

BCS Or Just BS: How College Football Could Crown The Wrong National Champion? Just Do The Math—Correctly!

C. E. Wynn Teasley, University of West Florida, USA
Martin Hornyak, University of West Florida, USA

ABSTRACT

The 2009 college football season is here, but there has been a continuing controversy swirling over how the Football Bowl Subdivision (FBS) selects its national champion. College football uses a multi-criterion decision matrix (MCDM) evaluation technique to determine which two teams will play for the national championship. We analyzed the BCS method of evaluating which teams play for the national. We found that Texas should have been ranked ahead of Oklahoma. We conclude that BCS utilizes a mathematically flawed evaluation method to decide which two teams will play each year in the national championship game. Just do the math--correctly!

Keywords: college football, Bowl Championship Series, decision matrix, multi-criteria decision matrix

INTRODUCTION

There has been a continuing controversy over how the Football Bowl Subdivision (FBS) or Division I in college football selects its national champion. It evaluates the best college teams using a combination of polls and scores to determine which two teams should be ranked at the top of the Bowl Championship Series (BCS) after all the regular season games have been completed. Based on 2008 (which is also true for this year as well) that evaluation employed two “human” polls. One poll was comprised of current Division I football coaches and the other poll was comprised of a combination of former coaches, players, and other sports figures. In addition to the human polls, a composite of computer rankings was compiled from six different “unbiased” scoring methods, after the highest and lowest computer ranks were eliminated. Thus, four different computer scores are utilized for each team. The basis of how the computer polls determined their scores is privileged information and they yield very divergent results.

The current BCS formula employs a multi-criteria decision matrix (MCDM) evaluation method, which we accept as one appropriate method for doing so. College football is a very big business. An error in this evaluation method might well cost the teams of a slighted conference millions of dollars (estimated at \$17.5 million) as well as other residual costs of not having the national champion team from its conference.

While many legitimate criticisms of the BCS have been raised to support a playoff in Division I college football, we accept the measures that the BCS uses for our analysis. We re-analyze, however, the method of evaluating those teams. What we show is that the method of evaluation by the BCS produces an erroneous mathematical result. Specifically, the BCS method of combining the computer scores by ranking the teams inversely and calculating the computer “poll” score as a percentage or rate of the possible total ranking produces the error. Converting scores from a higher level of measurement (interval) to a lower level (ordinal) as a common denominator for their summation changes the relationship (i.e., the difference) between the scores and may well end up selecting the wrong team for the national championship game. That is precisely what we found here.

HOW THE BCS WORKS

The Bowl Championship Series has changed through the years. Separate components, such as strength of schedule and number of losses have been dropped. The BCS is currently comprised of only three components mentioned above. The two human polls include the Harris poll, which replaced the Associated Press (AP) poll in the BCS when writers became dissatisfied with the BCS results, and the Coaches poll. There are 113 voting members in the Harris poll and 61 current head coaches vote in the Coaches poll. The scores in each are developed as a percentage or rate of the maximum possible poll score, with the number one team receiving 25 points and the remaining teams receiving a chronologically lesser number of points for the first twenty-five teams. In the Harris poll the maximum score would be 2825 ($25 * 113$) and in the Coaches poll the maximum score would be 1525 ($25 * 61$). If a team was ranked number one on all the ballots, it would receive a human poll score of 1.0 ($2825/2825$ or $1525/1525$) on both human polls, and they would receive a lower rate (e.g., .9875) for some combination of lower rankings on some of the ballots.

The computer rankings are actually converted in a similar fashion. There are six computer polls that yield scores which range from around 325 to small fractions, like .975. The BCS tabulates these scores similar to the way it tabulates the human polls with an inverted ranking—25 to 1—given to the top twenty-five teams—first to twenty-fifth respectively. It then drops both the highest and lowest scores from matrix, using only the four middle scores for each team. This is a technique often used with subjective judges in sports scoring, like the Olympics, where human error or bias might be mitigated by that technique. The computers, however, are not as susceptible to human error and are often called “unbiased” for that reason even though their programmers do inject biases from the criteria and formulas they choose to employ. The final computer score is the sum of the four inverted rankings for each team divided by one hundred since that would be the maximum score a team could achieve on the four rankings.

The third component of the BCS, the computer rankings or polls, is therefore based on only four scores for each team. This relatively small number of “ballots” contributes to the amount of error that is found in the computer poll scores that are subsequently tabulated.

