

Online Supplement for

Some Comparisons among Quadratic, Spherical, and Logarithmic Scoring Rules

J. Eric Bickel
 Department of Industrial and Systems Engineering,
 Texas A&M University, College Station, Texas 77843-3131
 Tel: 979.845.4347, Fax: 979.847.9005
 ebickel@tamu.edu

This online supplement provides additional detail regarding the rank order simulation results and the Stanford testing results.

3.1 Rank Correlations

Simulation Results

Table A adds to the information presented in Table 2 of the paper. First, simulation results for the case of 50 assessors are included, as are results for odd numbers of statements. Second, the standard deviations of the samples are provided.

The standard deviations of these estimates are relatively small in all cases, but increase with the number of statements and decrease with the number of assessors. Because the rank correlations are bounded above by 1, care must be taken in using the standard deviation results to estimate a range in which the rank correlations may fall.

Table A: Simulated Rank Correlations (500 Simulations)

Mean Statements	Quadratic vs Logarithmic Number of Assessors (N) =				Spherical vs Logarithmic Number of Assessors (N) =				Spherical vs Quadratic Number of Assessors (N) =				
	n	10	50	100	200	10	50	100	200	10	50	100	200
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.975	0.989	0.991	0.992	0.970	0.987	0.989	0.990	0.994	0.998	0.999	0.999	0.999
4	0.963	0.980	0.982	0.983	0.959	0.977	0.979	0.980	0.987	0.995	0.996	0.997	0.997
5	0.955	0.970	0.973	0.975	0.945	0.968	0.972	0.973	0.980	0.991	0.993	0.994	0.994
6	0.940	0.964	0.966	0.968	0.938	0.961	0.966	0.967	0.975	0.988	0.989	0.991	0.991
7	0.928	0.959	0.960	0.962	0.926	0.958	0.961	0.963	0.970	0.984	0.987	0.987	0.987
8	0.927	0.952	0.954	0.956	0.926	0.954	0.958	0.959	0.962	0.981	0.984	0.984	0.984

Std Dev Statements	Quadratic vs Logarithmic Number of Assessors (N) =				Spherical vs Logarithmic Number of Assessors (N) =				Spherical vs Quadratic Number of Assessors (N) =				
	n	10	50	100	200	10	50	100	200	10	50	100	200
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.031	0.004	0.003	0.002	0.029	0.005	0.003	0.002	0.009	0.001	0.000	0.000	0.000
4	0.041	0.010	0.006	0.004	0.039	0.008	0.005	0.003	0.020	0.003	0.001	0.001	0.001
5	0.043	0.012	0.008	0.006	0.048	0.011	0.006	0.004	0.026	0.005	0.003	0.002	0.002
6	0.063	0.015	0.010	0.007	0.052	0.013	0.008	0.005	0.028	0.007	0.004	0.003	0.003
7	0.069	0.019	0.012	0.008	0.063	0.014	0.009	0.006	0.037	0.008	0.005	0.004	0.004
8	0.077	0.020	0.014	0.010	0.059	0.015	0.010	0.006	0.045	0.010	0.006	0.005	0.005

Stanford Testing Detail

The results of five years of Stanford testing involving 1,030 students are presented in Table B. The number of students tested ranged from 166 in Year 2 to 250 in Year 5. The Min and Max columns are the minimum and maximum rank correlations obtained on any of the 15 questions for the listed scoring rule comparison. The Average column is the average of the 15 individual question rank correlations. The Exam Total is the rank correlation among students based on their overall exam scores.

The rank correlations are all quite high. The performance of Q-L and S-L are comparable on a by-question basis, with each obtaining the highest rank correlations in about half the cases. However, Q-L produced higher exam total rank correlations than S-L in every year. S-Q produced the highest rank correlations in all but one case (Year 2 – Min).

