

2005

## Minimizing Game Score Violations in College Football Rankings

B. Jay Coleman

University of North Florida, [jcoleman@unf.edu](mailto:jcoleman@unf.edu)

Follow this and additional works at: [https://digitalcommons.unf.edu/bmgt\\_facpub](https://digitalcommons.unf.edu/bmgt_facpub)



Part of the [Management Sciences and Quantitative Methods Commons](#)

---

### Recommended Citation

Coleman, B. Jay, "Minimizing Game Score Violations in College Football Rankings" (2005). *Management Faculty Publications*. 1.

[https://digitalcommons.unf.edu/bmgt\\_facpub/1](https://digitalcommons.unf.edu/bmgt_facpub/1)

This Article is brought to you for free and open access by the Department of Management at UNF Digital Commons. It has been accepted for inclusion in Management Faculty Publications by an authorized administrator of UNF Digital Commons. For more information, please contact [Digital Projects](#).

© 2005 All Rights Reserved

# Minimizing Game Score Violations in College Football Rankings

B. Jay Coleman

Department of Management, Marketing, and Logistics, Coggin College of Business, University of North Florida,  
4567 St. Johns Bluff Road, South, Jacksonville, Florida 32224-2645, jcoleman@unf.edu

One metric used to evaluate the myriad ranking systems in college football is retrodictive accuracy. Maximizing retrodictive accuracy is equivalent to minimizing game score violations: the number of times a past game's winner is ranked behind its loser. None of the roughly 100 current ranking systems achieves this objective. Using a model for minimizing violations that exploits problem characteristics found in college football, I found that all previous ranking systems generated violations that were at least 38 percent higher than the minimum. A minimum-violations criterion commonly would have affected the consensus top five and changed participants in the designated national championship game in 2000 and 2001—but not in the way most would have expected. A final regular season ranking using the model was perhaps the best prebowl ranking published online in 2004, as it maximized retrodictive accuracy and was nearly the best at predicting the 28 bowl winners.

*Key words:* recreation and sports; programming; integer.

*History:* This paper was refereed.

The National Collegiate Athletic Association (NCAA) conducts officially sponsored national football championship tournaments in all four of its divisions, except for Division 1-A (its highest division). It has determined the Division 1-A football champion via a poll or a mathematical-ranking system since 1926. The top-rated team(s) according to the two so-called major polls, the Associated Press (AP) nationwide poll of sportswriters and broadcasters, which began in 1936, and the USA Today (formerly USA Today/ESPN) poll of coaches, which began in its current form in 1991, is (are) generally regarded by the public as the national champion(s) each year (NCAA Online 2004). Moreover, since 1998, the Bowl Championship Series (or BCS, operated by the six most powerful Division 1-A conferences and Notre Dame) has employed opinion polls, such as these, along with a variety of mathematical-ranking systems as part of a formula to determine which two teams to select to participate in its officially recognized national championship game. The winner of this game is deemed to be the national champion according to the USA Today poll (*BCS Media Guide* 2003). The BCS also uses its ranking to determine eligibility for invitation to the BCS's other three games

(BCSfootball.org 2005a), with payouts of \$14 to 17 million to each team (BCSfootball.org 2005b), compared to a payout of \$7.4 million for the best non-BCS bowl (Staples 2005). The ramifications of these rankings (monetarily and otherwise) are therefore quite substantial.

However, the NCAA recognizes 12 different polls (or entities naming national champions) and 21 different mathematical-ranking schemes on its Web site, some of which have determined champions retroactively as far back as 1869 (NCAA Online 2004). Of these 33 systems, 21 were in use for 2003. Although the AP and the USA Today opinion polls tend to get the greatest publicity, 14 of the 21 systems the NCAA listed for 2003 were mathematical. In addition, many authors have developed their own mathematical approaches to ranking college football teams. Kenneth Massey, whose mathematical-ranking system is part of the BCS formula, maintains a Web site, which at the end of the 2003 season posted the results of 97 different ranking approaches (Massey 2004b). Massey also compiles and posts a consensus ranking derived from all 97 sources and provides links to any associated Web sites provided by the ranking authors.

For many of the most widely recognized mathematical systems (for example, six of the seven systems used in the BCS formula in 2003 (Callaghan et al. 2004)), neither the NCAA nor the authors of these systems provide detailed descriptions of the models, and apparently no descriptions have been published. This lack of description makes it difficult to assess the merits of each method, and it likely contributes to the public's disregard for the results and preference for the less objective opinion polls. Another likely contributor to the math systems' unpopularity is a somewhat deserved public skepticism regarding black-box computerized math models (Keener 1993).

Perhaps the leading cause of the media's and the public's disdain for mathematical-ranking systems is the common face-validity problem that surfaces when a team is ranked below an opponent that it defeated during the season, particularly when the teams are very close to each other in the rankings (that is, team A defeated team B, yet team B is ranked immediately or a few spots ahead of team A). It's even more disturbing when the two teams are in the top five or top 10 and greatly so when such a ranking influences which two teams play in the BCS national championship game. For example, at the end of the 2000 regular season, the BCS designated Florida State University (FSU) to play for the national championship (as the second-ranked team in the country, according to the BCS formula), even though Miami (FL) had defeated FSU during the course of the season and was immediately behind FSU in the ranking. The mathematical-ranking systems included in the BCS formula were the cause of the controversy, because the AP and USA Today/ESPN opinion polls used in the formula both ranked Miami second, immediately ahead of FSU. In a similar outcome in 2001, Nebraska was second in the final BCS ranking and thus was selected to play in the national championship game, even though third-ranked Colorado had defeated Nebraska by a large margin in the last game of the regular season. As in 2000, the third-ranked team in 2001 (Colorado) was ranked ahead of Nebraska by the two major opinion polls. In both cases, the influence of the math systems in the formula caused the reversal, and people's perception of the math models took a major public relations hit.

In the jargon of tournament theory, these reversals of game results are called violations in the ranking (Ali et al. 1986). There are two primary reasons why such reversals are common among the math models (and why they are common among opinion polls). First, to some extent, they are unavoidable. The nature of game results is such that many cyclic triads are always present, in which team A beats team B, team B defeats team C, and team C beats team A. In the presence of a cyclic triad, there is no way to rank the three teams involved to avoid a game score reversal. Second, designing a mathematical procedure to minimize the number of violations in a ranking has proven extremely difficult and computationally time-consuming for large-scale problems (Grotschel et al. 1984, Laguna et al. 1999). This difficulty partially explains why none of the math systems the BCS currently uses and none of the 97 methods Kenneth Massey listed in 2003 on his Web site have been designed specifically to minimize violations.

