Difficult Training Improves Team Performance:

An Empirical Case Study of US College Basketball

Heather Barry Kappes

London School of Economics and Political Science

Adam L. Alter

New York University

Griffin S. Edwards

University of Alabama at Birmingham

David J. Berri

Southern Utah University

Author Note

We are grateful to Gavin Kilduff, Chris To, and Severine Toussaert for their helpful comments on this research. Data and materials are available at <https://osf.io/cr8wx/>

Word count (excluding abstract and references): 7867

Address all correspondence to:

Heather Barry Kappes

Department of Management

London School of Economics and Political Science

London, UK WC2A 2AE

+44(0)20 7107 5418

Email: H.Kappes@lse.ac.uk

**Abstract**

One major challenge facing policymakers is to design education and workplace training programs that are appropriately challenging. We review previous research which suggests that difficult training is better than easy training. However, surveys we conducted of students and of expert sport coaches showed that many prescribed easy rather than difficult training for those they coached. We analyzed the performance of NCAA basketball teams in postseason tournaments to see whether the existing research, largely on individuals in short-term situations, would generalize to teams in the long run. Indeed, playing difficult nonconference (training) games modestly improved performance for NCAA teams in the postseason. Difficult training particularly benefitted teams that lost many nonconference games, and the effect of difficulty was positive within the range of difficulty NCAA teams actually encounter, making it clear that difficult training is superior. We suggest that our results can be generalized beyond sports, although with careful consideration of differences between NCAA basketball teams and other teams that may limit generalizability. These results suggest that policymakers might consider amplifying the difficulty of team training exercises under certain conditions.

Keywords: learning, training, practice effects, teams, self-versus-other

**Difficult Training Improves Team Performance**

Across many policy areas, practitioners need to design training that will best help people succeed. Governments and companies use education and training to build skills that improve performance and productivity. Maximizing the benefits of training requires identifying what kind of training builds skills most effectively. The world of work is increasingly one of teams (Deloitte, 2018), so the key policy question is what kind of training is most beneficial for team performance.

One crucial question about training is how difficult it should be. An analogy to medical vaccinations suggests that people should be exposed to milder trials (i.e., easy training) before they experience a full-blown challenge. Although early research suggested that this analogy holds for tasks that are completely novel or very challenging (Barch and Lewis, 1954), a larger body of evidence in the behavioral sciences suggests that people benefit more from particularly difficult training (Keith and Frese, 2008; Soderstrom and Bjork, 2015). However, this evidence largely stems from studies of individuals rather than teams, and looks at relatively short-term rather than longer-term outcomes.

To quantify the effects of difficult versus easy training on the real-world performance of teams over time—as opposed to performance on a short-term task or test that researchers have constructed—we looked to the domain of sports. Although sports require physical as well as intellectual skill, lessons from sports have been generalized to other settings including employees (Katz, 2001) and students (“Gridiron Grit,” 2017). We constructed a dataset consisting of all the nonconference (preseason), regular season, and conference games over a 10-year period in NCAA basketball. This data allowed us to observe the effects of difficult versus easy training on performance, in a team, over time.

**Difficult versus Easy Training**

We began our research by asking what expert practitioners believed about difficult versus easy training. Speaking to an audience of approximately 300 coaches and experts at the Australian Institute of Sport in November 2016, the second author asked them, “All else being equal, do you prefer easy training sessions or harder ones?” The example of archery was used to illustrate the idea (where the training target is closer—easier—or further away—harder—than the test target), while saying that the same distinction between easy and hard training would apply to other sports like basketball, soccer, etc. To avoid herding and social demand effects, audience members were asked to close their eyes and raise a hand to show their preference for easy versus difficult training. Approximately 85% of these experts (whose responses were counted by observers as well as the presenter) said they preferred easy training. Asked for clarification, coaches said they preferred the easy approach because they were concerned about exhausting or demotivating players. They only used hard training with very seasoned veterans and only infrequently.

The bulk of evidence from the behavioral sciences does not support these coaches’ preference for easy over difficult training. This evidence comes from multiple lines of research, dating back to early studies on transfer of training with Air Force personnel (Holding, 1962). Difficult tasks increase the likelihood that people will make errors, which facilitate learning because they provide informative feedback (Fisher and Lipson, 1986; Heimbeck et al., 2003), and encourage people to recruit additional cognitive and physical resources to overcome those errors (Alter et al., 2007). “Error management training” incorporates this observation and finds that encouraging participants to make many errors and to learn from them produces better performance on a later focal task, compared to training that prevents participants from making errors (Keith and Frese, 2008). By making errors, people improve in so-called *metacognition*, meaning that they get better at planning, monitoring, and evaluating their own efforts. Although errors provide useful feedback, seeing that one is performing poorly can be upsetting. During difficult training people gain opportunities to experience the emotions associated with making mistakes, and to practice regulating them. Thereby, enhanced emotion control that helps people persevere in the face of challenge may account for some of the downstream benefits of difficult training (Bandura, 1997; Duckworth et al., 2007). Indeed, recent research suggests that talented individuals who become top performers actually benefit from experiencing life stress, which helps with mental toughness and resilience (Collins and MacNamara, 2012); people in general report better mental health and well-being if they have experienced a moderate amount of lifetime adversity (Seery et al., 2010). Finally, consider that after easy training, the demands encountered during actual performance may seem overwhelming; those same demands may seem relatively low when compared to difficult training, thereby boosting confidence and enhancing success. This is the same mechanism musicians are drawing on when they practice playing a difficult piece at double speed, or when baseball players swing two bats as a warm up. These lines of research together make a compelling case that difficult training is superior to easy for helping people acquire new physical and intellectual skills (Bjork and Bjork, 2011; Soderstrom and Bjork, 2015). Indeed, this diverse research on training echoes the message from goal-setting literature which shows performance benefits of very difficult goals (Garland, 1983; Locke and Latham, 1990), including in team settings (Kleingeld et al., 2011).

