



## Quarterly Journal of Speech

Publication details, including instructions for authors and subscription information:  
<http://www.tandfonline.com/loi/rqjs20>

### Methods used in computing contest scores

Robert West <sup>a</sup>

<sup>a</sup> The University of Iowa

Published online: 05 Jun 2009.

To cite this article: Robert West (1919) Methods used in computing contest scores, Quarterly Journal of Speech, 5:4, 319-333, DOI: [10.1080/00335631909360758](https://doi.org/10.1080/00335631909360758)

To link to this article: <http://dx.doi.org/10.1080/00335631909360758>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

# THE QUARTERLY JOURNAL OF SPEECH EDUCATION

VOLUME V

OCTOBER, 1919

NUMBER 4

## METHODS USED IN COMPUTING CONTEST SCORES

ROBERT WEST  
The University of Iowa

THE last of the orators takes his seat; we listen to the familiar "rubber stamp" from the local principal, "the judges will now prepare their decisions"; and the orchestra bursts into melody. Then what happens?

In most contests the judges are instructed to return their cards to the usher as soon as possible "without leaving their seats." So with the blare of "Manhattan Beach" from the orchestra and the yelling of the delegations from the schools represented and other confusing sounds to confound the concentration of his mind, the poor judge attempts to render an honest verdict. He shields his paper from the prying glances of the inquisitive girls that sit behind him. Just as he has decided upon the first and second place, the usher taps him on the shoulder and enquires if he has completed his ballot. So he hastily blocks out the rest of his verdict and turns it in. In a few minutes some neutral appears on the platform with the final computation made on the basis of the judges' report cards. The decision is announced. The winning schools make the night hideous in their joy, and the rest of the audience say philosophically, "Oh, well, the ways of judges are past finding out."

And it may be so, that they are past finding out, but, at least, "we can approach to a knowledge of the infinite," as my philosophy professor used to say; and that is the purpose of this paper, presumptuous and profane as it may be. I want not only to discover what happens to the decision at the hands of the judge but

what happens even in that holy of holies where the cards are taken to be computed to determine the final grades. A most sacrilegious investigation, to be sure!

I suppose that, when the contest idea was first conceived, it was in connection with the scholastic grades of the pupils in the same school. At the end of the term the percentage grades for each pupil were averaged and the one having the highest average was said to be the best student. Aside from the fact that experts in educational measurements would condemn such methods on the basis of error in the fundamental psychology of the system, yet there is a still more serious objection to be found in the mathematics of the plan. The error back of the averages thus obtained is made more apparent by the step that school people then took in applying the percentage plan to the deciding of speaking and reading contests. They said: we will let each judge give a percentage grade to each contestant, just as the teachers mark the efforts of the pupils. The contestant having the highest average at the hands of the judges shall be the winner. Suppose, for example, to take a simple case, there are three contestants and three judges. The first judge marks A 80%, B 85%, and C 75%. The second judge agrees with the first, except that he exchanges the places given to A and C; he does not disturb the placing of B as the best of the three; he is, however, possessed of rather stricter standards than the first judge, for he gives a best grade of 80% to B, whereas the first judge gave that contestant 85%. The third judge looks at the contest from a slightly different angle; he stresses some qualities in the speakers that the other judges did not consider so important. In the mind of the third judge there is a considerable interval of distribution between A and B and between B and C. He gives A 95%, B 70%, and C 50%.

The summary card of such a contest would look like this:

		Contestants		
Judges	1st	A	B	C
	2nd	80%	85%	75%
	3rd	70%	80%	75%
		95%	70%	50%
Totals		245	235	200
Averages		81 2/3%	78 1/3%	66 2/3%

Thus we see that B, who was credited by the judges with two firsts and a second was defeated by A, who was given by the three judges a first, a second, and a third. Thus a manifest error has been made. Those who are familiar with the inner workings of contests will agree with me that such a distribution of grades by the judges is by no means unusual in its disagreement. It happens again and again, and, whenever the percentage grade system as here described is used, the liability to error is large. The error is by no means subtle. It consists in the inequality of the intervals on the scales used by the several judges for measuring the abilities of the contestants. To make the matter quite plain, suppose that I send out a questionnaire to 50 observers in the state of Iowa and you do the same for each of the states represented by the readers of this journal. Tell them that we want them to report the temperature each morning at 7 o'clock. Then, when the reports are in, we will decide on which state is the warmest by averaging the temperatures reported by the observers in each state. It is obvious that our results would be worthless unless these readings were taken on the same scale, either centigrade or fahrenheit. On one scale there are 180 degrees between boiling and freezing and on the other there are only 100 degrees. The results are not at all comparable. And no more are the scores of two judges, for one may grade on the basis that there is a difference of ten points between the first and the last in order of merit, and the other may use a scale of difference of 50 points; yet both may rank the contestants the same. We do not expect that the standards of the judges shall be the same, but we have a right to expect them to report their grades in comparable figures. If a contest is worth while, it is worth while to report the final results fairly.