However, given that the number of violations in the ranking is a retrodictive indicator of ranking quality (as Massey 2004a states, it's an indication of how well a system explains past results), Massey regularly reports a ranking violation percentage for each of the systems compiled on his site. This metric represents the percentage of all games played by the teams ranked in which the losing team is ranked ahead of the game's winner, without accounting for any adjustments for home-field advantage and without weighting any games more heavily than others. For example, for the 2003 season, the average violation percentage was 18.5 percent, with a range of 12.4 percent to 22.6 percent, for the 93 (out of 97) systems that ranked all 117 Division 1-A teams. The four omitted included the two major opinion polls that rank only a smaller number, such as a top 25 (Massey 2004b). In 2003, the total number of games played by the 117 teams in Division 1-A against each other was 699, meaning that these 93 systems produced an average of approximately 129 game score reversals, with a range from 87 to 158.

Do any of the current systems come close to creating a ranking that minimizes the number of game score reversals? Given the disfavor and controversy that surround systems that clearly do not directly address the issue, it seems a worthy question to

pursue. Yet no previous analyst has successfully done so. I developed a fairly simple mixed binary integer-programming model efficient enough to determine the minimum number of violations possible (and thus the maximum retrodictive accuracy) for a full season of Division 1-A college football. The model gains its efficiencies by exploiting aspects of the ranking problem in college football that other models have typically not addressed for minimum-violations rankings. I applied the model to the 10 complete college football seasons of 1994 through 2003 to determine the minimum violations for those seasons and compared the results to the ranking systems Massey compiled for 2000 through 2003. I then used the model as part of an iterative procedure to determine whether the consensus top five teams in any of these 10 seasons would have been changed if one had enforced a minimum-violations criterion. I also examined whether the participants in any of the BCS national championship games from 1998 through 2003 would have been altered by a minimum-violations restriction.

## Background

The minimum-violations problem is a version of the more general linear-ordering problem, for which Grotschel et al. (1984) and Schrage (1991) developed mathematical-programming methods, and which has wide-ranging applications outside the realm of sports, including the triangulation of input-output matrices in economics (Grotschel et al. 1984, Laguna et al. 1999). However, the college football application does not adhere strictly to the criteria for a pure linear-ordering problem. One major difference is that I allow teams to be tied in the final rankings, a less restrictive stance that yields a more efficient mathematical model than those developed for strict linear ordering. In other words, if teams A and B did not play each other during the season, it doesn't matter whether A is ranked ahead of B, B is ranked ahead of A, or A and B are tied in the ranking—none of the three scenarios will create a violation.

Another difference is that most previously offered models focused on developing the minimum-violations ranking for a so-called *complete tournament* (Ali et al. 1986, Cook and Kress 1990). From the perspective of sports, a complete (or round-robin) tournament is one in which all players (or teams) to

be ranked have played each other at least once. This is clearly not the case for Division 1-A college football, where the 117 teams currently play only 11 to 14 games, and this set of games may even include the occasional second game against the same team. The result is thus described as an incomplete tournament—and in this case, a very incomplete one.

Methods previously developed for linear ordering or determining a minimum-violations ranking have had these and other assumptions that complicate their use for ranking college football teams. Ali et al.'s (1986) mathematical-programming model and enumeration method have the restrictions of no ties in the rankings or the games and an assumption of a complete tournament. The former is currently true in college football but has not always been so (the 1996 institution of overtime abolished ties). To the extent that my model has applications in other environments in which ties are possible (for example, the National Football League), allowing for tied games would be advantageous. The developers of previous heuristic methods have made similar assumptions. For example, Cook and Kress (1990) allowed for tied games but also assumed a complete tournament. Goddard (1883, 1985) and Stob (1985) largely enforced similar restrictions. Although I could use Kendall's (1962) method of minimizing violations by ranking based on the number of wins or the iterated Kendall (IK) method of Ali et al. (1986) that breaks ties by examining opponent wins, the accuracy would be suspect given the incomplete-tournament format in college football. None of these heuristics could ensure optimality for the current problem.

Beyond these assumption problems, previous optimization models also suffer from serious efficiency issues when addressing college football. The largest complete-tournament, linear-ordering problems that authors reported solving with optimization models contain only 60 teams (or items) to be ranked (Grotschel et al. 1984), roughly half the number of Division 1-A teams in 2003. The reason for the intractability of large-scale problems is the potential exponential proliferation of constraints and binary integer variables associated with the growth in problem size (Ali et al. 1986, Grotschel et al. 1984, Schrage 1991). Although Grotschel et al. (1984) developed algorithms for solving large cases efficiently (and

using today's technology, they would be capable of solving much larger problems than the 60-item cases reported in 1984), these algorithms require programming effort far beyond that required by my model, which can be solved by standard commercially available linear-programming software. Even fairly efficient enumeration techniques, such as that Ali et al. (1986) developed, can be far too time-consuming for a problem the size of a full college football season.

In contrast, the presence of an incomplete tournament, and the allowance for ties in the ranking, yield definite modeling efficiencies when determining the minimum number of violations. We need binary variables for a pair of teams only if those teams played each other during the season, meaning that we require only 695 binary variables for the 2003 season (for a team that defeated the same opponent twice, we need only one binary variable; which explains why the 699 games in 2003 can be modeled with 695 binary variables). Moreover, my modeling approach differs from that used by Ali et al. (1986), Grotschel et al. (1984), and Schrage (1991), in requiring only a constraint for each pair of teams that actually played, instead of constraints involving pairs of teams regardless of whether they played. My model for the 2003 season therefore has only 695 constraints.

## Methodology

I used my mixed binary integer program (Appendix), model MinV, to determine the minimum violations for college football. The model defines a continuous variable representing the value (not the rank) for each team; this continuous variable can be thought of as the team's power index. Using a constraint for each game played, MinV sets a binary variable (representing a violation) equal to one if the value (power index) of the winning team is not greater than the value (power index) of the losing team by some arbitrary minimal amount. Otherwise, it sets the binary (violation) variable for that game to zero. For each tied game, MinV uses two additional continuous variables and two binary variables (instead of one): one of these variables will equal one if team A's power index is greater than team B's, and the other will equal one if the reverse is true. If the power indexes of the two teams are equal, both binary variables will equal zero.

The objective function represents the sum of all the binary variables (the sum of all the violations), counting any violation in a tied game as half a violation. Once MinV is solved, I simply rank order the teams according to their power indexes reported in the solution. If desired, ties in the ranking between teams that did not play each other can be broken arbitrarily (or using a secondary objective) with no change in the number of violations.