Why, then, might expert coaches vote against difficult training? One possibility is that they recognize the benefits outlined above but believe these benefits are cancelled out by exhaustion or demotivation following difficulty. Given that people tend to interpret the way they perform during training as an indication of their competence (Soderstrom and Bjork, 2015), difficult training probably leads players to perceive themselves as less skilled, which may dampen motivation. Indeed, it may be difficult to assess the relative benefits of training regimes in the moment: difficult training, during which people perform poorly, may feel ineffective, while easy training, which produces a sense of fluency and is enjoyable, might be interpreted as effective (Sitzmann et al., 2010). The net result of any trade-off between learning (enhanced) versus motivation (decreased) following difficult training cannot be quantified by previous research, because most of the research studies showing benefits of difficulty observe only short-term outcomes. If difficulty takes a physical toll over time, or dampens motivation to the extent that people disengage, it may take time to see the (negative) effects on performance. That is, perhaps the coaches are right, and previous research findings shouldn’t be generalized to contexts where training is repeated over time and long-term outcomes are important. Moreover, one might ask whether difficult training becomes unhelpful in situations where success versus failure can be quantified (i.e., training games which have a winner and a loser), since failure could be discouraging and might further dampen motivation. If the benefits of difficulty are erased in situations where people fail (i.e., lose difficult training games), then again, coaches may be right in avoiding it.

A second possible explanation for coaches’ voting against difficult training is that perhaps only moderate difficulty helps performance, and more extreme difficulty would hurt. It may be that the research studies showing benefits of difficulty on performance have not tested conditions extreme enough to show that high difficulty can backfire. Indeed, very high levels of lifetime adversity are associated with worse mental health than the moderate levels which appear beneficial (Seery et al., 2010). If extreme difficulty does hurt, then coaches might be making the right decision to avoid difficult training, given that it could be challenging to know ahead of time where the line between moderate and extreme difficulty lies.

A third possibility is that the same difficult training which benefits individual performance, as demonstrated in the research reviewed above, would hurt team performance. Resilient teams, those with “the capacity to withstand and recover from challenges, pressure, or stressors,” are not necessarily composed of resilient individuals (Alliger et al., 2015, p. 176), meaning that even if difficult training makes individual athletes more resilient, it might not help a team. When working as a team, confidence in one’s teammates is crucial for success (Bloom et al., 2003; Cohen and Bailey, 1997), and team members’ beliefs that the team can collectively succeed are strong predictors of performance outcomes (Gully et al., 2002). Perhaps difficult training decreases confidence in teammates or in the team as a whole, thereby weakening team bonds; as noted above, confidence and efficacy may be particularly at risk in situations where teams fail (i.e. lose difficult training games). Meta-analysis finds that specific difficult goals (rather than nonspecific goals) improve group performance, but if individuals within the group have the goal of maximizing their individual performance—as may be the case for some individual elite athletes hoping to turn professional—group performance suffers (Kleingeld et al., 2011). We are not aware of previous research directly comparing effects of difficult versus easy training (which is different from difficult versus easy goals) on team performance, but if teams respond differently to difficult training than individuals, then coaches who lead teams would be correct in avoiding difficult training. This is an important question for policy makers who make decisions about the workplace because most employees today work as part of a team rather than strictly individually (Deloitte, 2018).

Whereas the three lines of thought outlined above are reasons that expert coaches might be correct in preferring easy to difficult training, it is also possible that the expert coaches we surveyed are simply mistaken. In this case, it would be important to know whether this is an error they make only when designing training for others, or whether they would be similarly incorrect if designing training for themselves. People sometimes prescribe less optimal actions to others than they would to themselves (Hsee and Weber, 1997; Zikmund-Fisher et al., 2006), and coaches might be overly influenced by players’ enjoyment of easy training. Perhaps coaches are actually recommending less effective (easier) training to their athletes than they would prescribe for themselves. To further explore the beliefs about the relative effectiveness of easy and difficult training, and to investigate whether recommendations are different when made in the role of coach, we conducted a survey experiment. We asked a sample of student respondents that comprised both inexperienced athletes and elite competitors for their recommendations about training (1) to meet goals of motivation versus preparation (2) in an individual versus team sport (3) for themselves versus for others (i.e., in the role of coach). We varied these three things—the former two within-respondents and the latter between-respondents—to see how these factors affected the recommended training.

**Survey Experiment**

**Methods**

*Participants and Design*

Participants were 173 members of an undergraduate subject pool at a large American university who answered these questions as part of a longer experimental session. Gender and age were not recorded, but participants were asked about their experience in both individual and team sports; 74% reported that they had competed in sports, including 32% at an elite level.

The study used a 2 (goal: motivation versus preparation) x 2 (sport: running/individual sport versus basketball/team sport) x 2 (recommendations: for self versus as coach) mixed design. The first two factors were manipulated within-subjects and the third factor was manipulated between-subjects. The data for this and the next study are available at <https://osf.io/cr8wx/>.

*Procedure and Measures*

All participants answered four questions, two about running (*individual sport condition*) and two about basketball (*team sport condition*). For each sport they were asked separately about what strategy would be most useful for motivating further training (*motivation goal condition*) and which strategy would be most useful for training to do well (*preparation goal condition*). Half of the participants (*n* = 86) were randomly assigned to make recommendations for themselves; the other half (*n* = 87) to make recommendations in the role of “coach.” We manipulated the target of the recommendations between rather than within-subjects to prevent participants from anchoring on one set of recommendations when making the other—that is, to give the best opportunity to observe whether people were inclined to recommend different training for themselves versus for those they were coaching.

Participants were instructed, regarding running: “Let's say you'd like to compete in a race in a few months' time…” (*recommendations for self condition*) or “Please imagine that you are a coach being asked to advise some athletes about how to prepare for upcoming competition. Let's say you're coaching a runner who would like to compete in a race in a few months' time…” (*recommendations as coach* *condition*). Participants were then asked: “Which of the following strategies will be most useful for motivating yourself [the runner] to train for the race?” and “Which of the following strategies will be most useful for training yourself [the runner] to run quickly in the future race?” For both questions, participants chose between “Going out and running an extremely difficult training run” “Going out and running a very easy training run” and “These are equally useful.” Questions about basketball followed a parallel format.

**Results**

We recoded responses so that choosing difficult training was coded as 1 and choosing easy training or saying the two strategies were equally useful was coded as 0,[[1]](#footnote-1) and analyzed the responses using a 2 (goal: motivation versus preparation) x 2 (sport: running/individual versus basketball/team) x 2 (recommendations: for self versus as coach) mixed design. We used binary logistic regression analysis with Generalized Estimating Equations to account for correlations between the multiple responses submitted by each individual. In our first model, the probability of choosing difficult training versus another outcome was predicted by the main effect of each factor, all 2-way interactions, and the 3-way interaction (goal x sport x recommendation target).

