5-2013

# Toughness Predicts Performance In College Football

Joseph B. Rigoni
*University of Nebraska-Lincoln*, jbrigs3@yahoo.com

# TOUGHNESS PREDICTS PERFORMANCE IN COLLEGE FOOTBALL

by

J. Brandon Rigoni

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Psychology

Under the Supervision of Professor Daniel W. Leger

Lincoln, Nebraska

May, 2013

TOUGHNESS PREDICTS PERFORMANCE IN COLLEGE FOOTBALL

J. Brandon Rigoni, Ph.D.

University of Nebraska, 2013

Advisor: Daniel W. Leger

A premium college football player is estimated to generate over $1 million for his program, so optimal player assessment and selection are paramount. Lean body mass, back squat, and vertical jump have been the most predictive physical test scores, but such metrics typically account for less than 10% of variance in football performance. NFL scouts have tended to rely on vertical jump, 40-yard dash, and 20-yard shuttle scores, but interestingly, reliance on none of the physical tests conducted at the NFL Combine was predictive of team success; in fact, teams that relied on fewer total physical test scores tended to win more games. Dr. Tom Osborne suggested the Performance Index for interpreting physical performance metrics, but also insisted that psychological attributes such as toughness were equally important to his evaluation process since players deal with immense adversity during the course of a season. Toughness can be characterized as the ability to cope with stressors more effectively, and is measured in this study by examining cortisol reactivity. When physical test scores (i.e., lean body mass, hang clean, back squat, bench press, 10-yard dash, 20-yard shuttle, and vertical jump) were converted into Performance Index scores among 47 Division 1 freshman football players, players who contributed on the field differed from players who did not contribute only in vertical jump and 10-yard dash scores. A multiple regression model that included only physical test scores accounted for 28% of variance in football performance, whereas a model that included cortisol reactivity measured during a physical testing session, in addition to traditional physical

performance prediction metrics, accounted for 39% of variance. Results from this investigation suggest that cortisol reactivity may capture a distinctive attribute of players and predict on-field performance better than many individual physical performance variables. If cortisol reactivity, as a measure of toughness, can be effectively used to predict performance on the football field, then the metric should be included in performance prediction models that have traditionally included only physical attributes of players.

# DEDICATION

This dissertation is dedicated to all of the players who could have made a difference on the field had they been given an opportunity. May the tools we use to make recruiting decisions capture the essence of what it takes to be a successful football player.

# ACKNOWLEDGEMENTS

I would like to thank many people for their support. Tom Osborne has personified a life based on principle; Mike Arthur, an interminable appetite for learning; and James Dobson, a code of humble work ethic. Doak Ostergard and Dave Kennedy taught me more than they will ever realize. Cal Garbin single handedly inspired my interest in research methods and data analyses. Special thanks to committee member Chris Eskridge for thinking outside of the box, John Flowers for his genuine enthusiasm, and Fred Luthans for his commitment to being the best. My advisor, Dan Leger, has always encouraged productivity and practical skill development. Ultimately he helped me discover my unique role in the world of academia. Thanks to Doug Granger, Jim McClurg, Joe Moglia, Steve Kiene, and Clarence Thomas for the grace with which they have accomplished so much and the interest they have taken in me. Thank you to all of the secretaries, librarians, lab technicians, and IT support staff who make the University of Nebraska a thriving research environment. A special thanks to Prem Paul and Mike Zeleny for recognizing the value of collaboration between Academics and the Athletic Department. I thank my loving family, Dad, Mom, Bud, and Grandma. Most of all, I thank my wife Laura.

# TABLE OF CONTENTS

# LIST OF APPENDICIES, TABLES, & FIGURES

## APPENDICIES



## TABLES



## FIGURES

# Chapter 1

## INTRODUCTION:

## MARGINAL REVENUE PRODUCT OF A PREMIUM COLLEGE FOOTBALL PLAYER

*The Current Economic State of College Football*

Historians believe roots to the game of football began as early as 500 B.C. amidst Greek and Roman cultures (Johnson, 2001). Around 1600 A.D., the English created a more formal representation of the game. By 1866, Henry Chyadwick had unveiled football in the United States, but it wasn't until after the First World War that the game grew in popularity (Arnold et al., 1977). Today, however, football and all that it entails is quite different than what it was immediately following WWI; it is a culture of its own (Johnson, 2001).

