Pushing "Reset": The Conditional Effects of Coaching Replacements on College Football Performance*

E. Scott Adler, *University of Colorado, Boulder*

Michael J. Berry, *University of Colorado, Denver*

David Doherty, *Loyola University Chicago*

Objectives. We assess the effects of coaching replacements on college football team performance. *Methods.* Using data from 1997 to 2010, we use matching techniques to compare the performance of football programs that replaced their head coach to those where the coach was retained. The analysis has two major innovations over existing literature. First, we consider how entry conditions moderate the effects of coaching replacements. Second, we examine team performance for several years following the replacement to assess its effects. *Results.* We find that for particularly poorly performing teams, coach replacements have little effect on team performance as measured against comparable teams that did not replace their coach. However, for teams with middling records—that is, teams where entry conditions for a new coach appear to be more favorable—replacing the head coach appears to result in worse performance over subsequent years than comparable teams who retained their coach. *Conclusions.* The findings have important implications for our understanding of how entry conditions moderate the effects of leadership succession on team performance, and suggest that the relatively common decision to fire head college football coaches for poor team performance may be ill advised.

"We're in the era of PlayStation. If you don't like it, just hit 'reset.'"

Former University of Colorado football coach Dan Hawkins responding to a media question about the ease of replacing head football coaches

At the highest levels of competition, college athletics departments are extraordinarily dependent upon the revenue generated by their football programs to finance coaches, staff, facilities, and an array other athletic teams. Football team success affects television and bowl revenues, ticket and merchandise sales, and potentially alumnae contributions to athletics departments. When team performance is disappointing, athletics administrators and university officials are often pressed to find ways to reverse the trend. The prominent leadership role held by head coaches means that, fairly or unfairly, they are often blamed

*Direct correspondence to E. Scott Adler, Department of Political Science, UCB 333, Boulder, CO 80309-0333 ⟨E.Scott.Adler@Colorado.EDU⟩.

SOCIAL SCIENCE QUARTERLY, Volume 94, Number 1, March 2013

DOI: 10.1111/j.1540-6237.2012.00929.x

for poor records and are regularly fired or pressured to resign as a way of "fixing" the program. Over the last decade, on average approximately 1 in 10 teams annually replaced their coach for performance reasons.

The stakes involved in college football are high. A recent conservative estimate finds that, on average, football accounts for almost half of the generated revenue in Football Bowl Subdivision (FBS; formerly Division I-A) athletics departments (Fulks, 2009). Additionally, the costs of replacing a coach are typically substantial—often involving hundreds of thousands, if not millions of dollars in contract buyouts. However, there are no studies that empirically analyze the consequences of this particular variety of leadership succession. Little is known about whether replacing the coach is an effective strategy for improving performance.[1] To date, studies investigating leadership succession effects in sports focus almost exclusively on professional teams (for two exceptions, see Fizel and D'Itri, 1997, 1999). In this article, we present the first analysis of the effects of performance-based coaching replacements on the performance of college football teams.

Our examination of the understudied environment of college football has several innovations over previous research on leadership succession in sports. To start, we analyze a large numbers of teams—about 120 programs in the FBS. We employ matching analysis to compare teams that undergo a treatment (coach replacement) with similarly performing teams that do not. We also examine whether replacement effects are conditioned on whether team performance prior to the replacement was extremely poor or merely mediocre. This analysis allows for an assessment of the conditional effect of entry conditions on leadership succession and team performance. We find that while coaching replacements may provide a short-lived boost to performance among teams that have been performing particularly poorly, they have a deleterious effect on mediocre teams (those that won approximately 50 percent of their games in the year prior to the replacement). Our final innovation is that we track postsuccession performance over the five years following the replacement while all previous work—with the exception of Giambatista's (2004) study on professional basketball coaches—has only examined performance in the year immediately following the replacement.

The section that follows provides a brief survey of the research on sports leadership succession and describes a number of possible ways coaching replacement might be expected to affect team performance in college football. We then describe our data. Next, we present our empirical analysis, which reports our estimates of the effects of coaching replacements based on various matching techniques. We conclude with a discussion of the implications of our findings.

[1]Brown, Farrell, and Zorn examine how well the qualities of individual football coaches match the characteristics of NCAA programs, but do not assess the effects of coaching succession on team performance (2007).

Consequences of Leadership Replacement

The literature on leadership succession in organizations generally pursues one of two lines of inquiry—the causes of leadership succession or its effects on organizational performance (Kesner and Sebora, 1994; Giambatista et al., 2005). An area of research that has been particularly fruitful examines leadership changes in sports teams (Allen et al., 1979; Cannella and Rowe, 1995; Fabianic, 1994; Fizel and D'Itri, 1997, 1999; Gamson and Scotch, 1964; Giambatista, 2004; Pfeffer and Davis-Blake, 1986; McTeer et al., 1995; Rowe et al., 2005). There are a number of benefits to testing the impact of leadership succession on performance using sports teams. For one, unlike many other organizational scenarios, the leadership hierarchy is very clearly defined. Second, sports provide an objective and broadly agreed upon way to measure organizational performance—that is, wins and losses. Though there are other goals that can also be part of the performance objectives in the sports enterprise—for college athletics, these might be alumnae contributions, media attention for the university, or the ability to recruit higher quality student-athletes—these are generally considered ancillary to the primary objective of team performance. A final advantage of studying the effects of leadership in a sports environment is that all teams compete under uniform governing institutions regarding recruitment of actors and participation in activities that should lead to improved performance. College football is a particularly appealing context to test for coaching replacement effects as an average of nearly 20 FBS teams begin each fall with a new head coach.

Performance-based replacement of head college football coaches is dominated by the logic expressed by Army Athletic Director, Kevin Anderson, when he hired Rich Ellerson to replace fired head football coach Stan Brock in 2008: "One of our primary goals of the search was to find someone capable of turning around our program immediately and we are confident Rich is the perfect individual to accomplish that" (Associated Press [AP], 2008). That is, team underperformance is directly attributable to substandard coaching and leadership, and replacement of the coach should result in quicker recovery to satisfactory levels of achievement. In the organizational management literature, this approach is known as improved management or the "common sense" theory, and it asserts that the likely result of replacing failing leaders with what should be a better fitting manager is greater success in the future (Grusky, 1963; Gamson and Scotch, 1964; Huson et al., 2004). In the football context, this perspective would contend that schools that replace their head football coach due to poor performance will experience *better* team performance than similar teams that do not replace their coach. Previous research finds some support for this notion (Gamson and Scotch, 1964; Fabianic, 1994; McTeer et al., 1995).

However, another line of research contends that schools that replace their head football coach due to poor execution on the field will perform *worse* than similar teams that do not replace their coach. The notion is that firing a head coach may be more of a symbolic act and that fluctuations in performance are better attributed to factors out of the control of coaches: changes in the quality of opponents, loss of key players, or merely random shocks. Similar to the "vicious cycle" theory of managerial succession (Grusky, 1960, 1963; Brown, 1982), this perspective asserts further that coaching changes may disrupt the organization by altering the staff, procedures, coaching style, recruitment networks, and team culture. Ultimately, this may do more harm than good. Some empirical research also supports the contention that coaching replacements are, more often than not, detrimental to team performance (Fizel and D'Itri, 1999; Giambatista, 2004).

Some of the conflicting findings from previous research may be due to the fact that this work has not adequately captured basic differences among teams that replace their coach. One of the key contributions of our analysis is that we consider the possibility that the effects of coaching replacements are not homogeneous. Researchers exploring leadership success in other settings have asserted that entry conditions can be consequential for understanding the trajectory of unit performance over the course of their tenure (Hambrick and Fukutomi, 1991; Boal and Hooijberg, 2000; Pfeffer and Salancik, 2003). Thus, the effects of a coaching replacement may depend upon the context or environment faced by the new coach.