Three effects were significant at the *p* < .10 level in this model. First, there was a large effect of goal, Wald χ2(1) = 58.97, *p* < .001. As seen in Figure 1, participants were much more likely to recommend difficult training when the goal was to choose the training that would be best preparation for future performance, as opposed to motivating people to train further. Second, there was an effect of recommendation target, Wald χ2(1) = 2.90, *p* = .089. The probability of recommending difficult training was higher when making recommendations for self (.56) than when making recommendations as coach (.47). Finally, there was an interaction effect of sport x goal, Wald χ2(1) = 3.28, *p* = .07. As seen in Figure 1, participants were slightly more likely to choose difficult training as the more motivating, and slightly less likely to choose it as the more helpful, when making recommendations for the team sport (basketball) rather than the individual sport (running). The 3-way interaction (goal x sport x recommendation target) was not significant, Wald χ2(1) = .001, *p* = .98, and none of the other main effects or interaction effects in the model approached significance, *p*s > .38.

[Figure 1 about here]

A further analysis showed that, although respondents who had experience competing in sports at an elite level (approximately 32% of the respondents) were more likely to recommend difficult training than respondents without such experience (.58 versus .48), Wald χ2(1) = 3.24, *p* = .072, there was no evidence that elite experience affected how participants responded to the other manipulations. That is, none of the two-way interaction effects (elite experience by goal, by sport, or by recommendation target) were statistically significant, *p*s > .24.

**Discussion**

The results of the survey experiment speak to our introductory speculations as to why coaches might hesitate to recommend difficult training. First, participants strikingly preferred easy training for motivating athletes but difficult training for preparing them to perform well in the future; we had conjectured that this trade-off might be one reason that expert coaches chose easy over difficult training. It seems that people more generally, not only experts, suspect there is a trade-off between these two effects of difficult versus easy training. Second, we saw that, perhaps surprisingly, people were more likely to prescribe difficult training for themselves than in the role of coach. Although speculative, given the small and non-representative sample in our survey experiment, the difference between recommendations for self versus for others could explain why most expert coaches we surveyed earlier did not recommend difficult training to the athletes they coached. Lacking certainty about the athletes’ (in the survey question) level of commitment, perhaps coaches were too hesitant to use the method that they probably recognize (as our survey respondents did) produces superior performance, even if they would have been willing to use this method themselves.

Finally, we observed some differences in the recommendations that survey respondents made about running (an individual sport) versus basketball (a team sport). While acknowledging that there are other differences between these two sports which could also account for the differences, one interpretation is that some respondents believe difficult training is slightly less beneficial as preparation (while still being less motivating than easy training) in a team than an individual sport. If these beliefs are correct, then the behavioural sciences research which finds difficult training more effective would perhaps not be advisable to apply to teams. We turned to our NCAA basketball games data analysis to see if this is actually the case, in terms of real performance over time.

**NCAA Basketball Games Data**

To quantify the effects of difficult versus easy training on the real-world performance of teams, we constructed a dataset consisting of all the nonconference (preseason, which we treat as training), regular season, and conference games over a 10-year period in NCAA basketball. This dataset allowed us to observe the effects of difficult versus easy training on performance, in a team, over time.

In addition to the simple linear effect of training difficulty, we studied two further questions. First, we followed up the linear effect by testing for nonlinearity, to see whether the degree of difficulty—from moderate to extreme—matters. Second, we examined whether teams responded differently depending on whether they won or lost their nonconference (training) games. Given the secondary nature of the data, we are not able to directly observe the preceding effects, but our analyses do allow us to infer: whether difficult training produces superior performance because it a) simply sharpens skills (helping regardless of outcome); b) builds confidence only when teams win in the face of considerable difficulty (helping only when teams win); or c) builds coping skills in the face of failure (helping only when teams lose). The results of these two sets of further analyses, testing nonlinear effects and testing moderation by training game outcomes, allow us to refine our recommendations about easy versus difficult training for teams.

**Methods**

*Data*

We used a dataset with NCAA college basketball teams from the big five conferences for the ten seasons from 2003-2004 to 2012-2013 (N = 605). Those conferences include the ACC, Big 12, Big East, Pac 12 and SEC, and team performance is measured at the team by year level.

College basketball has nonconference games (like a preseason, which we treat as training games), a regular season (conference games), and a post-season. In the preseason, teams play nonconference games against teams that are not in their regional grouping, so results do not affect their conference ranking. In the post-season, each team plays a knock-out tournament against the other teams in its conference, and the strongest teams then go on to compete in the NCAA tournament. We mapped the questions outlined above onto this dataset by asking: to maximize tournament success, would it be better for a team to play nonconference games that are very difficult or very easy? Specifically, we tested for an effect of nonconference play on postseason performance to ascertain if a relatively strong (or weak) nonconference schedule had any statistical impact on post-season outcomes, and if so, whether that impact was positive or negative.[[2]](#footnote-2)

To capture the difficulty of each team’s nonconference (i.e. training) schedule, we rely on a measure known as Simple Rating System (SRS). SRS calculates a team’s quality by considering not only that team’s performance but also the performance of the teams it plays (Drinen, 2006). More specifically, it considers a team’s average margin of victory as well as the average margin of victory of the team’s opponents. Although the calculation of SRS requires one solve a system of equations, it is approximated by knowing a team's point differential and strength of schedule (SOS), or the combined winning percentages of the opponents the team played. For example, the 2012-13 Florida team that made it to the Elite Eight in the NCAA Tournament had a point differential (average margin of victory during the season) of 17.0. This team’s strength of schedule (SOS), or the combined winning percentages of the opponents Florida played, was 6.8, so Florida’s SRS score was 23.9 (there is some rounding).[[3]](#footnote-3) That same year, Auburn which lost the first game of the conference tournament and did not play in the NCAA tournament had point differential of -4.0 and an SOS of 4.2, so Auburn’s SRS score was 0.2. As Drinen (2006) notes, the interpretation of SRS is relatively easy: “If Team A's rating is 3 bigger than Team B's, this means that the system thinks Team A is 3 points better than team B.” Most schools in major conferences typically have an SRS greater than zero because their SOS are relatively high. It is not uncommon, though, for a mid-major school to have a negative SRS score.