Table B: Stanford Testing Rank Correlations ($n = 4$)

Year	N	Quadratic vs Logarithmic				Spherical vs Logarithmic				Spherical vs Quadratic			
		By Question			Exam	By Question			Exam	By Question			Exam
		Min	Max	Average	Total	Min	Max	Average	Total	Min	Max	Average	Total
1	167	0.978	0.999	0.989	0.991	0.974	0.998	0.988	0.985	0.995	1.000	0.998	0.996
2	166	0.906	0.997	0.983	0.971	0.982	0.997	0.991	0.960	0.948	1.000	0.993	0.993
3	244	0.986	0.999	0.995	0.978	0.979	0.999	0.993	0.965	0.994	1.000	0.998	0.994
4	203	0.984	0.999	0.994	0.972	0.987	0.999	0.995	0.957	0.990	1.000	0.997	0.995
5	250	0.987	0.999	0.994	0.985	0.986	0.998	0.993	0.974	0.996	0.999	0.998	0.993
	Avg	0.968	0.999	0.991	0.979	0.982	0.998	0.992	0.968	0.984	1.000	0.997	0.994

3.2 Rank Differences

Simulation Results

Table C adds to the information presented in Table 3 of the paper. First, simulation results for the case of 50 assessors are included, as are results for odd numbers of statements. Second, the standard deviations of the samples are provided.

In this case, the standard deviations are large relative to the average values. Increasing the number of statements or decreasing the number of assessors increases the standard deviation of ranking differences.

Table C: Simulated Rank Differences (500 Simulations)

Loss (10th Percentile)												
Mean Statements <i>n</i>	Quadratic vs Logarithmic Number of Assessors =				Spherical vs Logarithmic Number of Assessors =				Spherical vs Quadratic Number of Assessors =			
	10	50	100	200	10	50	100	200	10	50	100	200
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	-9.8%	-5.3%	-4.8%	-4.6%	-10.4%	-6.0%	-5.5%	-5.3%	-4.3%	-2.1%	-1.7%	-1.5%
4	-12.4%	-7.5%	-7.1%	-6.9%	-13.9%	-8.4%	-7.8%	-7.5%	-6.5%	-3.4%	-2.9%	-2.6%
5	-13.9%	-9.4%	-8.7%	-8.3%	-14.5%	-9.8%	-9.2%	-8.8%	-8.5%	-4.5%	-4.0%	-3.6%
6	-14.9%	-10.1%	-9.8%	-9.4%	-16.2%	-10.7%	-10.1%	-9.8%	-9.7%	-5.5%	-4.9%	-4.4%
7	-16.9%	-11.3%	-10.6%	-10.4%	-16.2%	-11.6%	-10.7%	-10.6%	-11.7%	-6.4%	-5.4%	-5.1%
8	-17.2%	-11.9%	-11.3%	-11.1%	-17.5%	-11.8%	-11.4%	-11.1%	-12.8%	-6.8%	-6.1%	-5.7%

Std Dev Statements <i>n</i>	Quadratic vs Logarithmic Number of Assessors =				Spherical vs Logarithmic Number of Assessors =				Spherical vs Quadratic Number of Assessors =			
	10	50	100	200	10	50	100	200	10	50	100	200
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	7.3%	1.5%	1.0%	0.6%	7.2%	1.5%	0.9%	0.6%	5.6%	0.6%	0.5%	0.2%
4	7.4%	2.2%	1.4%	0.9%	7.8%	2.0%	1.2%	0.8%	6.5%	1.2%	0.7%	0.4%
5	7.7%	2.5%	1.6%	1.2%	7.9%	2.2%	1.4%	0.9%	7.3%	1.5%	0.9%	0.6%
6	8.9%	2.8%	1.9%	1.3%	8.6%	2.4%	1.6%	1.1%	7.6%	1.8%	1.2%	0.7%
7	8.8%	3.1%	1.9%	1.4%	8.2%	2.5%	1.6%	1.1%	8.1%	2.2%	1.2%	0.9%
8	9.4%	3.1%	2.2%	1.5%	9.0%	2.6%	1.8%	1.2%	8.4%	2.4%	1.4%	1.1%