The next step in the evolution of the contest was a brilliant one. Those in charge of the contest decided that they would devise a plan that would make the results from the several judges comparable. To do that they inserted a clause in the rules that the first speaker must be given the same percentage grade by each of the judges. Thus, to take the example of the contest cited above, all the judges must grade A the same, say 80%. The card below illustrates exactly what the three judges would do if

they were required to comply with this rule. The distribution of grades preserves practically the same relations between the scores as reported by each judge.

		CONTESTANTS		
Judges	1st	A	B	C
	2nd	80%	85%	75%
	3rd	80%	90%	85%
		80%	55%	35%
Totals		240	230	195
Averages		80%	76 2/3%	65%

Thus again justice is defeated, and this time merely because, although the three scales have a common point of departure, yet their intervals above and below this point are unequal.

Another variation of this same nature is that of requiring that the lowest grade on the judge's report card shall be a certain minimum. For the purpose of illustration let us set down the table of scores of the same judges on the same contest referred to above following the minimum percentage rule, using for the minimum 55%, for example.

		CONTESTANTS		
Judges		A	B	C
	1st	60%	65%	55%
	2nd	55%	65%	60%
	3rd	100%	75%	55%
Totals		215	205	170
Averages		71 2/3%	68 1/3%	56 2/3%

Again the contestant who was considered by two of the judges to be worthy of receiving the first honors and by the other to be worthy of second loses to the contestant who was given one first, one second, and one third.

The following is still another variation with the same fundamental error. Suppose the judges were required to express their opinions about the several contestants in percentages giving the best one a certain maximum, say 90%.

		CONTESTANTS		
		A	B	C
Judges	1st	85%	90%	80%
	2nd	80%	90%	85%
	3rd	90%	65%	45%
Totals		255	245	210
Averages		85%	81 2/3%	70%

Thus this device, as do all the others, fails to remedy this error of unequal intervals in the different scales of percentage used by the several judges on a speaking or reading contest. It makes very little difference whether or not the different scales have a common point. It must be remembered, referring to the thermometer illustration, that even the centigrade and fahrenheit scales have coincident points; but that does not make it possible to report temperatures on either interchangeably without interpolation. To illustrate how the contestants would rank if the grades were interpolated to a common system, let us translate them into terms of the centigrade scale, considering that the difference between the best and the poorest, which ever they are, is the same as the difference between freezing and boiling. The summary would look like this:

		CONTESTANTS		
		A	B	C
Judges	1st	50	100	0
	2nd	0	100	50
	3rd	100	44	0
Totals		150	244	50
Averages		50	81 1/3	16 2/3

Thus it appears that when the scores are reduced to a common unit of expression they tell the true story of the relative abilities of the contestants.

The next step in point of time, though not the next logically, was a decided departure from the percentage scales and obviates their errors. It was the familiar rank method. Each judge is to rank the contestants in the order of merit with the best one as 1. The ranks for each contestant are then added, and the one having the lowest total rank is considered the winner, the one having the next lowest rank is considered second and so on down

the scale. This method smacks of laziness. Although it has the advantage of not possessing the error of the percentage method, it is at best little better than a crude short cut. Yet it seems as though that quality is one of the most important desiderata. We are perfectly willing to let the judges at a swine show take all the time they want and are anxious that the scores be compiled accurately. The reports may be delayed for hours, if necessary, in order to award the honors fairly. But in a contest of humans, we seem to want to get the agony over as soon as possible without much consideration as to accuracy. Hence this rank method.

Let us examine and analyze a typical summary card from judges marking on this basis. Suppose that we have a contest of eight persons, A, B, C, D, E, F, G, and H, judged by five judges. When the contest is over and the judges' ranks are in, what is the process that we really go through with to determine the final rank? If one will examine it carefully, it will be at once evident that it is simply a method of averaging the ranks given to each contestant by the several judges. The one who has the best average rank is declared to be the winner. That is what the process really is, although it does not seem to be such on first inspection, because the last step of the process, that of dividing the totals by the number of judges, has been omitted as being quite unnecessary for the purpose in hand. This, then is our summary card:

		CONTESTANTS							
		A	B	C	D	E	F	G	H
Judges	1st	7	2	4	1	5	6	3	8
	2nd	8	1	2	3	4	5	6	7
	3rd	7	6	5	1	3	2	8	4
	4th	8	6	5	1	3	2	7	4
	5th	7	5	1	2	3	4	6	8
Totals		37	20	17	8	18	19	30	31
Ranks		8th	5th	2nd	1st	3rd	4th	6th	7th

D, then, is declared to be the winner, because he has the best average rank. It makes no difference whether we in judging rank the best one the highest number or the lowest, the winner is to be found by taking the best average rank. We might, for instance, tell the judges to grade the best one 80% and the poorest one 10%.