The model considers multiple game situations in which the same two opponents each won once to be separate events, with separate binary variables and separate constraints. This implies that if the same two teams played each other twice, with each team winning one game, MinV will essentially force one team to be ranked ahead of the other (the model solution will determine which), and one violation will be counted in the objective function. In such a case, one could argue that it might be more appropriate to allow the two teams to tie in the final ranking, and if the solution does so, no violations should be assessed. However, I allowed a direct comparison to the reported ranking-violation percentages of the various systems summarized by Massey (2004b, c), which do not exhibit any ties in the ranking. (For the seasons Massey covered, the NCAA allowed no game ties because of the overtime rule. Thus, MinV's components that encourage ties in the ranking between teams that actually tied a game did not influence comparative results.)

I applied my model to the college football seasons from 1994 through 2003. I collected files for complete-season game scores for 1994 through 2002 from James Howell (2004), who is widely regarded as a reliable source for scores files, for example, by Peter Wolfe (2003), whose ranking system the BCS includes in its formula. I collected the file for 2003 game scores from Massey's (2004e) Web site. I also collected from Howell (2004) lists of the Division 1-A teams that participated in each of the 10 seasons. I parsed the scores files to eliminate data on games played by non-Division 1-A teams and on games played by Division 1-A schools against non-Division 1-A schools, so that the model analyzed only data on games played by two Division 1-A schools. I wrote code to read each season's game scores and write model MinV. I used LINGO Version 8.0 (LINDO Systems 2003), running

on a Pentium 4 (2.53 GHz) machine with 1 GB of RAM to determine optimal solutions to MinV for the 10 seasons.

## Results

MinV for each season generally determined the minimum number of violations within a few minutes (Table 1). For five of the 10 seasons, the original run of the model took less than 1.5 minutes. After some minor adjustments to one component of the model (Appendix), the running time in another two seasons (2001 and 2002) was under 40 seconds. The only real computational difficulty experienced was for 2000. Within five minutes of running time, the model identified a high-quality solution that was within one violation of the lower bound, but LINGO did not establish this as the optimal solution even after an extremely lengthy run. However, subsequent runs of an adjusted form of the model quickly established this value as the optimal solution (Appendix).

The minimum number of violations MinV established was much lower than that obtained by even the best-performing existing college-football ranking system (Table 2). For the 2003 season, the best retrodictive system Massey reported generated a violation percentage of 12.4 percent of the 699 games played, which translates into 87 game score violations. By comparison, MinV established a minimum number of violations of 63 (or nine percent of the games played). Thus, the best system in 2003 generated a number

of violations that was over 38 percent higher than the minimum identified by MinV. The overall average violation percentage of 18.5 percent (or about 129 violations) for the 93 systems reported for 2003 represented a 105 percent increase over the MinV result. I obtained similar results for 2000, 2001, and 2002, where MinV improved the best existing retrodictive systems by 45 violations (a total 82 percent higher than MinV’s minimum), 33 violations (a total 61 percent higher than MinV’s minimum), and 34 violations (a total 62 percent higher than MinV’s minimum), respectively. When compared to the average and worst-case performances of all published rankings, the improvements are far more extensive. Since 2000, one online source (Beck 2004b) has given an award of sorts for the best retrodictive ranking system; MinV would have won in a landslide for all past seasons.

## Impact on the National Champion and the Top Five

One criticism levied against the use of a minimum-violations-ranking criterion is that such models do not necessarily produce a unique ranking (Stob 1985). Multiple optima often exist, particularly for minimum-violations rankings for very incomplete tournaments. Thus, although model MinV identifies one ranking that minimizes the number of game score violations, there are likely others that would be more attractive if one pursues a secondary objective while holding the minimum number of violations constant.

Year	Division 1-A teams	Division 1-A games	Total variables (binary)	Total constraints	Pivots (minutes:seconds) Required by LINGO
1994	107	593	739 (606)	619	67,514 (0:25)
1995	108	597	731 (605)	614	68,432 (0:17)
1996	111	614	725 (614)	614	3,942,014 (7:23)
1997	112	628	740 (628)	628	94,737 (0:23)
1998	112	627	739 (627)	627	17,260 (0:06)
1999	114	633	745 (631)	631	356,585 (1:25)
2000	115	641	754 (639)	639	250,701,233 (613:08)
2001	117	652	769 (652)	652	39,860 (0:10)
2002	117	707	823 (706)	706	76,847 (0:38)
2003	117	699	812 (695)	695	1,272,280 (5:10)

**Table 1: MinV generally determined the minimum number of violations in a few minutes for each NCAA Division 1-A college football season from 1994–2003 (Appendix). For 2000, I terminated the branch-and-bound process at the point noted, with the number of violations within one of the lower bound. Subsequent analysis confirmed the solution’s optimality (Appendix).**

Year	Division 1-A games	Minimum number of violations	Ranking system average (%)	Ranking system minimum (%)	Ranking system maximum
1994	593	45.5 (7.7%)	—	—	—
1995	597	45.5 (7.6%)	—	—	—
1996	614	46 (7.5%)	—	—	—
1997	628	43 (6.8%)	—	—	—
1998	627	44 (7.0%)	—	—	—
1999	633	63 (10.0%)	—	—	—
2000	641	55 (8.6%)	18.3	15.6	23.1%
2001	652	54 (8.3%)	17.3	13.3	21.0%
2002	707	55 (7.8%)	17.5	12.6	21.2
2003	699	63 (9.0%)	18.5	12.4	22.6%

**Table 2: The minimum number of game score violations MinV identified for the NCAA Division 1-A college football seasons from 1994–2003 was far superior to the numbers that 58, 68, 75, and 93 different ranking systems (Massey 2004b, c) produced in the years 2000, 2001, 2002, and 2003, respectively. Massey did not report ranking violation statistics prior to 2000.**

However, using an extension of model MinV to pursue a secondary objective (for example, matching game score differentials as closely as possible) while holding the minimum violations constant, is much more mathematically inefficient and is likely a worthy avenue for future research.

However, it is fairly easy to extend the model to determine whether the top five teams and particularly the consensus national champion in each year would have changed if one had used a minimum-violations criterion. For the eight seasons of 1996 through 2003, Massey derived a consensus final ranking and a consensus national champion by averaging together all of the rankings from all the sources compiled for that year (Massey 2004b, c). I made matching

his consensus top five for those seasons a secondary objective. For 1994 and 1995, I used the top five teams according to the AP media poll as a surrogate secondary target. Using model MinV as part of an iterative process (Appendix), I identified a top five that (1) did not force the minimum number of violations to be surpassed, and (2) most closely matched the consensus (or, in 1994 and 1995, the AP) top five (Table 3).