To calculate the quality of the nonconference opponents, we take the average SRS score of the nonconference teams played. Average nonconference SRS scores in our sample ranged from -7 to +11.6 (*M* = 0.81, *SD* = 3.13), where higher scores indicate a more difficult pre-season schedule. Nonconference schedules are arranged by each school’s athletic director and coaches. While the games are not randomly assigned, they are often decided several years in advance. Since nonconference games are scheduled years in advance, long before the composition and strength of each team is known, the difficulty of nonconference games is largely exogenous.[[4]](#footnote-4) Additionally, unlike the NFL where players cannot compete until three years after they have graduated from high school, the NBA allows players to compete just one year after high school graduation. The turnover, particularly among star high school players, is therefore considerable, making it especially hard to predict the strength of any particular team even one year in advance thus making it difficult to try and tailor the nonconference schedule to the quality of the team.

To measure post-season success, we wanted to ensure all the relevant variation is exploited. To do this, we measured tournament success by counting the number of games played and adding one to the team that won the tournament. We do this separately for the knock-out tournament that each team plays against the other teams in their conference (i.e. conference tournament) and for the NCAA tournament. Historically the NCAA tournament consisted of 64 teams, giving the tournament 6 rounds of games. In 2001, though, a 65th team was added which gave the tournament a play-in game between two teams prior to the traditional 6 rounds. In 2011, three more teams were added, and the number of play-in games was increased to four. The teams selected for this tournament consist of the winners of 32 Division 1-A conferences. The remaining teams needed for the field are then selected by a committee. Ideally, these “at-large” teams are the “best” teams that failed to win an automatic conference bid.[[5]](#footnote-5) Our measure of games played in the tournament plus one for the winner allows us to identify all the teams that qualified for the tournament (which in the case of the NCAA tournament for some schools may, in it of itself, be considered a success) and still acknowledge the tournament winner.[[6]](#footnote-6)

To control for potential confounds, we extracted a range of variables from each team’s performance during its in-season games. Those variables were each team’s own regular season win/loss record and average point margin (see Pitts, 2016 for a similar approach), and team (or conference) fixed effects and season fixed effects. The team fixed effect variable is a dummy variable for each team. Each team’s fixed effect captures factors that influence performance and remain relatively stable across time, including brand power, loyalty of fan base, and quality of training facilities and recruiting and coaching staff. This set of dummy variables captures any idiosyncratic differences between the teams, including those mentioned above, that tend not to change over time. Similarly, certain model specifications include conference fixed effects in place of team fixed effects. Whereas team fixed effects absorb the time invariant idiosyncratic differences between teams, conference fixed effects account for the time invariant differences between across conferences. Conference fixed effects would be appropriate if there were reason to believe that were some sort of unobserved factor correlated with conferences that were driving the results—that is, if it were not so much the team unobservables that drive wins, but rather the unobserved quality/heritage/reputation of the conference. We report our main results with models that include both of these controls to show the insensitivity of the results to the level of fixed effects. Additionally, a season fixed effects variable—a dummy variable for each season—captures unobserved factors that are common to all teams and vary across time, including the tendency for players to improve as they become more experienced.

*Model Validity*

There are a couple of challenges to identification when trying to model the run a NCAA team makes in a tournament. Of primary concern is the nature of the dependent variable, which creates some challenges for OLS regressions. The data is discrete in nature—a team cannot win, for example, 1.23 games in either tournament. Additionally, the range of possible outcomes extends from zero to six or seven. Given the cutoffs at zero and seven, and the discreteness of the data, OLS is probably an inappropriate modeling technique. Since the data is discrete in nature and truncated at both ends, it mirrors a distribution more similar to a Poisson than a Normal, and so we estimate every specification using Poisson regression techniques with quasi-maximum likelihood (QML) estimation.[[7]](#footnote-7) Gourieroux, Monfort, and Trognon (1984) prove that the QML estimator is consistent and asymptotically normal when the Poisson model is correctly assumed, and Wooldridge (1999) shows that only correct specification of the condition mean is necessary for consistency. This allows the model to be distributionally misspecified, without regard to over or under dispersion, and still get consistent and efficient results. This method is especially useful in this study, as it allows for robust standard errors clustered at the state level (Wooldridge, 1999).

Another potential challenge to identification is the host of unobserved variables that may influence both our outcomes of interest and the difficulty of each team’s nonconference schedule (average opponent SRS). One advantage of the data in this domain, however, is that teams have very limited control over the difficulty of their nonconference schedules. As an example, there is very little correlation between a team’s previous year tournament performance and the next season’s non-conference strength of schedule. The unadjusted correlation between the previous year’s performance and next season’s average nonconference opponent SRS is 0.24 for the conference tournament and 0.34 for the NCAA tournament. Still, certain teams have a heritage of success in basketball, so to control for the possibility that stronger teams were more likely to construct difficult preseasons, each model includes team fixed effects to capture any non-observed, time-invariant factor that makes some teams more successful than others. The model proposed by Hausman, Hall, and Griliches (1984) suits this issue well as it allows for fixed effects.[[8]](#footnote-8)

*Model*

The baseline model we estimate is a Poisson fixed effects model which imposes a linearity assumption between SRS and tournament run. Generally, we estimate

(1)

where *r* is the count of the tournament run either in the conference tournament or NCAA tournament that varies by team, *I*, in season, *t*, *α* is the intercept, *X* is a vector of covariates that include regular season win-loss ratio and average point margin that vary by team and season, *F* is a matrix of either team or conference fixed effects (depending on the specification), *Y* is the matrix of season fixed effect and *e* is the error term. The independent variable of interest *s­it* is the measure of the average strength of schedule of the non-conference opposition for each team in each year as measured by SRS. In this approach, we allow *m(∙)* to take the form *exp(∙)* as proposed by Hausman and colleagues (1984), and use quasi-maximum likelihood (QML) estimation to estimate a fixed effects Poisson regression.

In addition to the specification in equation (1) we also relaxed certain assumptions to more directly address two behavioral questions. The first revised equation allows us to examine whether the benefits of difficult nonconference schedules diminish or reverse when the nonconference schedule is exceedingly difficult. To address this question, we revise equation (1) by including a quadric term in SRS.[[9]](#footnote-9)

(2)

We thus measure the degree to which there are any nonlinearities by performing a test of joint significant on the level and quadric terms of .