Improved management theory would contend that when an extremely poor performing team replaces its head coach, it is not likely to see much improvement because the team does not have the raw materials in the form of players, staff, and facilities to turn the program into a "winner." However, a mediocre team might have decent players. For these teams, a new coach may realistically be able to better manage these resources and, thus, improve team outcomes. In contrast, disruption theory might make the same assertion at the lowest end of the performance scale, but suggests the opposite outcome for better performing teams. That is, coaching replacements among extremely poor performing teams are unlikely to affect performance both because there is not much for the coach to work with *and* the team does not have far to fall. In other words, among these teams there may be little for the coach to "disrupt." However, when an average team replaces its coach, it risks disrupting the organization and harming performance. This leads us to two contrasting hypotheses:

Conditional Improved Management Hypothesis: *Teams that replace their head football coach due to extremely poor performance will perform similarly to teams that do not replace their coach. Teams that replace their head football coach due to mediocre performance will perform better than similar teams that do not replace their coach.*

Conditional Disruption Hypothesis: *Teams that replace their head football coach due to extremely poor performance will perform similarly to teams that do not replace their coach. Teams that replace their head football coach due to mediocre* *performance will perform worse than similar teams that do not replace their coach.*

Finally, Giambatista et al. (2005) note that "most of the sports setting studies used a one-year performance window, even though it is common knowledge among sports fans that most new high-profile coaches expect at least one year before transformation begins to produce results" (2005:966). This argument is particularly applicable to college athletics, as the team fielded by the newly hired head coach is largely comprised of the players recruited by the previous coach. It may take several years for the new coach to recruit a set of incoming players that better "fit" into his system. To account for this phenomenon, we examine the effects of coach replacement over the five years following the replacement.

Data Sources

To assess the effect of coaching replacement on team performance, we created a primary data set that includes information on all FBS football programs from a contemporary period: the 1997 through 2010 seasons. This 14-year time series captures the entire Bowl Championship Series (BCS) era, which began in 1998, up to 2010. Modeling each year since the inception of the BCS is important since this system constituted a major reformation of the structure of college football. Accordingly, the analysis that follows uses the team/year as the unit of analysis. There is an average of 117 universities fielding FBS football teams annually and a total of 1,643 cases in the data set. The overall win-loss record, conference win-loss record, and information about bowl appearances for each team was collected from a database of college football scores from 1985 to the present (Howell, 2009). To ensure accuracy, team records were checked by the authors against the records listed by ESPN.

The head coach for each team/year was identified using the *College Football Data Warehouse,* which facilitated identification of all instances where a coaching change took place—a total of 263 coaching changes occur in our data set, affecting football programs at 115 universities (DeLassus, 2012). Two coders identified the reason for each coaching change by entering the university and coach's name as well as year of replacement into Lexis-Nexis's search engine to find AP articles covering the replacement. Coaching changes were coded into five categories, which were then collapsed into replacements that were performance based and those that were not.[2] To ensure coding reliability, we randomly assigned nearly half of the 263 cases for examination by both coders. In almost all cases the categorization of individual coaching changes was straightforward (as indicated by the 94 percent rate of intercoder

[2]Nonperformance-based replacement categories (1) retired, (2) resigned to take a different coaching job, and (3) resigned for health reasons; performance-based categories (1) fired and (2) resigned amid pressure.

agreement). One hundred and fifty-five replacements were coded as performance based, with the remaining 108 coded as unrelated to performance. Summary data of coaching replacements by year are presented in Table A1 of the Appendix.

The rationales for the vast majority of replacements were clear. Most replacements unrelated to poor performance were either voluntary retirements (36 percent) or cases where a coach was hired away by another college or professional team (60 percent). There were also four cases (4 percent) where a coach left for health reasons. Voluntary retirements were typically easy to identify and most involved a coach leaving after at least eight seasons with the team. Reports of these replacements did not mention team performance as a reason for the coach's exit.

In most cases, identifying performance-based coaching changes was also straightforward. Seventy-nine percent of these replacements were clear cases where the coach was fired—a fact often mentioned in the headline of the article. The remaining performance-based replacements were cases where a coach left "under pressure." In most cases, this type of replacement was easy to identify as the report either explicitly mentioned performance as a reason for the replacement or noted that the university would continue paying the coach after his resignation. For example, an AP article discussing the resignation of John Thompson as the East Carolina University head coach reported: "East Carolina coach John Thompson, who resigned under pressure earlier this week, will be paid until Jan. 1, 2008, under the terms of a settlement with the school. Thompson agreed Wednesday to quit effective the end of the season after he was told he would be fired" (AP, 2004). Although Thompson technically resigned, the report made it clear that this resignation was a direct result of East Carolina's dissatisfaction with the team's poor performance and intent to fire him if he did not resign.

Assessing the Effects of Performance-Based Coaching Changes

Although teams that perform poorly are more likely to replace their coach, coaching changes are not deterministically driven by team performance. Each season many teams finish with losing records, but only some replace their coach. For example, of the 186 team/years in our data set where the team won fewer than 20 percent of its games, only 39 led to a performance-based coaching replacement. We take advantage of the fact that among similarly performing teams, some opt to replace their coach while others do not. This dynamic allows us to compare teams that were similar apart from the coaching replacement decision (or "treatment").

To this end, we conduct three types of analysis. Each approach uses propensity scores (predicted probabilities of a coaching change) as a way of identifying sets of teams that were similarly likely to replace their coach for performance-based reasons, but differed in that some replaced their coach while others did not. In the first approach—stratified matching—we conduct a series of *t*-tests to estimate the effects of a performance-based coaching change on subsequent team performance within blocks of teams that performed similarly during the previous season. Second, we conduct nearest neighbor matching. Here, we identify the most closely matching control case for each treated case and compare the performance among this smaller set of control cases—where entry conditions even more closely match those among the treated cases—with performance among the treated cases. In the third set of analyses, we also use propensity scores, but rather than matching cases, we use them as analytic weights in a series of ordinary least squares (OLS) regressions. This approach allows for a further examination of the possibility that the effect of performance-based coaching replacements depends on the entry conditions. In the next section, we describe the method we use to calculate our propensity scores and then present each of our analyses in turn.

Calculating the Propensity to Replace a Coach

Our proposed analysis requires that we identify sets of teams that were similarly likely to replace their coach for performance-based reasons. To identify similar teams, we adopt a timeline where T is the first year that we evaluate the effects of coaching replacement. For the treatment (coaching replacement) cases, this is the first year of the new coach's tenure. We compare treated cases with control cases that were similar at $T - 1$—that is, in the year that precipitated the replacement among the treated cases. Assuming that we have accounted for all of the factors that are likely to be correlated with performance *and* the decision to replace a coach, we can treat the decision to fire a coach as randomly assigned within these blocks of teams and assess the impact of this decision as though it were an experimental treatment (Rubin, 1973; Dehejia and Wahba, 1999).