The consensus national champion would not have changed in any of these seasons had a minimum-violations criterion been enforced. In all 10 years, the top team could have been assigned that position without increasing the minimum number of violations. I offer no comment on whether the consensus (or AP) national champions were the appropriate teams to

Year	Number 1	Number 2	Number 3	Number 4	Number 5
1994	Nebraska (1)	Penn State (2)	Colorado (3)	Alabama (5)	Texas A&M (8)
1995	Nebraska (1)	Florida (2)	Tennessee (3)	Florida St. (4)	Colorado (5)
1996	Florida (1)	Ohio St. (2)	Florida St. (3)	Arizona St. (5)	Nebraska (4)
1997	Nebraska (1)	Michigan (3)	Florida (4)	Florida St. (2)	Tennessee (5)
1998	Tennessee (1)	Ohio St. (2)	Florida St. (3)	Wisconsin (4)	Florida (5)
1999	Florida St. (1)	Nebraska (2)	Va. Tech (3)	Michigan (4)	Kansas St. (5)
2000	Oklahoma (1)	Nebraska (8)	Washington (6)	Miami (FL) (2)	Florida St. (3)
2001	Miami (FL) (1)	Tennessee (3)	Florida (2)	Oregon (4)	LSU (9)
2002	Ohio St. (1)	Miami (FL) (3)	Georgia (4)	Oklahoma (5)	Texas (6)
2003	LSU (1)	Southern Cal (2)	Oklahoma (3)	Georgia (4)	Miami, OH (6)

**Table 3: Although the national champions would not have been affected, the remainder of the consensus top five NCAA Division 1-A college football teams for the 1994–2003 seasons often would have been adjusted had a minimum-violations restriction been enforced (consensus rankings are in parentheses). I used Massey's (2004b, c) consensus rankings for 1996–2003 and the final Associated Press poll in 1994 and 1995. Massey did not report a consensus ranking for 1994 or 1995.**

recognize, only that they could have been named without surpassing the minimum violations.

Although the national champion would not have changed in any year, a minimum-violations criterion would have affected the remainder of the top five in many years. In only three seasons (1995, 1998, and 1999) could the consensus (or AP) top five have remained the same. In four seasons (1997, 2000, 2001, and 2002) the consensus ranking would have changed the second spot. Some of these second-spot shifts would have been minor: in 1997, Michigan (consensus third) would have taken Florida State’s spot at second place, and Tennessee (consensus third) would have replaced Florida in second place in 2001. However, some of the second-place changes would have been dramatic. The 2000 season would have seen the greatest move by one team: Nebraska would have moved six spots from eighth to second, replacing Miami (FL), which would have dropped to fourth. In 2002, Southern California, the hot team during the latter part of that season, would have dropped from consensus second to completely out of the top five with the minimum-violations restriction. (The impact on Southern Cal highlights the fact that a minimum-violations objective weights games early in the year the same as those later on.) Other significant shifts in the top five would have included Louisiana State University’s (LSU’s) move from consensus ninth up to fifth in 2001, Washington’s move from consensus sixth to third in 2000, and Texas A&M’s rise from consensus eighth to fifth in 1994.

### Impact on the National Championship Game

For followers of college football in recent years, the existence of the BCS and its designated national championship game has been a consistent source of debate. One of the reasons for the controversy is the BCS’s partial reliance on mathematical-ranking systems to determine which two teams should play in the game and the resulting face-validity problem. These problems were particularly acute at the end of the 2000 and 2001 seasons (as discussed earlier) and also at the end of the 2003 season. In 2003, the BCS selected Oklahoma and LSU to play in the national championship game, even though Southern Cal held

the top position in both the major opinion polls going into the 2003 bowl games.

In an effort to determine whether a minimum-violations criterion would have changed the participants in these years or in any of the BCS’s six years of 1998 through 2003, I solved MinV while examining only the regular season and conference championship games in each year (that is, all games prior to the postseason bowls). These are the games that the systems used in the BCS ranking formula examine when determining the two teams to select. Once I had solved MinV, I followed an iterative process (Appendix) similar to the one I used to identify the top five teams to determine whether the BCS could have selected its top two teams without surpassing the minimum number of violations. If not, I proceeded with the process to establish the two highest-ranked teams in the BCS rankings (*BCS Media Guide 2003*, *BCS Rankings Week 8 2003*) that would have allowed it to meet the the minimum number of violations (Table 4).

For the most part (in four of the six years), the BCS could have selected the two teams it had without surpassing the minimum number of violations. However, the minimum-violations criterion would have caused major changes in 2000 and 2001, although maybe not quite the changes I and likely many others would have expected. In 2000, the controversy swirled around the BCS picking FSU (second) ahead of Miami (FL) (third) even though Miami had beaten FSU during the season. However, what did not get as much attention was the fact that Washington (BCS fourth) had beaten Miami during the season as well.

Year	Minimum number of violations	Actual BCS participants	Participants with MinV adjustments
1998	41	Tennessee, Florida St.	Tennessee (1), Florida St. (2)
1999	58	Florida St., Va. Tech.	Florida St. (1), Va. Tech. (2)
2000	51	Oklahoma, Florida St.	Oklahoma (1), Washington (4)
2001	51	Miami (FL), Nebraska	Miami (FL) (1), Oregon (4)
2002	50	Ohio St., Miami (FL)	Ohio St. (1), Miami (FL) (2)
2003	59	Oklahoma, LSU	Oklahoma (1), LSU (2)

**Table 4: The Bowl Championship Series (BCS), begun in 1998, would have selected different participants in its national championship games in 2000 and 2001 if it had adjusted its final ranking to adhere to the minimum number of violations (actual BCS rankings are in parentheses).**



As a result, enforcing a minimum number of violations would have caused Washington to be the preferred pick to face Oklahoma (BCS number 1). In 2001, the controversy focused on Nebraska (BCS second) getting the nod over Colorado (BCS third), which had beaten Nebraska soundly at the end of the regular season. However, Colorado had come on strong at the end of the year after two earlier losses. The minimum-violations objective would have accounted for this fact by naming Oregon (BCS fourth) to face Miami (FL) (BCS number one) in the championship game. This result is particularly interesting because Oregon was the second team (behind Miami (FL)) in both the major opinion polls. Thus, the math systems that caused the BCS to select Nebraska (fourth in both opinion polls) over Oregon actually forced a deviation from the minimum-violations ranking. Finally, a minimum-violations ranking could not have soothed the turmoil in 2003, as it would have allowed the BCS's controversial picks of Oklahoma (number one) and LSU (second) to play in the 2003 national championship game, even though many observers felt Southern Cal (BCS third) should have faced LSU. The BCS could also have placed LSU and Southern Cal in the top two without surpassing the minimum violations.