The second revised equation allows us to examine whether teams benefit from difficult nonconference schedules differently depending on whether they win or lose those preseason games. While beating a strong nonconference opponent may boost confidence, teams might also acquire technical and psychological coping skills when they lose difficult nonconference games. To examine this question, we interact the SRS variable with the demeaned preseason win percentage, or:

(3)

Where *nwr* is the nonconference win rate. In this specification, the marginal effect of *sit* at the mean nonconference win rate is . The coefficient measures the effect of playing tougher preseason games *and* winning those games. That is, if is indistinguishable from 0 then winning versus losing nonconference games doesn’t change the effect of preseason difficulty on postseason success, but if has a non-zero value then its direction is informative about whether winning versus losing magnifies the effect of nonconference difficulty. A positive value on suggests that the more successful teams are against nonconference opponents, the stronger the effect of preseason difficulty (i.e. SRS has a bigger effect when more games than average are won) and a negative value on suggests the opposite, that the less successful teams are in nonconference play, the stronger the effect of nonconference difficulty (i.e. SRS has a bigger effect when more games than average are lost). Alternatively stated, measures the returns to scheduling harder nonconference games for different types of teams. If is negative, then the value to scheduling tougher nonconference games is greater for teams that don’t start out the season as well compared to teams that win most their nonconference games.

Equation 3 represents a slight departure from our previous models, as we aim to measure directly the effect of winning vs. losing nonconference games. Whereas previously in Tables 1 through 3 we include as a control the team’s win ratio, to address how winning/losing nonconference games affects post-season performance, we include only the win ratio for nonconference games in Table 4 as part of the necessary controls to measure this interactive effect.

**Results**

The effect of a harder nonconference schedule is reported in Table 1 for the conference tournament and Table 2 for the NCAA tournament. Column (1) in each table reports the results with no controls or fixed effects. Column (2) includes controls and team and season fixed effects, and column (3) includes controls and fixed effects for conference and season. Every coefficient is reported as a marginal change and robust standard errors are in parentheses.

[Table 1 about here]

[Table 2 about here]

Across all specifications, both tables show a positive effect of nonconference opponent SRS on tournament run. It is important to note that this result is robust to the type of tournament. Though they are very different tournaments—every team gets into the conference tournament while only a select few in each conference make the NCAA tournament—we see that tougher nonconference competition yields a similar effect in both the NCAA and conference tournaments. That is, as teams play harder training games, they tend to make it farther in each respective postseason tournament. The coefficients reported in each regression are marginal effects that suggest how many additional games a team would play in each tournament if they increased their nonconference opponent SRS score by 1. For instance, in column (1) of Table 2, if a team increased its average opponent SRS score by 3, we would anticipate that team would advance an additional 3\*0.147=0.441 games in the NCA tournament. Keep in mind that these regressions control for team effects and the win/loss ratio for each team. Accordingly, the effect we estimate suggests what would happen for the same team with the same regular-season record which may, in part, explain why the effect is not larger.

We provide some specific examples to give a feeling for the magnitude of these effects. In the 2009-2010 season, all else being equal, Oregon (average pre-season SRS = -0.53, 33rd percentile) would have advanced half a game further in the NCAA tournament had they played a more difficult pre-season schedule similar to Arizona’s (average pre-season SRS = 7.09, 97th percentile). In the 2012-13 season, Georgetown entered the NCAA tournament as a number 2 seed and ranked number 5 in the AP poll. They were upset in the first round of the tournament losing to Florida Gulf Coast by 10 points. Georgetown’s preseason schedule consisted of mostly mid-major teams with weak SRS scores (average opponent SRS of -0.198, 38th percentile). If they had replaced half of their weaker nonconference games against mid-major teams and instead played Duke, VCU, Baylor, BYU, Alabama and Wisconsin, their average opponent SRS would have increased to 12.2, and we estimate their tournament run would have extended to the Sweet Sixteen instead of losing in the first round.

In contrast, the 2006-07 Tennessee Volunteers made it to the Sweet Sixteen before losing by one point to eventual runners-up Ohio State. That year, Tennessee played the second most challenging nonconference schedule in their conference with an average opponent SRS score of 3.96 (84th percentile and second to only Kentucky in the SEC that played nonconference games averaging a SRS score of 8.29). If Tennessee had, instead, played the nonconference schedule of cross state rival Vanderbilt (average opponent SRS=-0.84, 28th percentile), we estimate their tournament would have ended a game earlier losing to higher ranked Virginia.

In Table 3, we examine the nature of the nonlinear relationship between nonconference opponent quality and tournament run. Each column in Table 3 represents a distinct regression with a different outcome variable. Every regression includes controls and team and season fixed effects and reports the *p*-value of a joint test of significance on the level and squared average opponent SRS score. As shown in Table 3, the nature of the nonlinear relationship seems to be consistent across tournaments—a positive relationship increasing at a decreasing rate—though the relationship is statistically weaker in the conference tournament with a joint test p-value of 0.14, and while the squared term in the NCAA model has a large standard error, a test of joint significance between the level and squared term suggests that taken as a pair, they jointly matter to the model. That is, playing tougher teams is better, but at a decreasing rate. Said another way, the effect of playing increasing tougher nonconference competition levels off as the teams become increasingly strong.[[10]](#footnote-10) Based on the results in Table 3, we estimate that the optimal level of nonconference difficulty is a set of teams with an average SRS score of 18.8, beyond which point increasing difficulty begins to hurt. It is worth noting that no team in our dataset plays a nonconference schedule this difficult (the max is 11.6 played by Kansas in the 2004-05 season). Theoretically teams *could* schedule tougher nonconference games so their average opponent strength approaches a SRS score of 18 since typically, the top four or five teams in each major conference average a SRS score at or around 18, but in practice this is highly unlikely.[[11]](#footnote-11) Thus we can say that within the range of observed and likely nonconference schedules, increasing the difficulty of nonconference schedules helps.