Gain (90th Percentile)												
Mean Statements <i>n</i>	Quadratic vs Logarithmic Number of Assessors =				Spherical vs Logarithmic Number of Assessors =				Spherical vs Quadratic Number of Assessors =			
	10	50	100	200	10	50	100	200	10	50	100	200
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	3.8%	4.7%	4.6%	4.5%	4.2%	5.5%	5.4%	5.3%	0.7%	1.9%	1.6%	1.5%
4	5.6%	6.8%	6.7%	6.7%	6.8%	7.7%	7.7%	7.7%	1.9%	2.9%	2.8%	2.6%
5	6.9%	8.3%	8.2%	8.1%	7.6%	9.1%	9.1%	9.1%	2.8%	3.9%	3.8%	3.6%
6	7.7%	9.1%	9.4%	9.2%	9.3%	10.0%	10.1%	10.1%	3.7%	4.8%	4.6%	4.4%
7	8.8%	10.2%	10.1%	10.1%	9.5%	10.7%	10.6%	10.8%	4.4%	5.7%	5.2%	5.1%
8	9.0%	10.6%	10.8%	10.8%	10.4%	11.2%	11.2%	11.3%	5.5%	6.0%	5.8%	5.7%

Std Dev Statements <i>n</i>	Quadratic vs Logarithmic Number of Assessors =				Spherical vs Logarithmic Number of Assessors =				Spherical vs Quadratic Number of Assessors =			
	10	50	100	200	10	50	100	200	10	50	100	200
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	5.4%	1.5%	1.0%	0.7%	5.3%	1.4%	0.9%	0.6%	2.6%	0.5%	0.5%	0.2%
4	6.1%	2.0%	1.4%	1.0%	6.0%	1.8%	1.2%	0.8%	4.0%	1.1%	0.6%	0.4%
5	6.6%	2.5%	1.6%	1.2%	6.0%	2.1%	1.3%	0.9%	4.7%	1.4%	0.8%	0.6%
6	7.1%	2.6%	1.9%	1.3%	7.0%	2.2%	1.5%	0.9%	5.2%	1.7%	1.1%	0.7%
7	6.8%	3.0%	1.9%	1.4%	6.3%	2.4%	1.5%	1.0%	5.6%	1.8%	1.2%	0.9%
8	7.5%	3.1%	2.2%	1.6%	7.2%	2.3%	1.5%	1.0%	6.2%	1.9%	1.4%	1.0%

Stanford Testing Detail

Table D summarizes the ranking differences for our Stanford dataset. The summary is divided into two sections: 10th percentile and 90th percentile results. The 10th percentile results represent ranking losses that were so extreme that only 10% of students would have experienced a loss of that magnitude or greater (more negative). The 90th percentile results represent gains in rank such that only 10% of the students would have gained more than this from a change in the scoring rule. The Min and Max columns are the minimum and maximum rank differences obtained on any of the 15 questions for the listed scoring

rule comparison. The Average column is the average of the 15 individual question rank differences. The Exam Total is the rank difference among students based on their overall exam scores.

As can be seen, changes in the scoring rule could result in significant ranking differences on an individual question basis. For example, in Year 1, 10% of the students would have lost an average of at least 5.9% in rank (approximately 10 out of 167 spots) on each of the 15 questions by being scored with Q instead of L. In Year 2, 10% of the students would have lost at least 21% in rank (35 spots) on one particular question by being scored with Q instead of L. These changes in rank by question would have translated into large ranking differences for the exam overall. For example, in Year 3, 10% of the students would have lost at least 9.4% in rank for the total exam by being ranked with S instead of L. Gains on individual questions are lower, but the overall increases in rank can be significant. For example, in Year 3, 10% of the students would have gained at least 9.8% in rank by being scored with S instead of L. Overall, S-L seems to produce the largest changes in rank, while S-Q yields the smallest.

