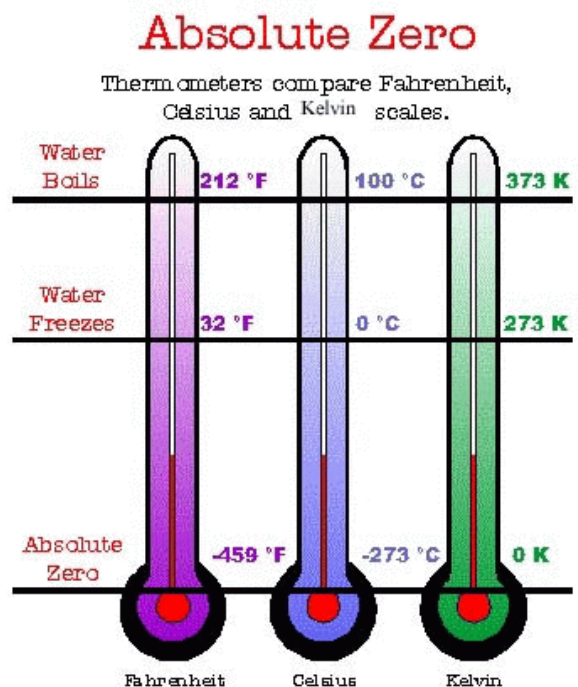
Question submitted by: W. Thomson of Lansing

Temperature is a physical quantity which gives us an idea of how hot or cold an object is. The temperature of an object depends on how fast the atoms and molecules which make up the object can shake, or oscillate. As an object is cooled, the oscillations of its atoms and molecules slow down. For example, as water cools, the slowing oscillations of the molecules allow the water to freeze into ice. In all materials, a point is eventually reached at which all oscillations are the slowest they can possibly be. The temperature which corresponds to this point is called absolute zero. Note that the

oscillations never come to a complete stop, even at absolute zero.

There are three temperature scales. Most people are familiar with either the Fahrenheit or the Celsius scales, with temperatures measured in degrees Fahrenheit ($^{\circ}\text{F}$) or degrees Celsius ($^{\circ}\text{C}$) respectively. On the Fahrenheit scale, water freezes at a temperature of 32°F and boils at 212°F . Absolute zero on this scale is not at 0°F , but rather at -459°F . The Celsius scale sets the freezing point of water at 0°C and the boiling point at 100°C . On the Celsius scale, absolute zero corresponds to a temperature of -273°C .

Scientists - especially those who study what happens to things when they become very, very cold - commonly use the Kelvin scale, with temperatures measured in Kelvin (K). This scale uses the same temperature steps as the Celsius scale, but is shifted downward. On this scale, water freezes at 273 K and boils at 373 K. Only on the Kelvin temperature scale does absolute zero actually fall at 0 K. The accompanying diagram compares the three temperature scales.



http://www.pa.msu.edu/~sciencet/ask_st/012992.html