Because the outcome of interest is team performance at time T, we must ensure that the treated and control teams we compare are similar at $T - 1$ on any factor that may affect performance at time T and also be correlated with coaching replacement. Past performance is likely to be the strongest predictor of future performance. This past performance captures a variety of relatively stable factors that may influence performance at time T, including institutional support for the team, reputation (which may improve recruiting capabilities), player quality, and other unidentified factors that may affect team success. We focus on four measures of absolute team performance: overall win percentage at $T - 1$, conference win percentage at $T - 1$, average win percentage from $T - 10$ to $T - 2$, and whether the team appeared in a bowl game at $T - 1$.[3] We also account for the possibility that the decision to expel a coach, as well as future performance, may be affected by whether a team meets

[3]We substitute overall win percentage for conference win percentage for the small number of independent teams (47 of the 1,205 team/years used in the analysis presented below).

performance expectations. For example, in 2004, the University of Wyoming and the University of Florida each won 58 percent of their games. However, over the previous four years Wyoming only won 19 percent of its games while Florida won 70 percent. While the two teams performed similarly in 2004 in absolute terms, accounting for historical performance, 2004 was a good year for Wyoming and a disappointing year for Florida. As such, Florida may be more likely to replace its coach than Wyoming. Importantly, if we expect that teams are likely to regress to their mean performance, we would expect Florida to outperform Wyoming in 2005. We calculated measures of the difference between overall win percentage (at $T - 1$) and average win percentage over the previous five years ($T - 6$ through $T - 2$). We also calculated a similar measure comparing performance to the average over the previous 10 years ($T - 11$ through $T - 2$). These measures capture how well teams performed relative to both recent and fairly long-term historical performance.[4] In order to ensure that treatment and control teams are as similar as possible, we also include a measure of how many years the coach had been in place at the beginning of the season.

We calculate propensity scores, which are estimates of the likelihood that a team will replace its coach in a given year, using a logistic regression model and regressing the dichotomous performance-based coaching replacement variable on the covariates discussed above using Becker and Ichino's (2002) *pscore* routine for STATA. The predicted values for each case generated from this model are the propensity scores. Among the control cases, these scores range from 0.001 to 0.658. Scores among the treated cases range from 0.022 to 0.470.[5] There is a great deal of overlap in the range of propensity scores in these two groups. However among control cases, there are a substantial number of propensity scores close to zero. Because there are not any treatment cases that performed similarly to these teams, we exclude these cases from our analysis and use only those control cases that fall within the range of scores observed among the treated cases (the region of common support).[6]

Columns 1 and 2 of Table 1 compare the mean values of the covariates used in the analysis for those teams that replaced their coach for performance reasons and those that did not, highlighting the differences in performance between

[4]A number of teams entered the FBS at a time that made it impossible to calculate average performance over the previous 5 or 10 years. In these cases, we used average performance over the team's full history in the FBS.

[5]The propensity score model is presented in Table A2 of the Appendix. Note that this model is a means of identifying propensity scores, which can be used to identify sets of treatment and control cases that are similar on the covariates used in this model. It is not intended to identify which factors are independently most predictive of coaching replacement. The distribution of propensity scores among treated and control cases is presented in Figure A1 in the Appendix.

[6]Specifically, there are 413 nontreated cases with propensity scores lower than the lowest propensity score among the treated cases. Among these control cases, the average win percentage at $T - 1$ was 76 percent. Almost 93 percent of these high-performing teams went to a bowl game in year $T - 1$. We also exclude the two untreated cases with propensity scores higher than the highest propensity score among the treated cases (Penn State in 2004 and 2005 has propensity scores of 0.658 and 0.598, respectively).

TABLE 1

Balance of Covariates in Treated and Nontreated Cases

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Full Sample		Common Support		Block 1		Block 2		Block 3		Block 4		Nearest Neighbor Matching	
	No Change	Change	No Change	Change	No Change	Change	No Change	Change	No Change	Change	No Change	Change	No Change	Change
Win percentage	0.528	0.331	0.436	0.331	0.593	0.571	0.456	0.448	0.300	0.296	0.152	0.155	0.334	0.331
($T-1$)	[0.2213]	[0.1669]	[0.1814]	[0.1669]	[0.0922]	[0.0949]	[0.1127]	[0.0959]	[0.114]	[0.1015]	[0.1153]	[0.1201]	[0.1769]	[0.1669]
Win percentage	0.505	0.462	0.484	0.462	0.527	0.564	0.479	0.464	0.444	0.442	0.446	0.451	0.456	0.462
($T-2$ to $T-10$)	[0.159]	[0.1531]	[0.1579]	[0.1531]	[0.1585]	[0.1807]	[0.1507]	[0.165]	[0.1519]	[0.1461]	[0.1506]	[0.124]	[0.1728]	[0.1531]
Win pct. differential	0.018	−0.112	−0.040	−0.112	0.063	−0.011	−0.025	0.006	−0.124	−0.134	−0.249	−0.248	−0.109	−0.112
(5 years; $T-1$)	[0.1953]	[0.1682]	[0.1782]	[0.1682]	[0.1463]	[0.11]	[0.1425]	[0.1498]	[0.1568]	[0.1423]	[0.144]	[0.1337]	[0.1773]	[0.1682]
Win pct. Differential	0.024	−0.136	−0.049	−0.136	0.068	0.010	−0.023	−0.021	−0.149	−0.149	−0.302	−0.305	−0.125	−0.136
(10 years; $T-1$)	[0.1993]	[0.1625]	[0.1708]	[0.1625]	[0.1285]	[0.127]	[0.1167]	[0.121]	[0.1297]	[0.1177]	[0.1144]	[0.1176]	[0.1683]	[0.1625]
Conf. win pct.	0.521	0.288	0.410	0.288	0.589	0.582	0.437	0.437	0.251	0.231	0.091	0.093	0.303	0.288
($T-1$)	[0.2567]	[0.1893]	[0.2065]	[0.1893]	[0.1207]	[0.0713]	[0.1197]	[0.0935]	[0.1212]	[0.1124]	[0.1172]	[0.1153]	[0.2089]	[0.1893]
Bowl appearance	0.498	0.155	0.330	0.155	0.720	0.625	0.204	0.325	0.027	0.000	0.011	0.028	0.153	0.155
($T-1$)	[0.5002]	[0.3629]	[0.4703]	[0.3629]	[0.4495]	[.5]	[0.4036]	[0.4743]	[0.1609]	[0]	[0.1054]	[0.1667]	[0.3611]	[0.3629]
Coach tenure	4.193	4.407	4.172	4.407	4.303	3.813	4.713	4.350	3.464	4.064	4.389	5.333	3.863	4.407
($T-1$)	[5.339]	[3.872]	[5.737]	[3.872]	[4.402]	[3.728]	[6.775]	[2.607]	[5.758]	[3.058]	[7.189]	[5.885]	[6.867]	[3.872]
Propensity score	0.087	0.178	0.115	0.178	0.039	0.042	0.089	0.093	0.179	0.187	0.308	0.319	0.165	0.178
	[0.0885]	[0.0995]	[0.0857]	[0.0995]	[0.0115]	[0.0124]	[0.0176]	[0.0185]	[0.0365]	[0.0388]	[0.0405]	[0.0562]	[0.0908]	[0.0995]
Observations	1,465	155	1,050	155	393	16	265	40	302	63	90	36	131	155

Cell entries are means. Standard deviations in brackets.

these groups and validating our coding of performance-based replacements. Columns 3 and 4 report means for each of the covariates restricting the sample to cases within the region of common support. The sizable differences in these measures observed in the full sample are reduced substantially among this restricted set of cases. However, some differences remain. We address these differences using both stratified and nearest neighbor matching.

Stratified Matching

After calculating propensity scores, the *pscore* routine uses a simple process to identify ranges of propensity scores—or blocks—where the mean propensity score among treatment and control cases is similar. This process yielded four blocks (propensity score ranges: Block 1: $0.022 \leq p < 0.062$; Block 2: $0.062 \leq p < 0.125$; Block 3: $0.125 \leq p < 0.250$; Block 4: $0.250 \leq p$). Note that these blocks delineate sets of teams where entry conditions varied from most favorable (Block 1) to least favorable (Block 4). After identifying blocks of cases, the routine tests for significant differences in covariate means between treatment and control cases within each block because balanced propensity scores do not guarantee balance on these underlying variables.[7] Summary statistics within each block are reported in columns 5–12 in Table 1. The table illustrates that the treated and control teams in each block performed remarkably similarly within each block.