RESEARCH ON THE BOWL CHAMPIONSHIP SERIES

The Bowl Championship Series (BCS) has drawn public criticism and complaints, especially from those favoring a playoff, almost from its inception. Because of college football’s status in the country’s popular culture, there has been much research directed at the BCS. Many scholars have questioned the statistical accuracy of the BCS and its computer scores and it is likely that much public criticism is based on the use of statistics and data which are proprietary and seems mysterious.

Stern (2004) has suggested a linear modeling approach for determining a football champion and he has recommended a “quantitative boycott” (2006) since the BCS has limited the data (such as “running up the score”) as a datum input in the computer models. A linear modeling approach has been supported by other analysts as well. (West and Lamsal, 2008)

Other statistical analysts have recommended approaches such as “random walkers” (Callaghan, Mucha, and Porter, 2008), “quadratic assignment” (Cassady, Maillart, and Salman, 2005), or reducing “retrodictive errors” (Coleman, 2005). Soares, et. al., (2007) employed a “genetic algorithm” to criticize the BCS results. Finally, Hopkins (1997) observed that “I don’t believe statistics has much to say about what kind of rating is most appropriate. The question of ‘What does Good mean?’ is not a question of statistical assessment.” While *good* might not be statistically assessed, predictive accuracy can be.

Since the BCS is now comprised of a combination of human polls and computer score rankings, some research has highlighted the subjective or perceptive biases of human rankings based on TV exposure (Campbell, Rogers, and Finney, 2007). On the other hand, one researcher has asked “do the computers know best?” Martinich (2002) found that the human polls were generally as accurate predictors of game outcomes as, so-called unbiased, computer scores were.

Some researchers have contested the legal or ethical nature of the BCS. Eckard (1998) notes that many economists regard the NCAA as a cartel that enforces restrictions on recruiting, eligibility and compensation that forces players to work for less than their talents would likely derive for their host institutions in an open market place. Being more direct, Schmidt (2007) claims that the BCS is a violation of the Sherman Anti-trust Act which basically serves to insure that “the rich get richer.”

The BCS has been an active topic for researchers from both a statistical and a legal standpoint. They have contended that other, often more statistically complex, models would be better predictors of which teams should play for the fabled national championship, while legalistic reviews have contended that the BCS maintains the hegemony of a few programs that perennially compete for the championship.

MULTI-CRITERIA DECISION MODELS (MCDMS) OR DECISION MATRICES

Multi-criteria decision models (MCDMs) have been used many times in the attempt to capture the impact of several criteria on a decision outcome (Nagel, 1984; Nagel, 1985; and Nagel, 1987) The purpose of this paper is to demonstrate how the misuse of measurement for developing an MCDM index can result in an inaccurate answer, which can have serious impacts and implications for those involved (Teasley, 1989). This paper will focus on the BCS as just one example of how this problem occurs in everyday life. MCDMs are frequently employed to make decisions where the use of one simple criterion is insufficient. The use of two or more criteria, however, typically requires the index maker to generate a common denominator in order for all the scores or measures to be combined accurately into one summative score. They are often tempted, as with the BCS, to reduce scores to ranks and then to tally the ranks. It is this process of converting interval level scores to a lower level of measurement (ordinal level) that distorts the relationships between the scores and can yield a false result.

THE 2008 BOWL CHAMPIONSHIP SERIES (BCS) FINAL RESULTS

At the end of the season, the BCS calculated the final results for the Division 1 college football season. Those results are reported in Table 1 below. That table shows that two teams—Oklahoma and Florida—are ranked first and second in the country and, as a result, they played each other in the national championship game on January 8, 2009 in Miami’s Orange Bowl. There are some subtle elements that might confound such an obvious result when one scrutinizes the BCS more closely.

Table 1: The BCS Pre-Bowl Game Standings in 2008

RK	Team	W-L	Harris			Coaches			Computer	BCS	
			RK	Points	%	RK	Points	%	Points	Prev	Avg
1	Oklahoma	12-1	2	2699	.9554	1	1482	.9718	1.0000	2	.976
2	Florida	12-1	1	2776	.9827	2	1481	.9711	.8900	4	.948
3	Texas	11-1	3	2616	.9260	3	1408	.9233	.9400	3	.930
4	Alabama	12-1	4	2442	.8644	4	1309	.8584	.8100	1	.844
5	USC	11-1	5	2413	.8542	4	1309	.8584	.7500	5	.821
6	Utah	12-0	7	2119	.7501	7	1134	.7436	.8600	6	.785
7	Texas Tech	11-1	8	2090	.7398	8	1132	.7423	.8700	7	.784
8	Penn State	11-1	6	2186	.7738	6	1193	.7823	.6600	8	.739
9	Boise State	12-0	9	1938	.6860	9	1034	.6780	.7300	9	.698
10	Ohio State	10-2	10	1858	.6577	10	1004	.6584	.5900	10	.635