Table D: Stanford Testing Rank Differences ($n = 4$)

		Loss (10th Percentile)											
Year	N	Quadratic vs Logarithmic				Spherical vs Logarithmic				Spherical vs Quadratic			
		By Question			Exam	By Question			Exam	By Question			Exam
		Min	Max	Average	Total	Min	Max	Average	Total	Min	Max	Average	Total
1	167	-12.0%	-0.6%	-5.9%	-4.8%	-15.6%	-0.6%	-6.8%	-5.4%	-1.8%	0.0%	-0.9%	-3.6%
2	166	-21.1%	-0.6%	-6.8%	-6.0%	-10.8%	-0.6%	-5.9%	-8.4%	-9.6%	0.0%	-1.8%	-4.2%
3	244	-7.8%	0.0%	-3.7%	-7.0%	-11.1%	0.0%	-4.4%	-9.4%	-2.9%	0.0%	-1.5%	-4.1%
4	203	-7.9%	-1.5%	-4.7%	-6.4%	-6.9%	-1.5%	-4.3%	-7.4%	-3.4%	-0.5%	-1.4%	-3.4%
5	250	-9.2%	-1.2%	-4.5%	-4.8%	-9.6%	-1.6%	-4.8%	-7.2%	-2.4%	-0.4%	-1.1%	-4.4%
	Avg	-11.6%	-0.8%	-5.1%	-5.8%	-10.8%	-0.9%	-5.3%	-7.6%	-4.0%	-0.2%	-1.3%	-4.0%

		Gain (90th Percentile)											
Year	N	Quadratic vs Logarithmic				Spherical vs Logarithmic				Spherical vs Quadratic			
		By Question			Exam	By Question			Exam	By Question			Exam
		Min	Max	Average	Total	Min	Max	Average	Total	Min	Max	Average	Total
1	167	0.0%	1.2%	0.4%	4.8%	0.0%	3.0%	1.0%	6.6%	0.0%	1.8%	0.6%	3.6%
2	166	0.0%	9.6%	1.3%	5.4%	0.0%	1.8%	0.8%	7.8%	0.0%	11.4%	1.3%	3.6%
3	244	0.0%	4.5%	1.1%	6.6%	0.0%	2.9%	1.7%	9.8%	0.0%	3.3%	1.4%	3.7%
4	203	0.0%	4.9%	1.3%	4.9%	0.5%	3.0%	1.3%	7.4%	0.0%	3.9%	1.3%	3.4%
5	250	0.4%	3.6%	1.0%	4.4%	0.4%	2.4%	1.3%	8.4%	0.0%	2.4%	0.7%	4.0%
	Avg	0.1%	4.8%	1.0%	5.2%	0.2%	2.6%	1.2%	8.0%	0.0%	4.6%	1.1%	3.7%

3.3 Score Equality

Stanford Testing Detail

Table E displays the maximum response difference for the Stanford dataset. If Q or S scoring had been used instead of L in Year 1 the average maximum response difference to achieve the same or higher score would have been 0.08 (see columns 3 and 8). These differences varied by question, and the maximum difference on one particular question in Year 1 was 0.13 under Q scoring and 0.10 under S

scoring. Spherical scoring tended to produce the largest and most extreme differences. In addition to their magnitude, the maximum response differences tended to involve a large number of students. For example, the average number of students involved in an incident comprising the maximum response difference in Year 1 was 5.2% for Q scoring and 7.5% for S scoring. However, on one question, 28.4% of the class would have been involved in such an incident in Year 5 under Q scoring and 30.7% in Year 2 under S scoring.

Table E: Stanford Maximum Response Differences and Score Equality Results ($n = 4$)

Year	N	Quadratic vs Logarithmic					Spherical vs Logarithmic				
		Response Difference		Students Involved per Question			Response Difference		Students Involved per Question		
		Average	Max	Min	Max	Average	Average	Max	Min	Max	Average
1	167	0.08	0.13	0.6%	21.6%	5.2%	0.08	0.10	0.6%	21.6%	7.5%
2	166	0.09	0.14	1.2%	16.9%	6.2%	0.07	0.10	1.2%	30.7%	9.8%
3	244	0.04	0.08	0.4%	8.6%	3.4%	0.11	0.21	0.4%	19.3%	6.7%
4	203	0.04	0.08	1.0%	20.2%	6.1%	0.11	0.20	1.5%	21.2%	7.5%
5	250	0.04	0.08	1.2%	28.4%	6.9%	0.13	0.21	0.4%	18.4%	9.6%
	Avg	0.06	0.10	0.9%	19.1%	5.6%	0.10	0.16	0.8%	22.2%	8.2%