### Retrodictive vs. Predictive Accuracy

A concern regarding any ranking system designed to optimize retrodictive accuracy is whether retrodictive accuracy actually leads to predictive accuracy. The optimization of predictive accuracy is a different problem—and ranking philosophy—from that MinV addresses. Models developed to address predictive accuracy often weight recent games more heavily than early games, use information from prior seasons (that is, start with a preseason ranking), include adjustments for home-field advantage, and consider margin of victory in developing the ranking (Wilson 2005b). Indeed, some retrodictive models do some of these things. However, the overall ranking philosophy, or the definition of what is fair, is the feature that most distinguishes predictive and retrodictive models. The general philosophy of predictive models is to determine which teams are strongest, whereas the philosophy of a retrodictive model would be to establish which teams had the most impressive season

(under a minimum-violations philosophy, measured by actually winning games) and the two are not necessarily the same (Wilson 2005b).

To understand the differences, suppose that in the opening game of the 2004 season, Virginia Tech (VT), which was not highly rated going into the season, had upset Southern Cal (USC), which was the top-rated team in the preseason polls, and suppose that the game had been played at USC. A retrodictive (minimum-violations) model would have ranked VT ahead of USC after that game, because VT had the more impressive season up to that point. However, a predictive model may very well have maintained USC ahead of VT despite the loss, based on preseason information and the expectation that USC would still end up being a stronger team over the course of the season. With either model, consternation among some observers would have been likely. For example, if USC were still ranked ahead of VT, many (those that favor a retrodictive (minimum-violations) philosophy) would view that ranking as unfair to VT. If USC were ranked behind VT, those that believe a ranking should reflect predictive accuracy might view that ranking as unfair if they thought that USC was still the stronger team. (Bihl 2002 discusses the final regular-season ranking of Florida State and Florida in 1996 in a similar vein.)

A similar issue arises with adjusting for home-field advantage. Suppose that VT had lost to USC by two points on the road. Had a ranking model included a home-field adjustment of three points (as most predictive and many retrodictive systems would), it essentially would have deemed VT the winner of the game for ranking purposes. Observers who believe actually winning the game regardless of situational factors is most important to determining rankings (those who would espouse a minimum-violations philosophy) would view that as unfair to USC. Those who believe that a ranking should reflect team strengths might agree that after accounting for playing on the road, VT was indeed the better team and that it would be unfair to rank them otherwise.

Which approach is best is in many ways irreconcilable because it's largely a matter of which ranking philosophy one considers fair. Both perspectives have their proponents. Of the six mathematical systems included in the BCS ranking in 2004, at least three

are retrodictive (Wilson 2005a), and one (Colley 2002) makes no adjustments for margin of victory, home-field advantage, or date of the game (as does MinV). The author of one of the BCS's predictive systems, Billingsley (2002), considers date of the game and makes minor adjustments for game site but notes the importance of fairness in recognizing head-to-head competition (that is, violations) in the ranking. For that matter, the BCS ranking itself has been adjusted in recent years to directly reflect head-to-head competition (Martinich 2002). The interest in retrodictive performance is further shown by Massey's inclusion of violation percentage on his compilation page, Beck's (2004b) retrodictive accuracy award, Hopkins' (1997) discussion, and the feedback regarding MinV I received during the 2004 season. In contrast, many do not favor a retrodictive or minimum-violations-ranking approach: for example, Martinich (2002) supports a predictive approach and the application of a prediction metric, and Bihl (2001) favors retrodiction but not necessarily a minimum-violations metric. However, even predictive proponents may be interested in a minimum-violations ranking for comparative purposes.

Exemplifying the interest in both philosophies and adjustments to retrodictive systems, Wilson (2005a) maintains a comprehensive online list of college football ranking systems. As of July 21, 2005, Wilson classified 102 systems as either retrodictive (56) or predictive (46). I examined the Web sites for the retrodictive systems and found that many consider such factors as margin of victory and home-field advantages, but at least 19 do not consider home-field advantages, and at least 18 do not consider margin of victory.

Clearly a range of views exist within and across the philosophies. The philosophy behind MinV is that winning the game is the only thing that matters, regardless of situational factors and regardless of whether the winning team may not be the one expected to perform better in later games. Thus, for example, in MinV, I don't include an adjustment for home-field advantage nor do I weight recently played games more heavily than earlier games. Given the objective of the model to determine the minimum violations, and how such a criterion would affect the

two major decisions based on rankings (who's number one at the end of the season, and who should play in the national championship game), I do not adjust for either factor. However, the model could easily be modified to do either or both. A single global home-field advantage variable could be added (or subtracted) from the left side of each constraint based on whether the home team won the game. More recently played games could be weighted more heavily by simply increasing the weight (the objective function coefficient) on the violation variables for later games. This weight could be incrementally increased for each week (or on each game played by a given team) over the course of the season. The foundation provided by MinV will allow for the future evaluation of the predictive (or retrodictive) impact of both these modifications.

Given the philosophy I employ, the MinV ranking may not make the best predictions of future games; as Martinich (2002) demonstrated, retrodictive accuracy does not necessarily result in predictive accuracy. However, no prior researcher has assessed the relationship between these two perspectives, because none have developed systems that optimize the retrodictive component (as measured by violation percentage) for college football. MinV allows that as a future research objective.

## Experience in 2004

My experience in regularly publishing MinV rankings online during the 2004 season gave me a partial look at this relationship and at MinV's potential predictive accuracy. Early in the 2004 season, I told Kenneth Massey about the MinV model and asked him to post weekly rankings from the model during the 2004 season. Starting on October 16, 2004 (or halfway through the season) and continuing through January 4, 2005 (the final nine times when rankings were widely published for all other systems), Massey (2004d) posted a weekly ranking from MinV on his ranking compilation page. At his suggestion, I developed a Web site linked from Massey's compilation page that provided the complete MinV ranking for the week and an overview of the number and percentage of violations for that week.

To develop the posted ranking for each week, I used game scores from Wolfe (2004) and solved

MinV to establish the minimum number of violations for the games through that weekend. Then, I solved the model iteratively by adding constraints to encourage a solution that would match the consensus ranking of those systems that Massey had compiled at the time I made the MinV runs, or Massey's own ranking, which was highly correlated (99.3 percent) to the consensus ranking and used twice, on November 13 and 20. (Massey's final consensus for each week was ultimately affected by rankings posted later in the week and by the MinV ranking itself.) I tried to encourage a ranking that matched the secondary target. However, I solved the model only 15 to 20 times with various sets of constraints added or deleted at each stage, focusing only on the top 25 or so teams. My objective was to obtain a reasonable-looking ranking for Massey to post to establish a minimum-number-of-violations benchmark against which to compare the violation percentage of the other compiled systems.

The MinV ranking was consistently at or near the bottom (of roughly 100 systems compared each week) in its correlation to the Massey consensus. For example, at the end of the regular season (prior to the bowl games), the posted MinV ranking's correlation to the Massey consensus was 90.4 percent, the lowest among the 99 systems compiled that week (Massey 2004d). The average was 97.1 percent, with a high of 99.8 percent. Clearly the posted MinV ranking each week reflected the minimum-violations criterion and not simply the consensus (or the Massey ranking) that I used as the secondary objective.