[Table 3 about here]

In Table 4, we address the degree to which the outcome of those nonconference games matters. Table 4 is organized similarly to the previous table. The coefficient of particular interest in this table is the interaction between nonconference opponent strength and the nonconference win rate. Across all specifications, the coefficient is negative, suggesting that playing tougher teams and losing actually provides more benefit for both tournament runs than playing tougher teams and winning. Consistent with work such as error management theory and acquiring the skills of emotion control, this result indicates that the lessons learned from losing, rather than winning, tough early games to strong teams pay off in the long run. The combined results of the analyses presented in Tables 3 and 4 are approximately visualized in Figure 2.

[Table 4 about here]

[Figure 2 about here]

**General Discussion**

In brief, what we found in this archival data allows us to suggest that skilled performers working in teams should use difficult training. The benefits this training provides enhance performance over time. Whereas coaches—and those like managers who make decisions about training teams of people other than themselves—might be inclined to err on the side of caution by prescribing easy training, doing so predicts worse team performance in the long-term.

Earlier in the paper, we identified four possible reasons why expert coaches might prefer easy training. The first three were reasons why these coaches might be correct in their preference for easy training; our results argue against these three explanations. First, we speculated that over time—over a longer period than has been examined in previous research—the benefits of difficulty might be counteracted by exhaustion or demotivation. Observing performance over an entire season, we find no support for that explanation; any exhaustion or demotivation is more than compensated for, since there is a positive effect of difficult nonconference games on postseason performance. Also relevant to this explanation, we examined whether teams responded differently depending on whether they won or lost their nonconference (training) games. In fact, we found that difficult nonconference (preseason) games predicted better tournament performance in general, and this effect was most pronounced among teams that lost many of their nonconference games. Thus, a fear that failure (i.e., losing games), perhaps by discouraging players or working against team cohesion, would over the long-term erase the benefits of difficult training, is also unfounded. Competitors are perhaps more resilient than coaches assume.

Second, we considered the possibility that only moderate difficulty, as opposed to great difficulty, helps performance. To address this explanation, we tested a nonlinear effect of difficulty to see whether difficulty above a certain level begins to hurt performance. This was not the case; the benefits of difficulty level off but do not backfire until extremely high levels of difficulty—levels above what even the most challenged teams in our dataset play. Thus, the desire to avoid extreme difficulty would not be a valid reason for coaches to prefer easy to difficult training.

Third, we considered that perhaps since previous research mainly examines effects of difficult training on individuals, difficulty might not have the same benefit for teams. Although we cannot say whether effects on teams are smaller than on individuals, our data shows a clear benefit to difficult training for teams, speaking against this explanation. Indeed, one might have wondered whether losing difficult training games hurts later team performance by making it harder for players to trust their teammates—in line with this speculation we observed in our survey experiment a slight drop in recommendations for difficult training for a team sport compared to an individual one. However, as noted above, losing difficult training games actually amplified the benefits of difficult training.

On the whole, our results speak in favor of difficult training for teams, when optimal performance over time is the objective. These results add to the body of research in behavioural sciences, mostly with individuals and short-term outcomes, showing the benefits of difficult training. However, it is worth considering some limitations to the generalizability of our results. One strength of this research is that pre-season game difficulty and postseason performance are objective indicators. However, difficult training games are not necessarily synonymous with difficult training in general (e.g., many hours of heavy weight-lifting), and we cannot speak to the relative effects of other sorts of difficulty in training. It is feasible that difficult pre-season games help in part by giving players a reason to try, which might otherwise be missing in this low-stakes setting; while consistent with the arguments we have outlined about why difficulty helps, this mechanism might mean that other forms of difficult practice do not show the same effects. Moreover, NCAA seasons have limited time frames; it is possible that difficult training would be less beneficial for teams that perform in short or seasonal bursts, whereas employees might be at heightened risk of burnout when faced with protracted extreme challenges (Sitkin et al., 2011). It is also possible that teams without expert coaches, where the time for feedback is limited, or where players know that coaches themselves set the difficulty of the training schedule, would not show the same benefits of difficulty. Without further research to test effects of difficulty in teams under these different variations, one must still exercise caution in designing training, even while recognizing the benefits of difficult training regimens.

**Conclusion**

Our investigation was motivated by the observation that although research in the behavioral sciences shows that difficult training is superior, this evidence does not appear to have been fully incorporated into practice. In the domain of sports, most of the expert coaches we surveyed preferred easy to difficult training. Although sports require physical as well as intellectual skill, lessons from sports have been generalized to groups including employees (Katz, 2001) and students (“Gridiron Grit,” 2017), where optimal training allows for reaping the maximum benefits from education. To what extent should our current results be generalized outside of sports? We believe rather broadly; we have no reason to predict that results would differ for teams that rely on intellectual rather than physical performance. According to large meta-analyses, practice effects in intellectual domains tend to be moderate in size (Hausknecht et al., 2007). Even if the overall effect of training is small, we still expect that difficult training helps performance—intellectual as well as physical—more than easy training.

Across various policy areas, such as in education and in the workplace, practitioners need to design training that will best help people succeed. As the nature of work changes, effective policy can help ensure that teams of workers acquire the skills they need, and remain resilient in the face of challenge (e.g., Van der Hoek et al., 2018). Previous research in the behavioral sciences, reviewed at the start of this paper, suggested that difficult training is the most useful, but we suspect that this evidence has not yet been fully incorporated into practice (although we have evidence for this gap only in sports). Focusing on the domain of sports, we identify several reasons why this may be the case (e.g. long-term effects, team versus individual performance, the possibility of extreme difficulty backfiring), reject these possibilities following analysis of archival data, and show that difficult training helps teams succeed. Interestingly, our survey experiment, with university student participants, suggests that most people recognize the merits of difficult training for performance—though they may overestimate its dampening effects on motivation—but hesitate to prescribe it to others to the same extent as they would do for themselves. Thus, whereas other research in behavioral science might close by suggesting that people try to gain distance from decisions (e.g. imagine giving advice to someone else rather than deciding for oneself), when it comes to training, we recommend the opposite. Policy makers should not hesitate to select the same difficult training for others as they would willingly use for themselves.

**References**

Alliger, G.M., Cerasoli, C.P., Tannenbaum, S.I., Vessey, W.B., 2015. Team resilience. Organizational Dynamics 44, 176–184.

Alter, A.L., Oppenheimer, D.M., Epley, N., Eyre, R.N., 2007. Overcoming intuition: metacognitive difficulty activates analytic reasoning. Journal of Experimental Psychology: General 136, 569.