The Harris Poll and the Coaches Poll scores result from the number of inverted ranking scores each team receives from the 113 and 61 voters respectively. It is interesting to note that Florida ranks ahead of Oklahoma in the Harris poll, but it ranks behind Texas in the computer rankings. Alabama had ranked number one going into the SEC playoffs, when it suffered its first loss to make all the top five teams have one loss each. But, since it lost last,

it dropped to the bottom of the top four teams. Also, Florida, which ranked fourth, jumped to second ahead of Texas because, in large part, Florida and Oklahoma had the good fortune to play—and win—an additional game in their respective conference championships. Therefore, these final standings are also a good example of “recency” error, especially when combined with the fact that Texas was the last of the top three teams to lose a game and Alabama was the last to lose of the top four teams.

REVISING THE BCS OUTCOME BASED ON USING PERCENTAGES OR RATES

As stated previously, the human polls and the computer rankings are comprised of inverted rankings, but the human polls include between 61 and 113 ballots. Thus, those composite rankings offer a more linearly and normally distributed result. The computer rankings, on the other hand, start with raw scores and convert those scores to ranks, which changes the relationships between the scores. Table 2 demonstrates what happens when ranks are utilized instead of percents or rates. This table includes the both the points and scores from the Harris Poll and the Coaches Poll in columns one through four respectively. The Points are the combined inverted rank scores for each of the top six teams and the Scores are the result of the Points divided by the maximum possible number of points. The fifth column is the Ranked Computer Scores as currently reported in the BCS and column six includes the Percent Computer Scores where all scores are recorded as a percent of the highest score. Column seven reports the BCS Score as currently reported based on the average of the Harris Poll Scores, the Coaches Poll Scores, and the Ranked Computer Scores. The last column shows the Revised BCS Score derived from the summation of the Harris Poll Scores, the Coaches Poll Scores and the Percent Computer Scores.

Table 2: Revised BCS Scores for the Top Six Teams with Percent Computer Scores

	1	2	3	4	5	6	7	8
Team	Harris Points	Poll Scores	Coaches Points	Poll Scores	Ranked Comp.	Percent Comp.	BCS Score	Revised BCS Score
Oklahoma	2699	0.9554	1482	0.9718	1.000	1.000	0.976	0.9757
Florida	2776	0.9827	1481	0.9711	0.890	0.975	0.948	0.9764
Texas	2616	0.9260	1408	0.9233	0.940	0.980	0.930	0.9432
Alabama	2442	0.8644	1309	0.8584	0.810	0.918	0.844	0.8804
USC	2413	0.8542	1309	0.8584	0.750	0.917	0.821	0.8765
Utah	2119	0.7501	1134	0.7436	0.860	0.941	0.785	0.8115

The most obvious finding from this revised BCS score is that Florida would actually rank first in Division I college football after the regular season ended. This occurred from the fact that the ranking scores did not reflect how close Florida was to Oklahoma and Texas with regard to the actual computer scores. Interestingly, Florida was favored by odds-makers, who have money at stake, to beat Oklahoma, which was ranked first in the BCS. This may add credence that the Revised BCS Score is more accurate or valid. Similarly, in the SEC championship game, Alabama was ranked first in the BCS with Florida ranked fourth. Then, the odds-makers predicted a ten point victory for Florida and it won by eleven. This raises the question of whether the polls should be based on the whole season’s accomplishment or on which team is playing best at season’s end.

What difference does converting scores to percents or rates as a common denominator matter? When the computer scores are converted that way the correlation between the raw scores and the converted scores is perfect ($r^2 = 1.0$). When the computer scores were converted to ranks, the correlation is not perfect ($r^2 = .91$), with about 9 percent difference or error in the result. The correlation between the 2008 BCS Score and the Revised BCS Score is, therefore, also less than perfect ($r^2 = .95$), with about five percent error. While that seems very small, there was less than five percent difference between the top three teams in the final BCS scores in Table 1. (.976 - .930 = .046). So, small fractions do matter—they matter a lot!