Even so, the MinV rankings would have yielded intriguing results as prediction tools during 2004. They would have been mediocre predictors during the regular season. Beck (2005a, b) computed and compared the weekly prediction accuracies of up to 47 ranking systems (not including MinV) each week of 2004; 42 of these systems made predictions every week in which I posted a MinV ranking. When I compared the accuracies of these 42 systems to MinV's accuracy for the last half of the regular season (October 23 through December 4), I found that MinV's accuracy would have compared as follows for the seven weeks (in number of standard deviations from the average number of correct predictions by Beck's 42 systems): 3.23, 1.17, 0.45, 0.60, 1.79, 0.15,

and 0.12. In other words, performance was typically below average but essentially average in four of the last five weeks. These results are not surprising, as these regular season games may be influenced by home-field advantages and recent performances, which were not built into MinV. The comparatively poor results for the first week are also not surprising, as the small sample size at that point (five or six games for most teams) adversely affects retrodictive systems more than predictive systems, which, unlike retrodictive systems, incorporate information from prior years. Without the week of October 23, the aggregate MinV predictions over the last six weeks (covering 234 games) were not statistically different from the average, even using a 10 percent significance level.

However, the results for bowl game prediction performance were markedly different. The comparison of bowl prediction performance is one of the best comparisons to make regarding predictive accuracies. By that time, the largest possible sample of games within the season is available on which to base predictions. The bowl games are typically more evenly matched than many of the games played during the season, making their prediction more difficult. They are played at largely neutral sites. Their separation by a few weeks likely reduces any carry-over effect of each team's recent performances. The sample of 28 bowl games also is reasonably sized (in contrast, the prediction accuracy for the week of December 4 was based on only 10 games). Finally, the two rankings for a season that most matter are the prebowl ranking and the postbowl ranking. The BCS (and other bowl officials) use the prebowl ranking in determining bowl match-ups, and observers use the postbowl ranking in determining team positions for the year.

For these rankings, MinV was extremely accurate; its prebowl ranking (through the games of December 4, 2004), published online in advance of all bowls, predicted the winners of 19 of the 28 games. Beck (2005b) and Trono (2005) compared the bowl prediction accuracy of published ranking systems: 47 by Beck and 99 by Trono. MinV was better than all 47 systems Beck evaluated, each of which successfully predicted 18 or fewer games. In Trono's comparison of 99 systems, including MinV, MinV tied for fourth in prediction accuracy. The most accurate system

predicted 22 of the 28 games, and the second- and third-place systems predicted 20 games. The 98 systems (other than MinV) averaged 15.76 games correct, with a standard deviation of 1.65 games, meaning that MinV's accuracy was nearly two standard deviations better than the average. (It would have been nearly 2.5 standard deviations better than the average of Beck's systems.) Jeff Sagarin's system, the leading predictive system for the season and for the season's second half according to Beck's statistics, correctly predicted 15 games, the same number that would have been correctly predicted by picking the Las Vegas betting favorites (Beck 2005b). By comparison, the Massey consensus ranking used as the secondary target would have predicted only 14 of the games correctly (Beck 2005b), meaning that it would have surpassed only six of the 99 systems Trono compared. Thus, even though the Massey consensus influenced the MinV ranking as the secondary objective, the minimum-violations objective clearly improved it considerably. In addition, the MinV bowl predictions would have gone 17 for 28 against the Las Vegas point spreads Beck (2004a) reported, which would have placed MinV fourth against the spread versus Beck's 47 other systems. The probability of achieving this level of success against the spread by chance is less than 10 percent.

The 2004 MinV prebowl ranking was the best published ranking when compared against all 47 systems Beck evaluated. It had the best predictive accuracy and the best retrodictive accuracy (an 8.2 percent violation percentage, whereas the next best system's violation percentage was 43 percent higher). No other system approach this combination, which implies that the MinV ranking essentially reconciled the retrodictive and predictive ranking philosophies at the final prediction point in 2004. Of the three systems Trono evaluated with better bowl-prediction performance, none approached the retrodictive accuracy of MinV, as their violation percentages were 59 percent, 96 percent, and 104 percent higher than MinV's (Massey 2004d). Also, according to Beck, the next best retrodictive system (as reported by Massey 2004d) had a bowl-prediction performance of 16 games correct (tied for 13th), with an against-the-spread record of 15 for 28 (tied for 18th). Even when accounting for the point spread, MinV's performance was well above the

chance level of success and among the top performers. These results indicate that pursuing a minimum-violations ranking (or retrodictive accuracy) can also lead to high predictive accuracy and that further evaluation of this relationship merits research.

## Additional Interfaces

People have shown interest in the MinV model. Kenneth Massey asked me to provide weekly MinV rankings during the 2004–2005 college basketball season, in addition to those for the football season. I developed another Web site for this purpose and posted MinV rankings 12 times during the season, including a pre-NCAA Tournament ranking covering all regular-season and conference-tournament games (the games the NCAA Tournament selection committee examines when determining its at-large invitees). Although the size of the basketball-ranking problem (with 326 teams playing 30 games instead of 117 teams playing 11–14 games) precluded identification of guaranteed optimal solutions in a reasonable time, the model's best pretournament ranking (Massey 2005) produced a violation percentage of 18.3 percent. The next best violation percentage among the 38 other ranking systems compared was 22.8 percent, or almost 25 percent higher.

Other Web sources also responded favorably to the introduction of MinV online during 2004. Eugene Potemkin (2005) included it in his Frank of rating systems. His site is linked by Massey. His system assesses the fairness of ranking systems from multiple points of view and overall. During the last half of the 2004 season, MinV consistently rated number one according to Potemkin's fairness metric when viewed from the teams' point of view and occasionally rated number one overall (out of as many as 105 ranking systems). Ray Waits and David Wilson, who operate another Web site devoted to analyzing ranking systems, included MinV among the nine systems on their INCAA Football Superlist, which include those developed by BCS participants Kenneth Massey, Jeff Sagarin, and Peter Wolfe (Waits 2005).

## Appendix

Let the variables be defined as:

$VIOLATION_i = 1$  if loser ranked above winner of nontied game  $i$ , 0 otherwise;

$VIOLATIONA_k = 1$  if team A ranked ahead of team B in tied game  $k$ , 0 otherwise;

$VIOLATIONB_k = 1$  if team B ranked ahead of team A in tied game  $k$ , 0 otherwise;

$TEAM_j$  value (not the rank) of team  $j$  (that is, the power index for team  $j$ );

$DIFFA_k$  difference in values (not the ranks) of the two teams A and B in tied game  $k$ , if A is ranked ahead of B, over and above the minimum difference allowed between two teams with different ranks; and

$DIFFB_k$  difference in values (not the ranks) of the two teams A and B in tied game  $k$ , if B is ranked ahead of A, over and above the minimum difference allowed between two teams with different ranks.