This matching process allows us to confidently attribute any observed differences between control and treatment cases in performance at time T and beyond to the coaching change. At this point, we proceed with analysis much as we would in the context of a randomized experiment. We can also provide preliminary evidence regarding whether the effects of performance-based coaching replacement differs across entry conditions by comparing the treatment effect across blocks.

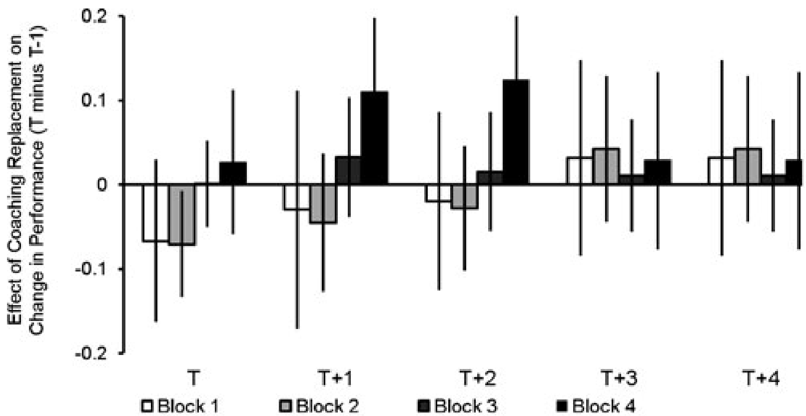
The entry conditions were most favorable for teams in Blocks 1 and 2. These teams performed best at $T - 1$, on average winning 59 and 45 percent of their games, respectively. In contrast, entry conditions in Block 3 were less favorable, and those in Block 4 were particularly unfavorable. The teams in Blocks 3 and 4 won only 30 and 15 percent of their games at $T - 1$. We focus our analysis here and below on the effects of coaching replacements on change in overall win percentage—how a team's performance changed from time $T - 1$ to time T. The findings are reported in Figure 1.

The bars in Figure 1 indicate the estimated difference in change in performance between treated and control teams in each block. Because there are a

[7] Given the number of statistical tests involved in checking for balance on all covariates (here, 28 tests, seven within each of four blocks), we applied a standard of balance of differences between treated and control cases not being significant at the $p < 0.01$ level. Our analysis found that this level of balance was achieved for all individual independent variables within each of the blocks. In fact, the only case where the more stringent standard of differences not being significant at the $p < 0.05$ level was violated was the five-year historical win percentage difference measure in Block 1.

FIGURE 1

Effects of Coaching Replacement on Change in Win Percentage (Stratified Matching)



Bars indicate difference in change in overall win percentage (T minus T–1) between teams that replaced their coach and those that did not (mean change among treated cases minus mean among control cases). Whiskers indicate 95 percent confidence intervals around difference. Because our data are truncated (at T=2010), the N is lower for estimates at T+N (number of cases: T=1205; T+1=1,117; T+2=1029; T+3=940; T+4=853).

limited number of treated cases in each block (ranging from 16 in Block 1 to 63 in Block 3), the standard errors (95% confidence intervals are indicated by the whiskers on each bar) on the estimates are quite large. Nonetheless, the pattern of estimated effects is telling. Among the teams with the most favorable entry conditions in Blocks 1 and 2, teams that replaced their coach won approximately 7 percent fewer games than those teams that did not replace their coach ($p = 0.19$ and 0.03, respectively). Additionally, contrary to claims that new coaches need a couple of years to rebuild a team, we find little evidence that Block 1 and 2 teams that replace their coach see a substantial improvement in performance after the new coach has had a season or two to rebuild. Instead, the pattern of effects suggests a regression to the mean where the negative impact of introducing a new coach diminishes over time for these teams. In contrast, there is some evidence that teams that replaced their coach with particularly unfavorable entry conditions (Block 4) saw larger improvements in performance at $T + 1$ and $T + 2$ than their counterparts who did not ($p < 0.05$ for each difference). However, this benefit evaporates at $T + 3$ and $T + 4$.[8]

The pattern of effects reported in Figure 1 suggests that the effects of coaching replacement depend on the entry conditions. In summary, the key findings are that coaching replacements, on average, appear to provide short-term benefits to teams that are performing extremely poorly. However, if anything, they have a deleterious effect on performance among teams where

[8]In additional analysis of the effects of coaching replacement on conference win percentage, we find a similar pattern of effects (see Appendix Figure A2).

entry conditions are most favorable. Importantly, this dispels the common rationale used by university athletic directors when firing the head coach, namely, that replacing the incumbent coach is a necessary step to improve on-field performance. Our findings demonstrate that the actual effects of such replacements are generally the opposite of what is intended.

Nearest Neighbor Matching

To check the robustness of these findings, we also conduct nearest neighbor matching. For each treated case, we identify the control case with the closest propensity score. Columns 11 and 12 of Table 1 report summary statistics for the cases used in this analysis and demonstrate that the treatment and control cases identified through this approach are nearly identical on all measures.[9] T-tests indicated that the performance-based coaching change treatment did not have an overall effect on subsequent win percentage. However, our expectations, as well as the findings presented in Figure 1, suggest that the effects of coaching replacements depend on the entry conditions a new coach faces. We examine this possibility using a series of OLS regressions, presented in Table 2.

In Table 2 column 1, we estimate the conditional effect of a coaching change on change in performance from $T - 1$ to time T by entering a linear interaction between the propensity score for a case and the treatment indicator. We also control for the covariates used in the propensity model to account for any remaining minor differences between treated and control cases. Consistent with the findings discussed above, the model indicates that among teams with more favorable entry conditions (those with propensity scores approaching zero), a coaching change leads to a decrease in change in wins between $T - 1$ and T of approximately 12 percent ($p < 0.01$). The positive, statistically significant coefficient on the interaction term ($p < 0.01$) indicates that this negative effect decreases among teams where entry conditions are less favorable (i.e., those with higher propensity scores).

In column 2, we estimate the conditional effect of coaching replacement using an interaction between the replacement treatment and indicators for each entry condition block. The negative coefficient on the coaching replacement variable indicates that among teams in Blocks 1, the estimated effect of a coaching change is an approximate 10 percent decrease in win percentage, relative to teams that did not replace their coach ($p = 0.052$). The coefficients on the interaction terms are positive, which indicates that the negative effect of coaching replacement among Block 1 teams is mitigated among teams with less favorable entry conditions. In columns 3–5 of Table 2, we estimate the conditional effect of coaching replacement on performance at $T + 1$, $T +$

[9]The difference in the number of treated and control cases stems from the fact that in some cases, the same control case is the closest match for more than one treated case.