If we then added the percentages, and gave first place to the one having the highest average percent, we would be doing exactly what we are doing above and the final results would be the same. The error in this method, however, consists in the assumption that in the minds of all of the judges, no matter in what order the contestants are ranked by them, the contestants are separated by equal intervals, that the first is the same amount better than the second in rank as the second is better than the third: that in the estimation of each of the five judges above, for instance, the degree of difference between any two consecutive ranks as reported by the judge is exactly one-seventh of the difference between the best and the poorest. That this is a very greivous fault is evident to any one who has ever had the experience of rendering a "minority report" in a contest on which he was asked to sit with a jury of untrained judges. In such a case, as often happens, the two laymen are in fair agreement that the contest is a very close one; while the expert sees great differences. The two laymen might give X and Y each a first and a second and not quarrel with each other, while the expert might give first to Z and consider that all the rest were grouped together at the bottom of the list as far inferior; and yet his vote on Z as compared with his second choice would have no more weight in the final computations than the quite indifferent preference of either of the laymen in favor of his first choice. In other words, if Cicero, Himself, were to appear on a program with high school orators, the judges could only show their preference for him by placing him one rank higher than the nearest high school lad, and it would appear that Cicero was as much better than the second in rank as the second was better than the third, that much better and no more; and, if, due to the idiosyncracies of some one or more of the judges, the first and second places should be important factors in the final ranking, that superiority of the great Roman over his imitators would have no more weight than as though he were but imperceptibly better. For that reason I say that this rank method seems at best to be but little better than a short cut.

One of the cleverest steps in the evolution of this problem is the next one, a development of the rank method. The two rank methods are based upon quite different assumptions. The first method, the one I have just explained, assumes that the only way



to arrive at a correct decision is to get the work of each contestant viewed by several judges and thus get the value of several different points of view. One man emphasizes one quality and a second looks for another, so that the final report incorporates the points of view of several judges. The second rank method, the one I am about to treat, assumes that concerning any one contestant the majority should rule, that, if he is given third place by three of the judges, seventh by one, and first by another, the "seventh" and the "first" are mistakes; he may not get third, but he should not be given too much advantage from the "first" nor too much discredit on account of the "seventh." Whereas the first rank method assumes that the final report should be conditioned by the opinion of every judge, the second system is built on the assumption that justice can only be done if, in case of a disagreement, the votes of the "off judges" be given as little weight as possible in the final summary. This system assumes, like the first, the equal distribution of abilities among the contestants from first to last, from the best to the poorest.

The method can best be explained by an example. Suppose the contest between the eight persons described above were judged by the second rank method. The first step is simply that of adding the ranks. The contestant having the lowest total of ranks is declared to be the winner. This is as fair as can be, considering that the results are obtained by the rank method. The next step is the clever one. Refer to the contest mentioned. Note that D is unquestionably the winner. His ranks are stricken out and the rest of the contestants are reranked with his place on each judge's list filled by the one standing next in preference. In other words, the attempt is to rank the contestants in accordance with each judge's wishes with the winner left out. The ranks, then, will run from one to seven. Referring to the ranks of the first judge: the 7 becomes a 6, the 2 is promoted to 1, the 4 advances to 3, the 5 changes to 4, and the 6 takes its place, the 3 becomes a 2, and the 8 is still in the last place, but is now 7. Now notice the changes in the ranks given by the second judge. Contestants A, E, F, G, and H all advance one place, but B and C do not. Why? Because no rank has been stricken out below theirs. The system assumes that the rank given to D in the final summary is correct. If, then, it is correct that D is the winner

the second judge must have made a mistake in giving B and C better places than D. Glance down the column of C's ranks. Notice that the fifth judge gave C a better rank than he gave D, another mistake. This last mistake, if the grades are to be computed by the first rank method, is a serious one; for it not only gives C the benefit of a 1 to add into his total, but it compels E to take a 3 instead of a 2. In short, computing by the first rank method, C gets the benefit of two mistakes and thus gets a better final grade than E. The supporters of the second rank system contend that this is unfair, on the grounds that the majority should rule.