Bandura, A., 1997. Self-efficacy: The exercise of control. Macmillan.

Barch, A.M., Lewis, D., 1954. The effect of task difficulty and amount of practice on proactive transfer. Journal of experimental psychology 48, 134.

Bjork, E.L., Bjork, R., 2011. Making things hard on yourself, but in a good way. Psychology in the Real World 59–68.

Bloom, G., Stevens, D., Wickwire, T., 2003. Expert Coaches’ Perceptions of Team Building. Journal of Applied Sport Psychology 15, 129–143. https://doi.org/10.1080/10413200305397

Cohen, S.G., Bailey, D.E., 1997. What makes teams work: Group effectiveness research from the shop floor to the executive suite. Journal of management 23, 239–290.

Collins, D., MacNamara, Á., 2012. The rocky road to the top. Sports medicine 42, 907–914.

Deloitte, 2018. 2018 Global Human Capital Trends.

Drinen, D., 2006. A very simple ranking system » Pro-football-reference.com blog » Blog Archive [WWW Document]. URL https://www.pro-football-reference.com/blog/index4837.html?p=37 (accessed 8.4.18).

Duckworth, A.L., Peterson, C., Matthews, M.D., Kelly, D.R., 2007. Grit: perseverance and passion for long-term goals. Journal of personality and social psychology 92, 1087.

Fisher, K.M., Lipson, J.I., 1986. Twenty questions about student errors. Journal of Research in Science Teaching 23, 783–803.

Garland, H., 1983. Influence of ability, assigned goals, and normative information on personal goals and performance: A challenge to the goal attainability assumption. Journal of Applied Psychology 68, 20.

Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: Theory. Econometrica: Journal of the Econometric Society 681–700.

Gridiron Grit: Angela Duckworth on Excelling in the NFL and Beyond, 2017. . Behavioral Scientist. URL http://behavioralscientist.org/gridiron-grit-angela-duckworth-excelling-nfl-beyond/ (accessed 8.3.18).

Gully, S.M., Incalcaterra, K.A., Joshi, A., Beaubien, J.M., 2002. A meta-analysis of team-efficacy, potency, and performance: interdependence and level of analysis as moderators of observed relationships. Journal of applied psychology 87, 819.

Hausknecht, J.P., Halpert, J.A., Di Paolo, N.T., Moriarty Gerrard, M.O., 2007. Retesting in selection: a meta-analysis of coaching and practice effects for tests of cognitive ability. Journal of Applied Psychology 92, 373.

Hausman, J.A., Hall, B.H., Griliches, Z., 1984. Econometric models for count data with an application to the patents-R&D relationship. national bureau of economic research Cambridge, Mass., USA.

Heimbeck, D., Frese, M., Sonnentag, S., Keith, N., 2003. Integrating errors into the training process: The function of error management instructions and the role of goal orientation. Personnel Psychology 56, 333–361.

Holding, D.H., 1962. Transfer between difficult and easy tasks. British Journal of Psychology 53, 397–407.

Hsee, C.K., Weber, E.U., 1997. A fundamental prediction error: Self–others discrepancies in risk preference. Journal of Experimental Psychology: General 126, 45–53. https://doi.org/10.1037/0096-3445.126.1.45

Katz, N., 2001. Sports teams as a model for workplace teams: Lessons and liabilities. The Academy of Management Executive 15, 56–67.

Keith, N., Frese, M., 2008. Effectiveness of error management training: a meta-analysis. Journal of Applied Psychology 93, 59.

Kleingeld, A., van Mierlo, H., Arends, L., 2011. The effect of goal setting on group performance: A meta-analysis. Journal of Applied Psychology 96, 1289.

Locke, E.A., Latham, G.P., 1990. A theory of goal setting & task performance. Prentice-Hall, Inc.

Pitts, J.D., 2016. Determinants of Success in the National Football League’s Postseason: How Important Is Previous Playoff Experience? Journal of Sports Economics 17, 86–111.

Seery, M.D., Holman, E.A., Silver, R.C., 2010. Whatever does not kill us: cumulative lifetime adversity, vulnerability, and resilience. Journal of personality and social psychology 99, 1025.

Sitkin, S.B., See, K.E., Miller, C.C., Lawless, M.W., Carton, A.M., 2011. The paradox of stretch goals: Organizations in pursuit of the seemingly impossible. Academy of Management Review 36, 544–566.

Sitzmann, T., Ely, K., Brown, K.G., Bauer, K.N., 2010. Self-assessment of knowledge: A cognitive learning or affective measure? Academy of Management Learning & Education 9, 169–191.

Soderstrom, N.C., Bjork, R.A., 2015. Learning versus performance: An integrative review. Perspectives on Psychological Science 10, 176–199.

Van der Hoek, M., Groeneveld, S., Kuipers, B., 2018. Goal setting in teams: Goal clarity and team performance in the public sector. Review of public personnel administration 38, 472–493.

Wooldridge, J.M., 1999. Distribution-free estimation of some nonlinear panel data models. Journal of Econometrics 90, 77–97.

Zikmund-Fisher, B.J., Sarr, B., Fagerlin, A., Ubel, P.A., 2006. A matter of perspective: choosing for others differs from choosing for yourself in making treatment decisions. Journal of General Internal Medicine 21, 618–622.

|  |  |  |  |
| --- | --- | --- | --- |
| Table 1: Nonconference opponent quality and conference tourney run | | | |
|  | (1) | (2) | (3) |
| Average Opponent SRS | 0.057‡ | 0.024 | 0.020^ |
|  | (0.013) | (0.015) | (0.011) |
| *Fixed Effects* | | |  |
| Team | | X |  |
| Conference | | | X |
| Season | | X | X |
| Observations | 605 | 605 | 605 |
| (Pseudo) R Squared | 0.01 | 0.06 | 0.05 |
| Notes: Each observation is measured at the team/year level. The reported coefficients are marginal effects, and robust standard errors are reported in parentheses. Each regression includes controls for the win-loss ratio, and point margin. The outcome is measured in number of tournament games played plus one extra for the tournament winner.  ^ p<0.10 † p<0.05 ‡ p<0.01 | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| Table 2: Nonconference opponent quality and NCAA tourney run | | | |
|  | (1) | (2) | (3) |
| Average Opponent SRS | 0.147‡ | 0.077‡ | 0.080‡ |
|  | (0.020) | (0.020) | (0.016) |
| *Fixed Effects* |  |  |  |
| Team |  | X |  |
| Conference |  |  | X |
| Season |  | X | X |
| Observations | 605 | 605 | 605 |
| (Pseudo) R Squared | 0.06 | 0.42 | 0.38 |
| Notes: Each observation is measured at the team/year level. The reported coefficients are marginal effects, and robust standard errors are reported in parentheses. Each regression includes controls for the win-loss ratio and point margin. The outcome is measured in number of tournament games played plus one extra for the tournament winner.  ^ p<0.10 † p<0.05 ‡ p<0.01 | | | |