TABLE 2

Effects of Coaching Replacement on Change in Win Percentage (Nearest Neighbor Matching)

	(1)	(2)	(3)	(4)	(5)
	Change in Win Percentage				
	(T minus $T-1$)	(T minus $T-1$)	($T+1$ minus $T-1$)	($T+2$ minus $T-1$)	($T+3$ minus $T-1$)
Performance-based replacement	−0.122	−0.102	−0.072	0.009	−0.063
(1 = yes)	[0.037][a]	[0.052]	[0.095]	[0.085]	[0.069]
Propensity score	−0.365				
	[0.551]				
Propensity score × PB change	0.599				
	[0.198][a]				
Block 2 (1 = yes)		−0.029	−0.132	−0.020	−0.116
		[0.058]	[0.072]	[0.083]	[0.077]
Block 3 (1 = yes)		−0.074	−0.169	−0.028	−0.100
		[0.088]	[0.103]	[0.101]	[0.100]
Block 4 (1 = yes)		−0.156	−0.296	−0.138	−0.209
		[0.123]	[0.129][b]	[0.139]	[0.144]
Block 2 × PB change		0.017	0.052	−0.020	0.117
		[0.072]	[0.101]	[0.106]	[0.077]
Block 3 × PB change		0.105	0.104	0.000	0.106
		[0.060]	[0.104]	[0.091]	[0.082]
Block 4 × PB change		0.165	0.183	0.119	0.125
		[0.068][b]	[0.105]	[0.099]	[0.100]
Win percentage ($T-1$)	−1.639	−1.578	−0.879	−0.639	−0.148
	[0.683][b]	[0.654][b]	[0.984]	[0.928]	[0.920]

TABLE 2—continued

	(1)	(2)	(3)	(4)	(5)
	Change in Win Percentage				
	(T minus T – 1)	(T minus T – 1)	(T + 1 minus T – 1)	(T + 2 minus T – 1)	(T + 3 minus T – 1)
Win percentage (T – 2 – T – 10)	1.367	1.261	0.461	0.264	–0.241
	[0.717]	[0.662]	[0.994]	[0.899]	[0.890]
Win pct. differential (5 years; T – 1)	0.152	0.210	0.209	0.046	0.013
	[0.201]	[0.165]	[0.255]	[0.234]	[0.248]
Win pct. differential (10 years; T – 1)	0.821	0.645	–0.262	–0.219	–0.976
	[0.814]	[0.672]	[0.947]	[0.836]	[0.906]
Conf. win percentage (T – 1)	0.008	–0.022	–0.309	–0.259	–0.205
	[0.152]	[0.140]	[0.148][b]	[0.172]	[0.184]
Bowl appearance (T – 1)	–0.005	0.000	0.029	0.023	0.021
	[0.036]	[0.040]	[0.044]	[0.054]	[0.060]
Coach tenure (T – 1)	0.001	0.002	0.003	–0.003	0.000
	[0.003]	[0.003]	[0.003]	[0.004]	[0.004]
Constant	0.176	0.201	0.404	0.303	0.318
	[0.145]	[0.112]	[0.118][a]	[0.112][a]	[0.128][b]
Observations	286	286	267	246	224
R-squared	0.249	0.254	0.350	0.354	0.394

Robust standard errors (clustered by team) in brackets.
[a]Significant at 1%.
[b]Significant at 5%.

2, and $T + 3$. The pattern of coefficients suggests that the negative effects of coaching replacements among teams with the most favorable entry conditions dissipate with the passage of time. Notably, the analysis in columns 3–5 shows no evidence that coaching replacements positively affect performance after an adjustment period.

Weighted Analysis

Finally, Barsky et al. (2002) propose an alternative approach to using propensity scores for this type of analysis. They suggest using OLS regression where control cases are weighted by the inverse of their propensity score and treatment case weights are set to one. In effect, this weighting scheme produces balance between treated and control groups on the measures used to calculate the propensity scores. This approach gives greater weight to control cases with higher propensity scores and lower weight to those with low propensity scores (i.e., those team/years where a coaching replacement neither occurred, nor was likely to occur).[10] Because we can include all of the cases across all blocks in this analysis (rather than analyzing cases within each block separately), this approach substantially improves estimate efficiency.[11]

In Table 3 columns 1–5, we replicate the models presented in Table 2 using all of the cases in the range of common support and the weighting approach proposed by Barsky et al. (2002), clustering standard errors by team. The findings are consistent with those presented and discussed above. As seen in column 1, for teams with propensity scores approaching zero, the decision to replace a coach leads to an almost 10 percent decrease in win percentage relative to the change in win percentage among similar teams that did not replace their coach ($p < 0.01$).

In column 2, we replace the linear interaction with a series of interactions between the coaching replacement indicator and indicators for Blocks 2–4. Again, the analysis indicates that among the teams in Block 1 where the entry conditions a new coach would face seem to be most favorable, the effect of coaching replacement is negative and statistically significant. Among these teams the model estimates that, on average, a coaching replacement results in an 8.5 percentage point decrease in win percentage ($p < 0.05$). The small coefficient on the interaction between the coaching replacement and Block 2 indicators indicates that the effect among teams in Block 2 was similar—a 7.3 percentage point decrease in win percentage ($p < 0.05$

[10] We demonstrate the effectiveness of this weighting approach in Table A3 of the Appendix. Specifically, we regress each of the covariates used in the propensity score model on the coaching change variable using these weights (and clustering standard errors by team). The statistically insignificant coefficient on the coaching change variable shows that these weights successfully balance the treated and control cases on these measures.

[11] For the sake of consistency, we restrict the sample to cases that fall within the range of common support.

TABLE 3

Effects of Coaching Replacement on Change in Win Percentage (Weighted Analysis)

	(1)	(2)	(3)	(4)	(5)
	Change in Win Percentage				
	(T minus $T-1$)	(T minus $T-1$)	($T+1$ minus $T-1$)	($T+2$ minus $T-1$)	($T+3$ minus $T-1$)
Performance-based replacement (1 = yes)	−0.099	−0.085	−0.063	−0.043	−0.063
	[0.034][a]	[0.041][b]	[0.066]	[0.057]	[0.051]
Propensity score	0.135				
	[0.378]				
Propensity score × PB change	0.438				
	[0.199][b]				
Block 2 (1 = yes)		−0.009	−0.044	−0.047	−0.060
		[0.032]	[0.035]	[0.039]	[0.042]
Block 3 (1 = yes)		−0.015	−0.098	−0.079	−0.094
		[0.056]	[0.062]	[0.066]	[0.068]
Block 4 (1 = yes)		−0.032	−0.202	−0.174	−0.177
		[0.087]	[0.089][b]	[0.094]	[0.109]
Block 2 × PB change		0.012	−0.002	0.009	0.054
		[0.051]	[0.079]	[0.070]	[0.061]
Block 3 × PB change		0.088	0.100	0.058	0.108
		[0.047]	[0.076]	[0.066]	[0.060]
Block 4 × PB change		0.112	0.164	0.156	0.111
		[0.058]	[0.076][b]	[0.072][b]	[0.078]
Win percentage ($T-1$)	−1.216	−1.043	−0.853	−0.409	−0.111
	[0.515][b]	[0.532]	[0.712]	[0.721]	[0.712]

TABLE 3—continued

	(1)	(2)	(3)	(4)	(5)
	Change in Win Percentage				
	(T minus $T-1$)	(T minus $T-1$)	($T+1$ minus $T-1$)	($T+2$ minus $T-1$)	($T+3$ minus $T-1$)
Win percentage ($T-2$–$T-10$)	1.018	0.788	0.424	−0.070	−0.398
	[0.509][b]	[0.523]	[0.704]	[0.697]	[0.682]
Win pct. differential (5 years; $T-1$)	0.044	0.141	0.226	−0.050	0.083
	[0.168]	[0.126]	[0.181]	[0.175]	[0.191]
Win pct. differential (10 years; $T-1$)	0.707	0.346	−0.304	−0.398	−1.027
	[0.585]	[0.543]	[0.708]	[0.684]	[0.706]
Conf. win percentage ($T-1$)	0.045	−0.015	−0.257	−0.262	−0.184
	[0.133]	[0.106]	[0.127][b]	[0.145]	[0.158]
Bowl appearance ($T-1$)	0.002	0.009	0.103	0.069	0.036
	[0.027]	[0.033]	[0.037][a]	[0.042]	[0.042]
Coach tenure ($T-1$)	−0.002	−0.001	0.000	−0.001	0.001
	[0.003]	[0.002]	[0.002]	[0.003]	[0.003]
Constant	0.089	0.149	0.333	0.377	0.364
	[0.098]	[0.073][b]	[0.080][a]	[0.079][a]	[0.088][a]
Observations	1,205	1,205	1,117	1,029	940
R-squared	0.234	0.235	0.346	0.344	0.372

Cell entries are OLS regression coefficients using weights as described in text. Robust standard errors (clustered by team) in brackets.
[a]Significant at 1%.
[b]Significant at 5%.

for linear combination of coefficients). The coefficients on the interactions between coaching replacement and Blocks 3 and 4 suggest that these negative effects are mitigated among teams that were performing particularly poorly before the coaching replacement ($p = 0.066$ and 0.057, respectively).

