

## Creating Standardized Scores

Because judges will not all respond in the same way to your paper scoring system, it is important to have a way of compensating for this. The easiest and generally accepted way of doing this is by **“standardizing” the scores for each judge** so that they are comparable across judges.

To do this, however, each judge should review the same number of papers (and at least five papers). The standardization process will ensure that each judge's distribution of scores has the same mean and standard deviation. A common means of standardizing is to use the z-score. With z-scores, each judge will end up with a set of scores so that the mean of that set is 0 and the standard deviation is 1.

(For the purposes of this example, assume that each paper is read by three judges, each judge reads five papers and each paper is evaluated on 10 criteria.)

To compute a z-score you will do the following:

1. **For each judge**, compute **an overall mean and standard deviation** for all the criteria evaluations for all the papers reviewed by that judge. Thus, if each judge read five papers and evaluated the papers on 10 criteria, you would create an overall mean and standard deviation for that judge for all 50 items. (A spreadsheet program will do this for you quickly and efficiently.)
2. Next **subtract the judge’s overall mean** for his/her 50 criteria items from his/her raw score for each criterion item **and then divide that number by the judge’s standard deviation** for all 50 criteria items. This gives you a new z-score for each item:

$$\text{z-score} = \frac{\text{raw score} - \text{mean}}{\text{standard deviation}}$$

3. Next **group the evaluations by paper**. You will have 30 z-scores (3 judges X 10 criteria). **Sum over those 30 z-scores** to create an overall z-score for the paper.\*

Below is an example of how to do this\*\* As is readily apparent, the first two judges had the same ranking order for the papers but they used a different scoring “system.” Their scores, when standardized, are the same and the third judge’s scores receive the same weightings as the first two. Hence now all three judges' scores are comparable. To create the overall measure for each paper you would sum over all three judges' standardized scores.

\*Of course you can create an “average” z-score by dividing by the number of evaluations, 30 in this example. This is desirable in that it gives you a number that is intuitively easier to understand.

\*\*Here it is assumed that each paper is evaluated on only one criterion rather than several as is the more common case. The example applies, but you have to sum over all of the criteria.

### Judge A's Scores

Paper Number	Raw Score	Z-Score	T-Score*
1	2	-1.3	37
2	4	-.6	44
3	6	0	50
4	8	.6	56
5	10	1.3	63
Mean	6	0	50
S.D.	3.2	1	10

### Judge B's Scores

Paper Number	Raw Score	Z-Score	T-Score*
1	6	-1.3	37
2	7	-.6	44
3	8	0	50
4	9	.6	56
5	10	1.3	63
Mean	8	0	50
S.D.	1.6	1	10

### Judge C's Scores

Paper Number	Raw Score	Z-Score	T-Score*
1	5	0	50
2	6	.6	56
3	7	1.3	63
4	3	-1.3	37
5	4	-.6	44
Mean	5	0	50
S.D.	1.6	1	10

### Overall Means

Paper Number	Raw Score	Z-Score	T-Score*
1	4.3	-.8	41
2	5.7	-.2	48
3	7	.4	54
4	6.7	-.1	50
5	8	.7	57

\* A t-score is another type of standardized score. It has a mean of 50 and a standard deviation of 10. It is computed by multiplying the z-score by 10 and then adding 50.