The following is the complete reranking of the contestants to determine the second place:

		CONTESTANTS							
		A	B	C	D	E	F	G	H
Judges	1st	6	1	3		4	5	2	7
	2nd	7	1	2		3	4	5	6
	3rd	6	5	4		2	1	7	3
	4th	7	5	4		2	1	6	3
	5th	6	4	1		2	3	7	5
Totals		32	16	14		13	14	27	24
Ranks					1st	2nd			

Reranking for third place:

		CONTESTANTS							
		A	B	C	D	E	F	G	H
Judges	1st	5	1	3			4	2	6
	2nd	6	1	2			3	4	5
	3rd	5	4	3			1	6	2
	4th	6	4	3			1	5	2
	5th	5	3	1			2	6	4
Totals		26	13	12			11	23	19
Ranks					1st	2nd	3rd		

Reranking for fourth place:

		CONTESTANTS							
		A	B	C	D	E	F	G	H
Judges	1st	4	1	3				2	5
	2nd	5	1	2				3	4
	3rd	4	3	2				5	1
	4th	5	3	2				4	1
	5th	4	2	1				5	3
Totals		22	10	10				19	14
Ranks				4th	1st	2nd	3rd		

Here the rule for the settling of the tie must be employed. The rule is quite in harmony with the theory upon which the system is built: in case of a tie let the majority of judges decide the preference. One will notice that C is considered better than B by the third, fourth and fifth judges, while only two of the judges, the first and the second, vote for B in preference to C. One will readily see the advantage of this method of settling a tie as compared with the percentage method as usually employed in settling a tie in ranks.

Reranking for fifth place:

		CONTESTANTS							
		A	B	C	D	E	F	G	H
Judges	1st	3	1					2	4
	2nd	4	1					2	3
	3rd	3	2					4	1
	4th	4	2					3	1
	5th	3	1					4	2
Totals		17	7					15	11
Ranks			5th	4th	1st	2nd	3rd		

Reranking for the last three places:

		CONTESTANTS							
		A	B	C	D	E	F	G	H
Judges	1st	2						1	3
	2nd	3						1	2
	3rd	2						3	1
	4th	3						2	1
	5th	2						3	1
Totals		12						10	8
Ranks		8th	5th	4th	1st	2nd	3rd	7th	6th

The last few places can easily be solved by inspection, though not by simple addition.

The following table illustrates how the two methods may differ in the final places:

	A	B	C	D	E	F	G	H
First rank method	8	5	2	1	3	4	6	7
Second rank method	8	5	4	1	2	3	7	6

The correlation between the two methods, as computed by the formula  $r = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$ , is a positive .905, whereas, if both of

the methods were correct, or at least if they told the same story in the end, the correlation would be 1,000.

The next step in the development of methods of computation is a combination of the rank method and the percentage method. It seems quite strange that this step did not at once follow the efforts to devise a satisfactory percentage method by making the ratings of the several judges comparable. To explain: suppose we require the judge to use both a maximum and a minimum percentage. Let us say to him, "Mark the contestant whom you think the best 100% and the one whom you regard the poorest 10%; and grade the rest accordingly placing each in such a place on the scale as will show his degree of excellence, as compared with the best at 100% and with the poorest at 10%." Naturally at any given contest the poorest differs from the best by an amount that will be the same for each of the judges. They may not all appreciate the difference, but the difference remains nevertheless. This plan, then, assumes that the difference is always 90%. If there are ten contestants, there are to be arranged by each judge nine intervals of difference between succeeding contestants as ranked in order of merit. The differences will not all be the same, hence the intervals on the scale at which the judge will place the several contestants will be unequal. It may be that the second in rank will be marked 99% or 50%. In other words, each judge's card will show not only his order of preference but also the degree by which he prefers one contestant to another from first to last, and these different intervals of the scale will be expressed in units that can be compared with the units of measurement employed by the other judges on the same contest. The final places can thus be determined by a simple average.

But this system has its difficulties. In the first place, the system of marking in the schools has so trained us that we hesitate to grade one so high as 100% or so low as 10%. In fact, we dislike very much to grade a contestant below "passing," which is about 70% or 75%. Then, too, we are very careful about grades in the 80s or 90s, but very careless about grades below 70%, yet these lower grades may well be important factors in determining even the first place on the contest. Trained as we are in handling of percentages we cannot conceive of the scale as extending above 100% or below 0%, so it seems absurd to us, when we are judging

a close contest, to call one perfect and one only one-tenth as good. If the judges could rid themselves of all associations connected with the percentage scale and think of the points at which absolute perfection and absolute worthlessness are located as being entirely dependent upon the contest in question, then the system would be quite the best thing yet devised. You must tell the judge that the best one on the contest may be 20 degrees below perfection or he may be 99 degrees below it or 7,000 degrees below, and that the poorest may be but five degrees better than the point of absolute worthlessness, or he may be 205 degrees better; but that we are centering our attention on just that portion of the scale that would include the best and the poorest, and for the purposes of convenience we are numbering those points from 10 to 100.