In 2002, the National Collegiate Athletic Association (NCAA) signed an 11-year, $6 billion contract with CBS Television Network (Langelett, 2003). In 2011, NCAA Division 1 public schools reported total revenues of over $7 billion (Schnaars et al., 2012), and the University of Texas alone boasted total revenues over $150 million (Wieberg, Upton, & Berkowitz, 2012). College football comprises a significant portion of most athletic departments' budgets and has become highly commercialized (Zimbalist, 2001). NCAA athletics is big business and college football is the engine that drives soaring revenues.

Teams are composed primarily of individual players recruited by the team and it is assumed that a team's total skill level and ability to generate revenue is a function of the individual skill levels of players that make up the team (Brown & Jewell, 2004). Those teams adept at evaluating players' abilities and enticing them to accept a scholarship acquire superior talent and consequently generate a higher-quality product (i.e., successful teams) and additional revenues. Investments in recruiting among collegiate football programs have increased steadily along with revenues (Zimbalist, 2001). Since the NCAA limits each institution to 85 football scholarships, significant time and resources are spent evaluating prospects and managing principles of economic scarcity within the player recruitment market.

*Marginal Revenue Product of a Premium College Football Player*

A logical question coaches might ask is, "What is the monetary value associated with acquiring a premium player?" Stated differently, this question can be operationalized as, "How much does revenue increase as a function of the number of premium players on a given team?" Such a question can be empirically tested and researchers, primarily in the field of sports economics, have conducted investigations by collecting relevant econometric data and controlling for a number of market factors.

Dr. Robert Brown and colleagues have addressed this topic in a series of studies. In 1993, Brown used the number of college players (from the 1988 and 1989 seasons) that were eventually drafted by the National Football League (NFL) to represent the number of premium players on college teams. He then regressed college team revenues on the number of future NFL draftees. The resultant coefficient, estimated to be the marginal revenue product of a premium college football player, was calculated to be over $500,000. This value represented the expected increase in team revenues that would be realized as a result of acquiring a single premium college football player, when other factors are held constant (Brown, 1993).

Unfortunately, these results were based on a small sample size (i.e., 39 college football teams) and broad calculations of team revenues. At that time, the U.S. Department of Education did not require schools to report any athletic financial data. Even when schools did file information, accounting methods were not always consistent, rendering much of the data useless (Brown, 2010). In 2004, Brown and Jewell revised the gross estimates using data compiled by *The Kansas City Star*. Among 87 college football teams, when revenues were aggregated across all sources, marginal revenue products were estimated to be just over $400,000 per premium player (Brown & Jewell, 2004).

In 2010, Brown updated estimates once again using more recent data (i.e., data from the 2004 – 2005 season) and calculated a premium college football player's marginal revenue product to be over $1 million. In other words, college football teams were estimated to earn roughly $1 million more per premium player than they would have earned without such players. After controlling for inflation, this value is nearly 30% greater than previous estimates (Brown, 2010).

*Intentions of Current Research*

Given the popularity of modern college football and the value of acquiring premium players, one might expect a significant body of literature devoted to tests that can be used to predict success on the football field. However, a surprisingly small amount of empirical research has been conducted. Football is arguably one of the least studied sports in the United States and little scientific evidence has been published in regard to the physical demands of the game (Johnson, 2001).

Many athletes have enhanced their performance on the field as a result of improving characteristics like strength, power, and speed (Pearson & Gehlsen, 1998), and formal systems of evaluation for such attributes have emerged as part of strength and conditioning culture (Arthur & Bailey, 1998). In fact, many view performance testing as an essential aspect of any strength and conditioning program (Zatsiorsky & Kraemer, 2006). Football coaches consider performance testing, as conducted by strength and conditioning professionals, an invaluable part of the player evaluation process (Johnson, 2001).

Physical characteristics of Division I football players are typically measured using a battery of tests designed to address different aspects of performance (e.g., strength, agility, speed, etc.). Single-repetition maximum strength tests and related assessments such as sprint and jump tests are believed to reflect the physical performance characteristics that represent potential to excel as a football player even though these metrics are not direct measures of on-field performance (Fry & Kraemer, 1991). Results from such tests are often used to predict playing status (Johnson, 2001).