The pattern of findings in columns 3 through 5 are broadly consistent with those presented in Figure 1. The unfavorable effect of coaching replacements in situations where entry conditions appear to be most favorable diminishes gradually over time but does not reverse. Teams that performed particularly poorly at $T - 1$ (Block 4) appear to see a brief improvement in performance two and three years into the new coach's tenure relative to comparable teams that did not replace their coach. The linear combinations of the coefficient on the coaching replacement variable and the interaction between this variable and the Block 4 indicator are 0.101 ($p < 0.01$) and 0.113 ($p < 0.01$) in columns 3 and 4. However, this positive effect wanes the fourth year of the new coach's tenure (column 5; linear combination of coefficients = 0.049, $p = 0.431$).

Rolling the Dice?

One additional possibility is that poorly performing teams are playing a game of "high risk, high reward." That is, that coaching replacement may harm team performance on average, but in a small number of teams that are particularly fortunate, a new hire may improve performance substantially. In contrast, teams that do not replace their coach may find themselves in a predictable pattern of disappointing performance year after year. If this explanation is valid, it would be manifest in larger variance in performance among teams that replace their coach, a possibility we test by comparing variance in win percentage among control and treatment cases at time T. We make this comparison within the full region of common support as well as within each of the three blocks using Levene's (1960) test of the equality variances in two groups as well as the alternative formulations of this test proposed by Brown and Forsythe (1974). The p-values range from 0.096 to 0.987, indicating that variance in performance is similar for treated and controlled teams (full results presented in Appendix Table A4). Put simply, we find little support for the notion that performance-based coaching replacement is a high-risk/high-reward strategy.

Leadership Succession and College Football

This article addresses the question of whether leadership succession in underperforming college football programs is a good strategy for improving team performance. Through an examination of the performance of elite-level college football programs between 1997 and 2010, we demonstrate that the condition of the organization at the time of leadership replacement is critically important in determining the consequences of replacement. These findings support the conditional disruption hypothesis. Specifically, they demonstrate that coaching replacements do not immediately affect performance among teams where entry conditions are particularly unfavorable (though we find some evidence of a boost in performance in a new coach's second and third year among these teams). However, among teams where entry conditions appear to be most favorable, the choice to replace leadership is detrimental, rather than helpful, to team performance.

The innovations in this study offer new insights into our understanding of leadership succession. For instance, our work suggests an explanation for the mixed findings in previous research on leadership succession. Most prior work does not account for the possibility that the effects of changes in leadership are conditional. Our work demonstrates the need to consider organizational entry conditions as an important explanatory factor when examining the effects of leadership succession. By properly accounting for entry conditions at the time of leadership succession, scholars may be better able to estimate the effects of such changes and also gain theoretical traction for understanding how this type of organizational change affects performance.

While our results are robust to a number of different analytic approaches, like all research, the analysis presented here has limitations. Perhaps most importantly, although the various measures of past performance we use in our analysis are likely to account for the factors that most strongly affect both performance-based coaching replacements and subsequent performance, we are not able to account for (or even identify) every potential confound.

One possibility is that coaching replacements affect the quality of player recruitment that, in turn, affects performance. Similarly, fluctuations in the financial resources available to a team may be associated with both team performance and decisions about coaching replacements. We explore these possibilities in supplementary analysis (presented in Appendix Table A5). Specifically, we compared recruitment class ratings from 2002 to 2007 from Rivals.com and do not find any statistically significant differences in recruitment ratings at $T-1$ and T between treated and control teams within any block ($p > 0.10$ for all comparison). Similarly, examining logged team expenses at T and $T-1$ using data from 2001 to 2007 (Department of Education, Office of Postsecondary Education, 2009), we do not find any differences between treated and control teams within any blocks.

An additional possibility is that the quality of replacement coaches varies with the performance of the team. If this is the case, it seems reasonable to expect that teams in Block 1 would be able to attract higher quality coaches than the lower performing teams in the other three blocks. To the extent that this is the case, our findings may underestimate the actual difference in the effect of replacement between these classes of teams. A variety of other factors may also affect the likelihood of coaching replacement, including the amount of time left on a coach's contract, whether a university has just made a sizable expenditure hiring a coach in a different sport, and patterns of alumni giving. However, in order to bias our estimates of the effects of coaching replacement, these factors would also have to affect subsequent team performance independent of the performance measures we use in our analysis.

We also acknowledge that we cannot directly identify the mechanisms that drive the aggregate effects we observe. The effects we find may stem from the time it takes for a new coach to adapt to his new environment, the disruption caused by changes in on-field strategies, or a number of other factors. Examining these mechanisms is an important avenue for future research.

As with any statistical analysis, we cannot rule out the possibility that some specific instances of coaching replacements truly benefit a team. This is certainly a possibility and there is little doubt that many commentators, school administrators, and other observers believe that coaching changes are often responsible for turnarounds in team performance. However, it is important to bear in mind that the fact that a team's performance improves following a coaching replacement does not necessarily mean that the coach should be given credit for the improvement. Many teams that perform poorly one year improve the following year without replacing their coach. In fact, while 3 of the 16 teams in Block 1 (19%) that replaced their coach saw improvements in win percentage at time T, 158 of the 393 teams in this block that did not replace their coach (40%) improved. In summary, to the extent that some coaching changes are more effective than others, our evidence suggests that on average teams are either unable to determine whether the coach is responsible for unsatisfactory performance or do a poor job of selecting a replacement.

Our findings have important practical implications for the high stakes environment that is contemporary college football. When a college football team's performance is disappointing, the first and often only remedy administrators, fans, and sports writers turn to is firing the coach. This is usually an expensive approach to solving the problem.[12] In fact, the concern of sky-rocketing head coaching salaries was the key finding in a 2009 Knight Commission on Intercollegiate Athletics report based on interviews with 95 FBS university

[12]Fulks (2009) finds that coaching salaries and benefits constitute about one-third of total athletic department expenditures, with head football coaches comprising nearly one-third of all salaries. In 2009, the average head coaching salary for all FBS teams was more than \$1.3 million annually, which corresponds to a 46 percent increase over the prior three years (Wieberg and Upton, 2007; Wieberg et al., 2009). Moreover, in recent years, it has become common practice for poorly performing teams to fire the head coach prior to the expiration of his contract, forcing the university to "buyout" the remainder of the fired coach's contract often at considerable expense (Wieberg, 2008).

presidents (Knight Commission on Intercollegiate Athletics, 2009). Despite the fanfare that often accompanies the hiring of a new coach, our research demonstrates that at least with respect to on-field performance, coach replacement can be expected to be, at best, a break-even antidote. These findings, coupled with the significant costs universities typically incur by choosing to replace a head football coach, suggest that universities should be cautious in their decision to discharge their coach for performance reasons.

REFERENCES

Allen, Michael Patrick, Sharon K. Panian, and Roy E. Lotz. 1979. "Managerial Succession and Organizational Performance: A Recalcitrant Problem Revisited." *Administrative Science Quarterly* 24(2):167–80.