To obviate all this explanation, however, a much simpler method has been devised. We have already seen that in using the rank method if we call the best one of ten contestants 10 and the poorest 1 and then give the best final rank to the one having the highest total, we have done precisely the same thing as though we had given the best one 1, as we actually do, and the poorest one 10, and then had given the best final rank to the one having the lowest total. The grades throughout the contest would be the same with these two methods. Let us, then, apply that principle to the problem in hand. Let us suppose that we have in mind an infinite number of contestants arranged in order of preference from the perfect to the worthless. Somewhere on the scale the best one in this contest belongs and somewhere there is a place where the poorest fits into the sequence. Let us then think of 48 places equally distributed between the best one on this contest and the poorest. Let us call the best one first and the poorest one 50th. Then let the rest be placed each in one of the 48 remaining places, in such a way, separated by such distances, as will properly express the relative degrees of difference obtaining among the several contestants. The final ranks can be determined in the usual manner employed in the first rank system, by simply adding the ranks and without reranking giving out the final ranks, the best to the lowest total, and the second to the next lowest total, etc., on to the poorest. In case of a tie, which, by the way, will seldom happen, decide the issue by consulting the records of the two persons tied to determine the

preference of the majority, as suggested above in the description of the second rank method. This is the theory of the plan. Let it be said, however, that all this explanation need not be given to the judge. The instructions to him are simple, indeed. The following ballot is presented as the most satisfactory method of deciding the final result of a contest yet worked out:

JUDGE'S BALLOT

PART I

The judge should carefully fill out Part I, before he considers Part II. The judge should indicate on the first blank line of this form the contestant whom he considers the best; on the second blank line the contestant whom he considers the second best; and so on down the page to the contestant whom he considers the poorest. In other words, the names of the contestants should be arranged in order of merit with the best at the top and the poorest at the bottom and the rest in their proper ranks between. No two should be ranked the same.

Number in order of appearance	Name	Selection
The Best		
Second		
Third		
Fourth		
Fifth		
Sixth		
Seventh		
Eighth		
Ninth		
The Poorest		



PART II

After the judge has carefully ranked the contestants in the order of their relative merit, he should next indicate on the scale below exactly *how much* better or worse each contestant is than the others. Do it in this manner:

Write the word "second" in the space that is just the right distance from the upper end of the scale, where your first choice is located, and also from the lower end, where your last choice is located, to show the excellence of your second choice as compared with your first and last.

Then do likewise with your third choice: give it a place on the scale by writing "third" in the space that will show your third choice in its proper relation not only in order of merit but also in comparative excellence to the three contestants whose places are already fixed.

Then place the fourth, fifth, etc., until all are so located on the scale that their positions will indicate the comparative excellence of each.

Do nothing with the best and poorest in your rank order of preference: your voting them as your first and last choices automatically fixes them at either end of the scale. Place the rest carefully.

Make as many changes in your work as necessary. Ask for a second blank, if you need one. Accuracy is more to be desired than speed.

Rank	Scale	Rank	Scale
The Best	1		16
	2		17
	3		18
	4		19
	5		20
	6		21
	7		22
	8		23
	9		24
	10		25
	11		26
	12		27
	13		28
	14		29
	15		30

Rank	Scale		Rank	Scale	
	31			41	
	32			42	
	33			43	
	34			44	
	35			45	
	36			46	
	37			47	
	38			48	
	39			49	
	40		Poorest	50	

When the blanks are turned in to the officials in charge of the contest, they need only to prepare a table with the names of the several contestants at the top. Take up the papers of the first judge. Note to whom he gave the first place and credit that person with 1 ; then note to whom he gave second and credit that contestant with the score upon the scale opposite which the word "second" appears, etc., until all the contestants have been credited with their proper scores. Then proceed with the papers of the other judges taking one report at a time. Add the scores and rank the contestants as suggested above.

Thus we have arrived at a method of computing the scores that (1st) enables the judge to check up his ratings of the several contestants by judging with his eye the intervals that separate them; that (2nd) allows the judge not only to rank the contestants, but to rate them; and that (3rd) shows finally to the contestant not only how he stands in relation to the others in the contest, but how much better or worse he is than they.