Performance testing may provide coaches with valuable information regarding physical characteristics of prospective players, but which other attributes might contribute to success on the football field? The concept of toughness is familiar to many applied practitioners and subject matter experts working within the game of football, but few have undertaken to investigate toughness empirically. This dissertation is intended to apply an evidence-based approach to investigate what many coaches intuitively understand: there is more to success on the football field than mere physical ability. Toughness predicts performance in college football.

*Preview*

Chapter 2 evaluates physical performance prediction metrics that have appeared most frequently in published literature. Each variable was evaluated for the extent to which it has predicted performance on the football field by means of meta-analyses. After average weighted effect sizes for each test were calculated, the chapter concludes by answering the question, “What do these results suggest about thought processes and procedures with respect to assessment and selection in the game of football?”

Chapter 3 conveys insights from applied practitioners who are considered exceptionally skilled at selecting talent in the game of football, namely NFL scouts. Scouts responded to surveys that posed the question, “On which tests do you rely most when determining which players to draft?” Responses were tallied in order to identify the tests relied upon most by NFL scouts to predict performance on the football field. In order to distinguish tests on which more successful teams tend to rely, the number of wins by each franchise during the subsequent 3 seasons was correlated with whether or not the team’s scouting department had designated a specific performance test useful. The extent to which winning teams relied upon a given test was interpreted to suggest the utility of specific tests (or lack thereof). Consistent with the meta-analyses in Chapter 2, this investigation was intended to delineate combinations of traditional physical performance prediction metrics that best predict performance on the football field.

Chapter 4 presents a structured interview with subject matter expert Dr. Tom Osborne. Osborne was extraordinarily successful as a Division 1 head football coach despite some inherent recruiting disadvantages associated with his program. During the interview, Osborne described recruiting considerations in depth. He had vast experience evaluating players based on physical characteristics; however, he suggested that psychological attributes were equally important to his selection process. Osborne explained his unique perspective with emphasis on his psychology background. He is considered an authority in multiple fields (i.e., both the game of football, which has historically emphasized physical performance metrics, and the field of psychology) and had developed an integrative approach that proved successful. Osborne's input expanded upon what was learned from meta-analyses of published literature *(Chapter 2)* and NFL scouting perspectives *(Chapter 3)* and guided the programs of research found in Chapters 5 and 6.

Chapter 5 proposes the Performance Index as a favorable conversion of traditional physical performance prediction metrics. The Performance Index is a decathlon-type scoring system that rewards points on an increasing basis as athletes approach a world-class performance. Importantly, the calculation controls for body weight, therefore allowing objective comparisons of players from different positions. Performance Index conversions, as pound-for-pound indications of physical ability, were applied to physical performance tests that had been recorded among freshman football players from the University of Nebraska. Each test score, after being converted to a Performance Index score, was evaluated for the extent to which it predicted players' contributions on the football field. The utilities of different test scores were compared to decipher physical performance metrics that are most predictive for the population.

Chapter 6 discusses an alternative measure of toughness that could be used to evaluate prospective players. The theory of toughness embodies a reciprocal flow between psychological and physiological systems and emphasizes a distinction between appraising circumstances as being challenging (i.e., potentially taxing but likely to lead to positive outcomes) or threatening (i.e., unpredictable, uncontrollable, and likely to lead to adverse consequences). A relationship has been observed between appraisals and the correspondence of two primary physiological stress response systems (i.e., the autonomic nervous system and the HPA-axis). Salivary cortisol reactivity in response to a testing session was measured among the group of freshman football players at the University of Nebraska as an indication of HPA-axis activation, and by extension, “toughness”. Cortisol reactivity was evaluated for the extent to which it predicted players' contributions on the football field. In addition, a nested model comparison was conducted in order to determine whether measuring the construct referred to as toughness could improve models that have historically included only traditional physical performance prediction metrics to predict performance on the football field.

Football is arguably the most popular sport in the United States. This may be due to the superior strength and speed of players relative to the general population, especially at the Division I and professional level (Johnson, 2001). Physical characteristics that are conventionally believed to predict success on the football field should be subjected to rigorous scientific evaluation so that tests can be optimally applied. In addition, psychological attributes that might promote success on the football field should be assessed for the extent to which existing performance prediction models can be improved.