20/20—THE VALUE OF HINDSIGHT

The final BCS scores in Table 1 reflect the results of conference championship playoffs. This was especially true for the games in the Southeastern Conference and in the Big 12. Those two games were viewed by

many as a virtual semi-final playoff for the national championship title game since they involved three of the four of the top ranked teams in the country. Texas, however, was not involved because it finished lower in the last regular season BCS and that was the determining tie-breaker in the Big 12 between Texas, Oklahoma, and Texas Tech. In those two games, Florida won impressive victory over previously top-ranked Alabama, while Oklahoma crushed Missouri, scoring over sixty points for the fifth straight game. In the regular season, Texas had beaten both Oklahoma and Missouri, the two teams that played in the Big 12 championship, but it was denied the chance to play for the Big 12 championship by finishing lower in the BCS standings than Oklahoma and, consequently, it likely lost the chance to play for the national championship as well. Table 3 reports the findings from converting the BCS computer scores to percentages or rates. It reports the actual computer scores from each of the six services as reported by the BCS, along with the high score on each dimension at the end of the twelve game regular season. It then converts those scores as percentages of the highest score reported at that time and subtracts the highest and lowest scores to tabulate the Average computer percent score with both all six computer service scores and with the highest and lowest scores removed just like the current BCS does. Finally, there is a Revised BCS Score reported by both Oklahoma and Texas after adding the Average computer percent score (minus the high and low scores) to the two human polls.

Table 3: Revised BCS Scores and Final Regular Season BCS Results

Team	Sagarin	Anderson	Billingsley	Colley	Massey	Wolfe	
Oklahoma	94.92	0.786	322.418	0.8940	2.56	10.215	
Texas	94.61	0.803	313.293	0.9446	2.53	10.186	
High poll score	94.92	0.819	325.926	0.9446	2.56	10.215	
Percentages							Average
Oklahoma	1.000	.960	.989	.946	1.000	1.000	.983
Texas	.997	.980	.961	1.000	.989	.997	.987
Subtract high/ low scores							Average
Oklahoma	-----	.960	.989	-----	1.000	1.000	.987
Texas	.997	.980	-----	-----	.989	.997	.991
BCS Revised	Harris Poll	Coaches Poll	Computer Percent Score	Revised BCS Score			
Oklahoma	.9094	.9161	.9872	.9375			
Texas	.9115	.9154	.9909	.9393			

The results in Table 3 show that Texas had a higher overall computer score average than did Oklahoma as well as a higher overall Revised BCS Score when the computer polls were converted to percentages or rates. That is clearly different from the result in Table 1, which shows Oklahoma had the higher computer score when they are first converted to ranks (ordinal level data) before they are compiled. Texas did not play in the championship game and the wrong team, Oklahoma, got the chance to become the national champion even though the computer scores ranked Texas higher when the scores are measured and summed more accurately with interval level scores , as opposed to ordinal level.

SUMMARY AND CONCLUSIONS

The Bowl Championship Series (BCS) evaluation approach to deciding which teams play for a national championship in Division I college football is very important in a number of ways. First, the championship playoff games pay each team, and its conference, \$17.5 million. If a BCS conference has one team in the national championship then it is likely to have another team in another BCS bowl with a similar sized payout. In addition, championship teams reap many more millions from adoring fans who purchase memorabilia and merchandise to support their championship teams. It has also been shown that winning breeds more winning and that money from future financing and recruiting acts much like a cartel or a monopoly.

Most people favor a playoff. A 2005 Gallup Poll showed that sixty-five percent of college football fans favored a playoff given the results of the BCS in recent years. Last year, the University of Georgia President proposed a playoff arrangement that was rejected and subsequently, US President Barack Obama has called for a playoff system. So, the pressure to “get it right” on the BCS should mean that they want the most accurate system possible.

The BCS uses a series of polls—Harris, Coaches, and Computers—that involves samples of 113, 61, and 4, respectively. It is this last “poll” that tends to involve errors in computing a final result. What we have shown is that converting the computer scores to inverted ranks in 2008 distorted the intervals or relationships between the scores and yielded a nine percent difference or error with regard to correlations between the actual scores and the ranked scores. Small percentages mattered a lot in the outcome. The top six teams in the computer polls were within nine percentage points of each other. Overall, the error in the computer rankings generated a five percent error in the overall BCS, and the top three teams were all within five percent of each other. Texas ranked third, but it did not benefit from playing an additional playoff game as both Oklahoma and Florida did.