Let the problem parameters be defined as:

$G$  total number of nontied games;

$i$  subscript for a given game in the set of nontied games;

$T$  total number of tied games;

$k$  subscript for a given game in the set of tied games;

$N$  total number of teams;

$j$  subscript for a given team in the set of all teams;

$W_i$  subscript of winning team in nontied game  $i$ ;

$L_i$  subscript of losing team in nontied game  $i$ ;

$A_k$  subscript of one team in tied game  $k$ ;

$B_k$  subscript of the other team in tied game  $k$ ;

$S$  minimum difference desired between  $TEAM_{W_i}$  and  $TEAM_{L_i}$  (chosen arbitrarily); and

$M$  sufficiently large positive number.

The minimum number of violations was determined using model MinV, as follows:

$$\text{MinV } \text{Min } Z = \sum_{i=1}^G VIOLATION_i + \sum_{k=1}^T (0.50VIOLATIONA_k + 0.50VIOLATIONB_k) \quad (1)$$

$$\text{s.t. } TEAM_{W_i} - TEAM_{L_i} \geq M(VIOLATION_i) - S, \quad i = 1, \dots, G, \quad (2)$$

$$\begin{aligned} TEAM_{A_k} - TEAM_{B_k} &\geq DIFFA_k - S(VIOLATIONA_k) \\ S(VIOLATIONB_k) - DIFFB_k &\geq 0, \\ k &= 1, \dots, T, \end{aligned} \quad (3)$$

$$DIFFA_k - M(VIOLATIONA_k) \geq 0, \quad k = 1, \dots, T, \quad (4)$$

$$DIFFB_k - M(VIOLATIONB_k) \geq 0, \quad k = 1, \dots, T, \quad (5)$$

$$TEAM_j \geq 0, \quad j = 1, \dots, N, \quad (6)$$

$$DIFFA_k, DIFFB_k \geq 0, \quad k = 1, \dots, T, \quad (7)$$

$$VIOLATION_i \in \{0, 1\}, \quad i = 1, \dots, G, \quad (8)$$

$$VIOLATIONA_k, VIOLATIONB_k \in \{0, 1\}, \quad k = 1, \dots, T. \quad (9)$$

Equation (2) forces a violation if a game winner's value doesn't exceed the loser's. Equations (3), (4), and (5) force half a violation if teams in a tied game don't have identical values.  $TEAM_j$ ,  $DIFFA_k$ , and  $DIFFB_k$  differ depending on the chosen values of  $M$  and  $S$  (meaning that  $TEAM_j$  can't be interpreted directly), although the resulting ranking is unaffected. For each additional win by a team against the same opponent, one  $VIOLATION_i$  and one constraint (2) can be dropped, if the first win's  $VIOLATION_i$  coefficient in (1) is incremented by one.

The size of  $S$  affected solution efficiency but not predictably. MinV solved quickly using  $S = 1.00$  for all years except 2000, 2002, and 2003, where runs using  $S = 1.00$  and  $S = 0.10$  didn't solve within 10:00. The 2002 and 2003 models solved with  $S = 10.0$  within the pivots and times in Table 1. For 2000, MinV never solved for any attempted value of  $S$ , although it found a  $Z$  within one violation of the lower bound within 5:00. I established this solution's optimality by adding constraints to force undefeated Oklahoma's  $TEAM_j$  to exceed each of its four closest competitors'  $TEAM_j$  by  $S$  (such constraints for undefeated teams don't threaten optimality). MinV then solved within 568,856 pivots and 2:09.

For Table 3, if MinV's top five didn't match the consensus, I added constraints to force consensus number one's  $TEAM_j$  to exceed the second's by  $S$ , and similarly to force  $2 > 3 > 4 > 5 >$  other highly-ranked teams. I re-solved MinV and added more constraints for number five until MinV matched the consensus top five without increasing  $Z$  (implying that mini-

mizing violations wouldn't have changed the consensus) or  $Z$  increased. If the latter, then I identified the top five that most closely matched the consensus by first adding a few constraints to the original MinV to encourage the consensus first team to have the greatest  $TEAM_j$ . I re-solved MinV and added more constraints for the consensus first team until it had the highest  $TEAM_j$  or  $Z$  increased. If the latter, I repeated the process for the consensus second team (after dropping all constraints added for the first) and then for the consensus third team, and so forth until the team that had the best possible consensus rank had the best  $TEAM_j$  without increasing  $Z$ . Once I identified the top team, I repeated this iterative process (while keeping the constraints added in the previous step) to fill spots 2, 3, 4, and 5, using all remaining teams in order of their consensus ranking. For example, if the top spot was assigned to a team other than the consensus number one, then the consensus number one was the first team considered for the second spot.

In the process I used for Table 4, I focused only on prebowl games, matching the BCS rankings (not the consensus ranking) and analyzing just the top two spots.

### Acknowledgments

I thank the three referees, the associate editor, the editor-in-chief, and the managing editor for their extremely valuable contributions to this article, as well as Kenneth Massey, James Howell, Peter Wolfe, Eugene Potemkin, and Ray Waits for providing excellent data and feedback.

### References

Ali, I., W. D. Cook, M. Kress. 1986. On the minimum violations ranking of a tournament. *Mathematics* 32(6) 660–672.

Beck, T. 2004a. College football prediction tracker. Retrieved July 18, 2005 from [tbeck.freeshell.org/fb/predncaa.html](http://tbeck.freeshell.org/fb/predncaa.html).

Beck, T. 2004b. College football ratings PT awards. Retrieved June 30, 2004 from [tbeck.freeshell.org/fb/awards2003.html](http://tbeck.freeshell.org/fb/awards2003.html).

Beck, T. 2005a. NCAA prediction archive. Retrieved July 18, 2005 from [tbeck.freeshell.org/fb/ncaaarchive/ncaaarchive.html](http://tbeck.freeshell.org/fb/ncaaarchive/ncaaarchive.html).

Beck, T. 2005b. Season total final results. Retrieved July 18, 2005 from [tbeck.freeshell.org/fb/results.txt](http://tbeck.freeshell.org/fb/results.txt).

Bihl, J. 2001. On ranking violation percentage. Retrieved July 22, 2005 from [homepages.cae.wisc.edu/~dwilson/rsfc/rate/bihl.txt](http://homepages.cae.wisc.edu/~dwilson/rsfc/rate/bihl.txt).

Bihl, J. 2002. An interview with Jeff Bihl. Retrieved July 22, 2005 from [yourlinux.tripod.com/Sports/interviews/B/jbihl.html](http://yourlinux.tripod.com/Sports/interviews/B/jbihl.html).