|  |  |  |
| --- | --- | --- |
| Table 3: Nonlinear relationship between average opponent SRS and tourney run | | |
|  | C Tourney + | NCAA + |
|  | (1) | (2) |
| Average Opponent SRS | 0.032^ | 0.086‡ |
|  | (0.017) | (0.030) |
| Average Opponent SRS Squared | -0.004 | -0.002 |
|  | (0.003) | (0.004) |
| P value of joint significance | 0.14 | 0.00 |
| Observations | 605 | 605 |
| (Pseudo) R Squared | 0.06 | 0.42 |
| Notes: Each observation is measured at the team/year level. The reported coefficients are marginal effects, and robust standard errors are reported in parentheses. The outcome is measured in number of tournament games played plus one extra for the tournament winner. Each regression includes team and season fixed effects, and a control for the win-loss ratio, and point margins. ^ p<0.10 † p<0.05 ‡ p<0.01 | | |

|  |  |  |
| --- | --- | --- |
| Table 4: Interactive effects between SRS and nonconference win rates | | |
|  | C Tourney + | NCAA + |
|  | (1) | (2) |
| Average Opponent SRS | 0.025^ | 0.107‡ |
|  | (0.015) | (0.024) |
| Demeaned Nonconference Win Rate | -0.453 | 3.617‡ |
|  | (0.505) | (0.697) |
| Average Opponent SRS\*NWR | -0.081 | -0.216^ |
|  | 0.075 | 0.113 |
| Observations | 605 | 605 |
| (Pseudo) R Squared | 0.06 | 0.43 |
| Notes: Each observation is measured at the team/year level. The reported coefficients are marginal effects, and robust standard errors are reported in parentheses. The outcome is measured in number of tournament games played plus one extra for the tournament winner. Each regression includes team and season fixed effects, and a control for the win-loss ratio, and point margins. ^ p<0.10 † p<0.05 ‡ p<0.01 | | |

*Figure 1.* The probability of recommending difficult practice in the survey experiment, as a function of the goal of the training (motivating people to continue versus preparing people to succeed) and the type of sport (running/individual versus basketball/team). Error bars are ± 1 standard error.



*Figure 2.* The non-linear effects of difficulty of nonconference games (SRS) and the extent to which these effects vary depending on results of the nonconference games, in the archival data.

Note: Because this figure visualizes the results of two separate sets of analyses (see Table 3, Table 4), this figure is only an approximation. For precise estimates of the magnitude of the relevant relations, see the relevant Tables. The scale on the x-axis is unlabeled here to prevent misinterpretation.

1. We used this approach because we were specifically interested in factors that affect the likelihood of perceiving difficult training as the most useful. Across questions, between 10% and 21% of respondents indicated that the difficult and easy training were equally useful. If we treat “they are equally useful” responses as missing and compare only recommendations for difficult versus easy training, conclusions are identical regarding the effect of goal, χ2(1) = 36.34, *p* < .001, and the effect of recommendation target, χ2(1) = 2.97, *p* = .085. However, the interaction effect of sport x goal is no longer statistically detectable, χ2(1) = .20, *p* = .66. Thus, readers should bear in mind that our inferences about differences in the recommendations that survey respondents made about running (an individual sport) versus basketball (a team sport) rely on comparing the belief that difficult training is superior to all other beliefs. [↑](#footnote-ref-1)
2. Another outcome one might consider is regular season performance. Regular season games are not a good outcome because they vary from conference to conference. Even among the top conferences, some conferences in any given year are harder than others. The NCAA tournament is the same for everyone, so that gives us the same test for teams even in different conferences. [↑](#footnote-ref-2)
3. see https://www.sports-reference.com/cbb/conferences/sec/2013.html [↑](#footnote-ref-3)
4. As an example, BYU basketball had, prior to the 2017-2018 season, 100% of the 2018-2019 nonconference games scheduled, 90% of the 2019-2020 games scheduled, and 81% of the 2020-2021 nonconference season scheduled. [↑](#footnote-ref-4)
5. see http://www.ncaa.com/news/basketball-men/article/2017-03-12/march-madness-bracket-how-68-teams-are-selected-division-i [↑](#footnote-ref-5)
6. While alternative measures of tournament success exist, each has shortcomings. A simple count of games played in the tournament fails to recognize the tournament winner, and a straight count of games won in the tournament would treat the teams that earned a spot in the tournament but lost the first game identically to the teams that failed to reach the tournament. Doing the latter would leave almost half of the variation in our data misspecified. While we choose the number of games played plus one for the tournament winner as our preferred outcomes, the results are completely robust to the alternate specification where we count only games played. [↑](#footnote-ref-6)
7. We replicate our main results using OLS and find no major difference relative to the Poisson regressions in the main results. [↑](#footnote-ref-7)
8. Since our variable of interest, average opponent SRS, varies within group, we chose a fixed effects Poisson model over a random effects model, as the fixed effects model does not carry the assumption that the unobserved error is not correlated with any of the independent variables. The results are insensitive to random effects specifications. [↑](#footnote-ref-8)
9. Our models suggest a quadratic relationship to be most appropriate. This is confirmed by semiparametric regressions where the team and season fixed effects are parametrically modeled and the relationship between SOS and tournament run is allowed a flexible nonparametric form. The semiparametric models confirm the appropriateness of a quadratic relationship. [↑](#footnote-ref-9)
10. While we present these results with parametrized nonlinearities, semiparametric regressions that allow for full functional form flexibility between tournament run and preseason opponent SOS report the same quadratic result. [↑](#footnote-ref-10)
11. For example, the top 5 teams in the 2006-07 ACC averaged a SRS score of 19.1. [↑](#footnote-ref-11)