Going back, in 20-20 hindsight, we found that Texas should have been ranked ahead of Oklahoma at the end of the twelve game regular season when all teams had played the same number of games. If that had happened, then Texas would have re-played Missouri, which it had previously beaten handily, for the Big 12 championship and it probably would have played for the BCS National Championship in Miami’s Orange Bowl. On that basis, the BCS probably did get the wrong teams into the national championship game, and as a result, it may not have crowned the right team as national champion for 2008.

Even so, we are not condemning the use of MCDMs or decision matrix evaluations in general or even the BCS specifically. While other analysts have urged more sophisticated statistical techniques, like linear models or quadratic equations, we simply recommend that the scores be summed in the most accurate method possible. The BCS may not be BS, but it does incorporate a mathematical “glitch” that can be easily fixed. While there may eventually be a future playoff, the BCS should make this simple statistical tweak until that time. Just do the math—correctly!

AUTHOR INFORMATION

C. E. Wynn Teasley is a Distinguished Service Professor in the Department of Management and MIS at the University of West Florida. He has taught at that university for thirty-three years in the disciplines Public Administration, Criminal Justice and Management. He has authored or co-authored one book and over 35 refereed journal publications and has served on the editorial boards of six journals.

Martin J. Hornyak is an associate professor in the Department of Management/MIS department of the University of West Florida, Pensacola, Florida. He has authored and co-authored over 30 refereed journal articles and conference papers. He has worked in industry and government in the project management field and has been teaching full-time at the college level for over twenty years.

REFERENCES

1. Callaghan, T, Mucha, P.J., and Porter, M.A., “The Bowl Championship Series: A Mathematical Review, *Physics*, 2008, February, pp. 1-12.
2. Campbell, N. D., Rogers, T.M., and Finney, R. Z., “Evidence of Television Exposure Effects in AP Top 25 College Football Rankings,” *Journal of Sports Economics*, 2007, 8, 4, 425-434.
3. Cassady, C. R., Maillart, L. M., and Salman, S., “Ranking Sports Teams: A Customizable Quadratic Assignment Approach,” *Interfaces*, 2005, 35, 6, pp. 497-510.
4. Coleman, J. B., “Minimizing Game Score Violations in College Football Rankings,” *Interfaces*, 2005, 35, 6, pp. 483-496.
5. Eckard, E. W., “The NCAA Cartel and Competitive Balance in College Football,” *Review of Industrial Organization*, 1998, 13, pp. 347-369.

6. Hopkins, M., "High Correlations in Large Cluster Rating Systems," An Internet Note, <http://homepages.cae.wisc.edu/~dwilson/rsfc/rate/Hopkins2.txt>, 1997, downloaded December 5, 2008
7. Martinich, J., "College Football Rankings: Do the Computers Know Best?" *Interfaces*, 2002, 32, 5, pp. 85-94.
8. Nagel, S. S., *Public Goals, Means, and Methods*, New York: St. Martin's Press.
9. Nagel, S. S., "P/G% Analysis: An Evaluating-Aiding Program." *Evaluation Review*, 1985, 9: 209-214.
10. Nagel, S. S., "Evaluation Analysis with Microcomputers," *Public Productivity Review*, 1987, 10: 67-80.
11. Schmit, J.D., "A Fresh Set of Downs? Why Recent Modifications to the Bowl Championship series Still Draw a Flag Under the Sherman Act," *Sports Lawyers Journal*, 2007, 14, pp. 219-250.
12. Soares, C., et. al., "Bowl Championship Series Vulnerability Analysis," Proceedings of the 2007 Conference on Diversity in Computing, 2007 Association for Computing Machinery (ACM) Conference.
13. Stern, H. S., "Statistics and the College Football Championship," *The American Statistician*, 2004, 58, 3, pp. 158-175.
14. Stern, H. S., "In Favor of A Quantitative Boycott of the Bowl Championship Series," *Journal of Quantitative Analysis in Sports*, 2006, 2, 1, pp. 1-4.
15. Teasley, C. E., "Some Subtle and Some Not-So-Subtle Ways Analysts Can Determine Computer Assisted Outcomes," in *Journal of Management Science and Policy Analysis*, winter 1989, pp. 163-172.
16. Teasley, C.E. and M. Hornyak, "The Glitch That Stole Christmas From The Pac-10," *Journal of Business & Economics Research*, 2005, 3, 3, pp. 39-47.
17. West, B. and Lamsal, M., "A New Application of Linear Modeling in the Prediction of College Football Bowl Outcomes and the Development of Team Rankings," *Journal of Quantitative Analysis in Sports*, 2008, 4, 2, pp. 31-46.

NOTES