Billingsley, R. 2002. In search of a national champion. Retrieved July 22, 2005 from [www.cffc.com/Archives/SearchRofRNC.htm](http://www.cffc.com/Archives/SearchRofRNC.htm).

BCSfootball.org. 2005a. BCS bowl eligibility. Retrieved July 28, 2005 from [www.bcsfootball.org/index.cfm?page=eligibility](http://www.bcsfootball.org/index.cfm?page=eligibility).

BCSfootball.org. 2005b. BCS bowl facts. Retrieved July 28, 2005 from [www.bcsfootball.org/index.cfm?page=facts](http://www.bcsfootball.org/index.cfm?page=facts).

BCS Media Guide. 2003. Retrieved July 9, 2004 from [www.bcsfootball.org/mediaguide.pdf](http://www.bcsfootball.org/mediaguide.pdf).

BCS Rankings Week 8. 2003. ABCSports.com. Retrieved July 9, 2004 from <http://sports.espn.go.com/ncf/abcsports/BCSRankingsWeek8>.

Callaghan, T., P. J. Mucha, M. A. Porter. 2004. The Bowl Championship Series: A mathematical review. Retrieved June 29, 2004 from [www.math.gatech.edu/~mucha/Reprints/bcsprop\\_side.pdf](http://www.math.gatech.edu/~mucha/Reprints/bcsprop_side.pdf).

Colley, W. N. 2002. Colley's bias free college football ranking system: The Colley matrix explained. Retrieved July 22, 2005 from [www.colleyrankings.com/matrate.pdf](http://www.colleyrankings.com/matrate.pdf).

Cook, W. D., M. Kress. 1990. An  $M$ th generation model for weak ranking of players in a tournament. *Operations Research* 41(12) 1111–1119.

Goddard, S. T. 1983. Ranking in tournaments and group decision making. *Mathematics* 29(12) 1384–1392.

Goddard, S. T. 1985. Reply on IRankings from round-robin tournaments. *Mathematics* 31(9) 1195.

Grotschel, M., M. Junger, G. Reinelt. 1984. A cutting plane algorithm for the linear ordering problem. *Operations Research* 32(6) 1195–1220.

Hopkins, M. 1997. Rating systems comparison. Retrieved July 22, 2005 from [homepages.cae.wisc.edu/~dwilson/rsfc/rate/hopkins2.txt](http://homepages.cae.wisc.edu/~dwilson/rsfc/rate/hopkins2.txt).

Howell, J. 2004. James Howell's college football scores. Retrieved October and November 2003 from [www.cae.wisc.edu/~dwilson/rsfc/history/howell/](http://www.cae.wisc.edu/~dwilson/rsfc/history/howell/).

Keener, J. P. 1993. The Perron-Frobenius theorem and the ranking of college football teams. *SIAM Review* 35(1) 80–93.

Kendall, M. 1962. *A Course of Combinatorics*, 3rd ed. Hafner, New York.

Laguna, M., R. Marti, V. Campos. 1999. Intensification and diversification with elite tabu search solutions for the linear ordering problem. *Operations Research* 26(12) 1217–1230.

LINDO Systems, Inc. 2003. *Global Optimization*. Chicago, IL.

Martinich, J. 2002. College football rankings: Do the computers know best? *Operations Research* 32(5) 85–94.

Massey, K. 2004a. About the ranking comparison. Retrieved June 29, 2004 from [www.mratings.com/cf/aboutcomp.htm](http://www.mratings.com/cf/aboutcomp.htm) key.

Massey, K. 2004b. College football ranking comparison. Retrieved June 29, 2004 from [www.mratings.com/cf/compare.htm](http://www.mratings.com/cf/compare.htm).

Massey, K. 2004c. College football ranking comparison. Retrieved June 30, 2004 from [www.mratings.com/cf/compareXXXX.htm](http://www.mratings.com/cf/compareXXXX.htm), where XXXX represents any year 1996 through 2002.

Massey, K. 2004d. College football ranking comparison. Retrieved July 19, 2005 from [www.mratings.com/cf/compare2004-15.htm](http://www.mratings.com/cf/compare2004-15.htm).

Massey, K. 2004e. Retrieved January 12, 2004 from [www.masseyratings.com/data/cf.gms](http://www.masseyratings.com/data/cf.gms).

Massey, K. 2005. College basketball ranking comparison. Retrieved July 21, 2005 from [www.mratings.com/cb/compare2005-18.htm](http://www.mratings.com/cb/compare2005-18.htm).

NCAA Online. 2004. Past Division 1-A football national champions. Retrieved June 30, 2004 from [www.ncaa.org/champadmin/iaFootballRpastRchamps.html](http://www.ncaa.org/champadmin/iaFootballRpastRchamps.html).

- Potemkin, E. 2005. Rank of rating systems. Retrieved July 21, 2005 from [rsport.netorn.ru/cf/rankrank.htm](http://rsport.netorn.ru/cf/rankrank.htm).
- Schrage, L. 1991. *Multi-Game Score Ratio System*. The Scientific Press, San Francisco, CA.
- Staples, A. 2005. Outback bowl, SEC remain partners. Retrieved September 26, 2005 from [outback-bowl-news.newslib.com/story/2217311/](http://outback-bowl-news.newslib.com/story/2217311/).
- Stob, M. 1985. Rankings from round-robin tournaments. *Mathematics* 31(9) 1191–1195.
- Trono, J. 2005. Predictive bowl comparison of 99 rating systems (2004). Retrieved July 18, 2005 from [academics.smcvt.edu/jtrono/OAFBBS/Systems2004.htm](http://academics.smcvt.edu/jtrono/OAFBBS/Systems2004.htm).
- Waits, R. 2005. NCAA football superlist. Retrieved July 18, 2005 from [homepages.cae.wisc.edu/~dwilson/superlist/](http://homepages.cae.wisc.edu/~dwilson/superlist/).
- Wilson, D. 2005a. American college football rankings. Retrieved July 21, 2005 from [www.nutshellsports.com/wilson/](http://www.nutshellsports.com/wilson/).
- Wilson, D. 2005b. Intention of college football ratings. Retrieved July 22, 2005 from [homepages.cae.wisc.edu/~dwilson/rsfc/rate/intention.html](http://homepages.cae.wisc.edu/~dwilson/rsfc/rate/intention.html).
- Wolfe, P. 2003. Personal communication.
- Wolfe, P. 2004. 2004 college football NCAA and NAIA scores. Retrieved periodically during the 2004 season from [www.bol.ucla.edu/~prwolfe/cfootball/scores.htm](http://www.bol.ucla.edu/~prwolfe/cfootball/scores.htm).