Associated Press. 2004. "Forced Out Coach to be Paid until Jan. 1, 2008." *Associated Press Online*.

———. 2008. "Army Hires Cal Poly's Ellerson as Football Coach." *Associated Press Online* December 26.

Barsky, Robert B., John Bound, Kerwin Charles, and J. P. Lupton. 2002. "Accounting for the Black-White Wealth Gap: A Nonparametric Approach." *Journal of the American Statistical Association* 97 (459):663–73.

Becker, Sascha, and Andrea Ichino. 2002. "Estimation of Average Treatment Effects Based on Propensity Scores." *Stata Journal* 2:358–77.

Boal, Kimberly B., and Robert Hooijberg. 2000. "Strategic Leadership Research: Moving On." *Leadership Quarterly* 11(4):515–49.

Brown, M. Craig. 1982. "Administrative Succession and Organizational Performance: The Succession Effect." *Administrative Science Quarterly* 27(1):1–16.

Brown, Morton B., and Alan B. Forsythe. 1974. "Robust Tests for the Equality of Variances." *Journal of the American Statistical Association* 69:364–67.

Brown, Todd, Kathleen A. Farrell, and Thomas Zorn. 2007. "Performance Measurement & Matching: The Market for Football Coaches." *Quarterly Journal of Business & Economics* 46(1):21–35.

Cannella, Albert A., and W. Glenn Rowe. 1995. "Leader Capabilities, Succession, and Competitive Context: A Study of Professional Baseball Teams." *Leadership Quarterly* 6(1):69–88.

Dehejia, Rajeev H., and Sadek Wahba. 1999. "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94:1053–62.

DeLassus, David. 2012. "College Football Data Warehouse." Available at ⟨http://www.cfbdatawarehouse.com/index.php⟩.

Department of Education, Office of Postsecondary Education. 2009. "Equity in Athletics Data Analysis." Available at ⟨http://ope.ed.gov/athletics/⟩.

Fabianic, David. 1994. "Managerial Change and Organization Effectiveness in Major League Baseball: Findings for the Eighties." *Journal of Sport Behavior* 17(3):135–48.

Fizel, J. L., and D'Itri, M. P. 1997. "Managerial Efficiency, Managerial Succession and Organizational Performance." *Managerial and Decision Economics* 18:295–308.

———. 1999. "Firing and Hiring of Managers: Does Efficiency Matter?" *Journal of Management* 25(4):567–85.

Fulks, Daniel L. 2009. *2004–08 NCAA Revenues and Expenses of Division I Intercollegiate Athletics Programs Report*. Indianapolis, IN: NCAA.

Gamson, William A., and Norman A. Scotch. 1964. "Scapegoating in Baseball." *American Journal of Sociology* 70(1):69–72.

Giambatista, Robert C. 2004. "Jumping Through Hoops: A Longitudinal Study of Leader Life Cycles in the NBA." *Leadership Quarterly* 15:607–24.

Giambatista, Robert C., W. Glenn Rowe, and Suhaib Riaz. 2005. "Nothing Succeeds Like Succession: A Critical Review of Leadership Succession Literature Since 1994." *Leadership Quarterly* 16:963–91.

Grusky, Oscar. 1960. "Administrative Succession in Formal Organizations." *Social Forces* 39:105–15.

———. 1963. "Managerial Succession and Organizational Effectiveness." *American Journal of Sociology* 69(1):21–31.

Hambrick, Donald C., and Gregory D. S. Fukutomi. 1991. "The Seasons of a CEO's Tenure." *Academy of Management Review* 16(4):719–42.

Howell, James. 2009. "Division I-A Historical Scores." Available at ⟨http://www.jhowell.net/cf/scores/ScoresIndex.htm⟩.

Huson, Mark R., Paul H. Malatesta, and Robert Parrino. 2004. "Managerial Succession and Firm Performance." *Journal of Financial Economics* 74(2):237–75.

Kesner, Idalene F., and Terrence C. Sebora. 1994. "Executive Succession: Past, Present & Future." *Journal of Management* 20(2):327–72.

Knight Commission on Intercollegiate Athletics. 2009. *Quantitative and Qualitative Research with Football Bowl Subdivision University Presidents on the Costs and Financing of Intercollegiate Athletics Report of Findings and Implications*. Baltimore, MD: Art & Science Group.

Levene, Howard. 1960. "Robust Tests for Equality of Variances." Pp. 278–92 in I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann, eds., *Contributions of Probability and Statistics: Essays in Honor of Harold Hotelling*. Menlo Park, CA: Stanford University Press.

McTeer, William, Phillip G. White, and Sheldon Persad. 1995. "Manager/Coach Mid-Season Replacement and Team Performance in Professional Team Sport." *Journal of Sport Behavior* 18(1):58–69.

Pfeffer, Jeffrey, and A. Davis-Blake. 1986. "Administrative Succession and Organizational Performance: How Administrator Experience Mediates the Succession Effect." *Academy of Management Journal* 29:72–83.

Pfeffer, Jeffrey, and Gerald Salancik. 2003. *The External Control of Organizations: A Resource Dependence Perspective*. Stanford, CA: Stanford University Press.

Rowe, Glenn W., Albert A. Canella, Debra Rankin, and Doug Gorman. 2005. "Leader Succession and Organizational Performance: Integrating the Common-Sense, Ritual Scapegoating, and Vicious-Circle Succession Theories." *Leadership Quarterly* 16:197–219.

Rubin, Donald B. 1973. "Matching to Remove Bias in Observational Studies." *Biometrics* 29:159–83.

Wieberg, Steve. 2008. "Huge Buyouts for College Coaches Causing Concern." *USA Today*. November 7.

Wieberg, Steve, and Jodi Upton. 2007. "College Football Coaches Calling Lucrative Plays." *USA Today*. December 5.

Wieberg, Steve, Jodi Upton, A. J. Perez, and Steve Berkowitz. 2009. "College Football Coaches See Salaries Rise in Down Economy." *USA Today*. November 7.

Appendix

FIGURE A1

Distribution of Propensity Scores

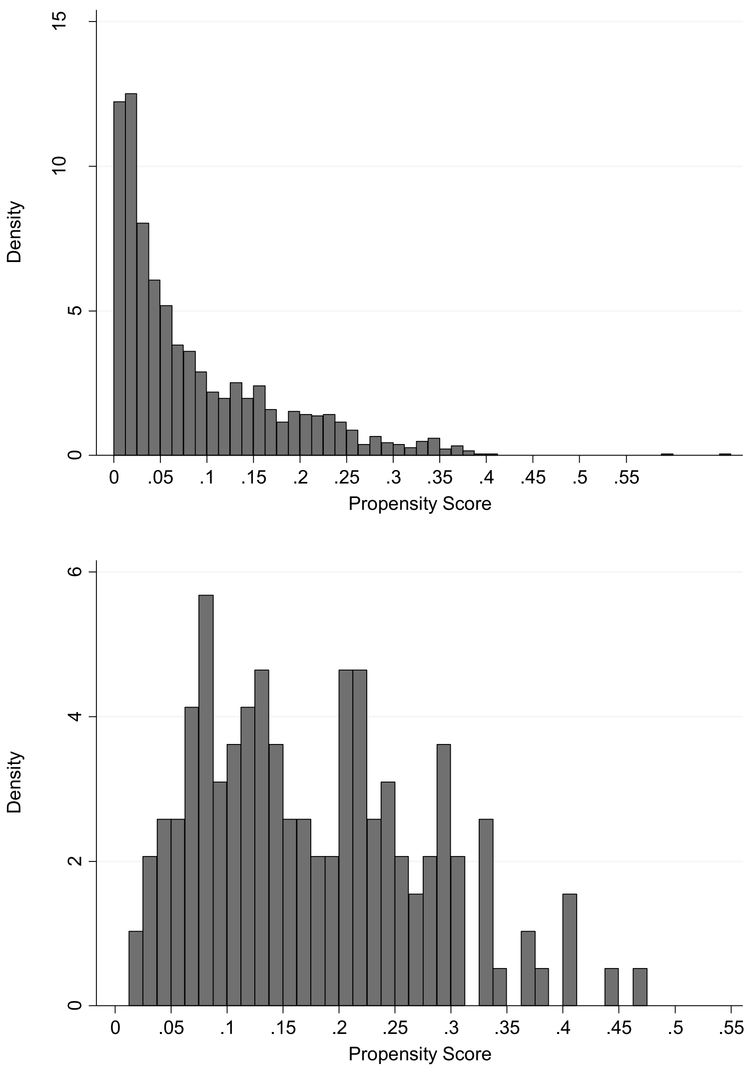
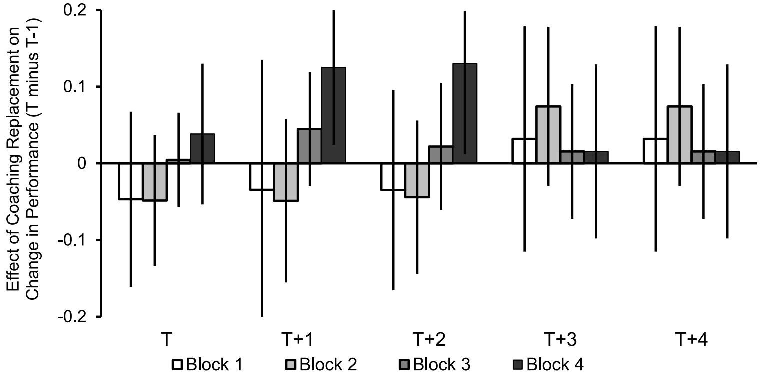


FIGURE A2

Effects of Coaching Replacement on Change in Conference Win Percentage (Stratified Matching)



Bars indicate difference in change in conference win percentage (T minus T-1) between teams that replaced their coach and those that did not (mean change among treated cases minus mean among control cases). Whiskers indicate 95 percent confidence intervals around difference. Because our data are truncated (at T=2010), the N is lower for estimates at T+N (number of cases: T=1205; T+1=1117; T+2=1029; T+3=940; T+4=853).

TABLE A1

Effects of Coaching Replacement (Stratified Matching)

Year	No. of FBS Teams	No. of FBS Teams Changing Coaches	No. of Performance-Based Coaching Changes	No. of Nonperformance-Based Coaching Changes	Percent of FBS Teams Changing Coaches
1997	112	24	11	13	21.4
1998	112	15	8	7	13.4
1999	114	20	9	11	17.5
2000	116	13	9	4	11.2
2001	117	25	17	8	21.4
2002	117	13	10	3	11.1
2003	117	19	13	6	16.2
2004	120	14	12	2	11.7
2005	119	22	15	8	18.5
2006	119	13	5	8	10.9
2007	120	23	10	13	19.2
2008	120	18	10	8	15.0
2009	120	22	15	7	18.3
2010	120	22	11	11	18.3
Total	1643	263	155	109	16.0

Data gathered by authors.

TABLE A2

Propensity Score Model

	Performance-Based Replacement (1 = yes)
Win percentage ($T-1$)	2.344
	[2.891]
Win percentage ($T-2-T-10$)	−2.767
	[2.815]
Win pct. differential (5 years; T −1)	1.244
	[0.732]
Win pct. differential (10 years; T −1)	−4.581
	[2.607]
Conf. win percentage ($T-1$)	−1.277
	[0.448][a]
Bowl appearance ($T-1$)	−0.194
	[0.152]
Coach tenure ($T-1$)	0.025
	[0.009][a]
Constant	−0.691
	[0.170][a]
Observations	1620
Log-likelihood	−436.994
Pseudo R-squared	0.145

Cell entries are coefficient estimates from probit model. Standard errors in brackets.
[a]Significant at 1%.

TABLE A3

Demonstration of Balance Using Weights

	(1) Win Percentage ($T-1$)	(1) Win Percentage ($T-2-T-10$)	(2) Win pct. Differential (5 year; $T-1$)	(3) Win Pct. Differential (10 year; $T-1$)	(4) Conf. Win Percentage ($T-1$)	(5) Bowl Appearance ($T-1$)	(6) Coach Tenure ($T-1$)
Performance-based replacement (1 = yes)	0.022	0.004	0.015	0.018	0.024	0.034	0.364
	[0.015]	[0.012]	[0.016]	[0.016]	[0.016]	[0.029]	[0.634]
Constant	0.309	0.458	–0.126	–0.153	0.264	0.121	4.043
	[0.013][a]	[0.015][a]	[0.008][a]	[0.009][a]	[0.014][a]	[0.017][a]	[0.660][a]
Observations	1,205	1,205	1,205	1,205	1,205	1,205	1,205
R-squared	0.004	0.000	0.002	0.003	0.004	0.002	0.001

Cell entries are OLS regression coefficients. Robust standard errors (clustered by team) in brackets.
[a]Significant at 1%.

TABLE A4

Effect of Coaching Replacement on Variance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Common Support		Block 1		Block 2		Block 3		Block 4	
	No PB Change	PB Change	No PB Change	PB Change	No PB Change	PB Change	No PB Change	PB Change	No PB Change	PB Change
Standard deviation (win percentage at T)	0.214	0.210	0.195	0.206	0.204	0.213	0.195	0.199	0.186	0.220
Measure of central tendency	Tests of the equality of variances (p)									
Mean	0.911		0.905		0.691		0.861		0.096	
Median	0.987		0.949		0.693		0.977		0.109	
10% trimmed mean	0.911		0.972		0.691		0.927		0.126	

Cell entries in first row are standard deviations of win percentage at time T. Other cell entries are p-values of tests of the equality of variance across groups.

TABLE A5

Recruitment Class Ratings (2002–2007) and Athletic Expenses (2001–2007)

Variable	Block 1		Block 2		Block 3		Block 4	
	No PB Change	PB Change	No PB Change	PB Change	No PB Change	PB Change	No PB Change	PB Change
Average Rival.com star ratings ($T-1$)	2.585	2.810	2.467	2.552	2.337	2.372	2.419	2.354
	[0.5155]	[0.6419]	[0.4724]	[0.5551]	[0.4748]	[0.3918]	[0.4571]	[0.3771]
Average Rival.com star ratings (T)	2.584	2.631	2.490	2.351	2.306	2.335	2.368	2.226
	[0.5337]	[0.5378]	[0.498]	[0.4074]	[0.4246]	[0.4254]	[0.4122]	[0.4055]
Football expenses (logged; $T-1$)	15.791	15.723	15.672	15.776	15.601	15.559	15.705	15.746
	[0.5979]	[0.6175]	[0.5408]	[0.6786]	[0.5836]	[0.5499]	[0.4748]	[0.4599]
Football expenses (logged; T)	15.784	15.854	15.679	15.709	15.519	15.587	15.637	15.816
	[0.5486]	[0.6792]	[0.5524]	[0.6917]	[0.5784]	[0.565]	[0.5071]	[0.553]
Observations	393	16	265	40	302	63	90	36

Standard deviations in brackets. No statistically significant differences in recruitment ratings or logged expenses at $T-1$ or T. Logged changes in expenses are significantly larger for treated groups in Block 3 ($p = 0.016$). Number of observations varies across rows.