This dissertation explores: 1) the traditional physical performance prediction metrics that are most predictive of performance on the football field, the extent to which such metrics can be relied upon when making decisions, and the way in which one can incorporate such metrics during evaluation processes; and 2) the extent to which measuring the construct referred to as "toughness" can improve models used to predict performance on the football field. With this information, strength and conditioning professionals and football coaches will be able to draw more accurate conclusions regarding the predictive validity of various tests and ultimately make better recruiting decisions. In summary, I am attempting to better measure the essence of what it takes to be successful in the game of football.

Chapter 2

# META-ANALYSIS OF TRADITIONAL PHYSICAL PERFORMANCE PREDICTION METRICS

*Traditional Physical Performance Prediction Metrics*

Within today's college football recruiting landscape, success on the field is generally attributed to players' physical characteristics (Hyllegard, Radlo, & Early, 2001). One can conceptually organize the physical measures that have been recorded most frequently in published literature according to 4 categories. "Anthropometric" tests include those related to body composition such as height, body weight, body fat percentage, lean body mass, etc. "Strength" measures correspond with exercises completed as part of a strength training program, which would typically take place in a weight room. Strength tests include power clean, hang clean, back squat, bench press, etc. "Functional" measures, by contrast, generally require more space. These tests might be completed on an outdoor football field or inside an indoor practice facility as they involve increasingly dynamic motions. Examples of functional tests include the 10-, 20-, and 40-yard dash, vertical jump, broad jump, 20-yard shuttle, etc. The category "Other" includes all measures that don't conceptually fit nicely into any of the previously mentioned categories (e.g., resting blood pressure). It should be noted that the category "Other" comprises a very small percentage of measures observed in published literature.

A systematic search of references provided by colleagues and through bibliographic references in articles revealed 16 publications that predicted performance on the football field using traditional physical performance prediction metrics. The metrics that were most frequently cited included: Height (11 studies), Body Weight (13 studies), Body Fat Percentage (7 studies), Lean Body Mass (5 studies), Power Clean or Hang Clean (4 studies), Back Squat (9 studies), Bench Press (12 studies), 10-Yard Dash (4 studies), 40-Yard Dash (12 studies), 20-Yard Shuttle (a.k.a. Pro Agility Run) (6 studies), and Vertical Jump (14 studies). Several studies included multiple effect sizes for each metric as samples of players were frequently assigned to groups (e.g., by position). While a small degree of variability among studies was evident, as a result of technological advancements, convention or other reasons, most researchers followed very similar testing procedures:

- Height

  "Height was measured with a meter stick to 1 decimal place…" (Stuempfele, Katch, & Petrie, 2003, p. 239).

- Body Weight

  "… body mass was measured on a Healthometer balance beam scale…" (Stuempfele, Katch, & Petrie, 2003, p. 239).

- Body Fat Percentage

  "Prior to underwater weighing, 3 trials of seated vital capacity (ATPS) were determined according to manufacturer's directions using a Medgraphics metabolic cart. Residual lung volume was estimated from vital capacity (BTPS) (residual lung volume = vital capacity * 0.24) according to the report of Wilmore, which revealed very close agreement between body composition measurements using measured vs. estimated residual lung volume. Body mass in water was assessed by hydrostatic weighing in the seated position in a 91- x 91- x 183-cm aluminum tank. Subjects performed 10 successive trials of underwater weighing, with approximately a 1-minute rest interval between trials following procedures described previously. Ten repeated weighings (using an average of the last 3 trials) produces a "true" underwater weight score. For white players, percent body fat was calculated using the equation of Siri, where black players, the Schutte equation was applied, where % fat = (437.4 / density g * $ml^{-1}$ – 392.8)." (Stuempfele, Katch, & Petrie, 2003, p. 239).

- Lean Body Mass

  "… body composition was determined using skinfold measurements and the formula of Jackson and Pollock." (Iguchi, J., Yamada, Y., Ando, S., Fujisawa, Y., Hojo, T., Nishimura, K., Kuzuhara, K., Yuasa, Y., Ichihashi, N, 2011, p. 3375).

- Power Clean

  "1. The bar was to be lifted from the floor, not the hang position, using one continuous motion.

  2. The bar was to successfully attain the proper rack position, and was not to be pushed up from the chest.

3. Squat cleans were not desired." (Fry & Kraemer, 1991, p. 127).

- Back Squat

"1. The athlete was to attain a position with the thighs parallel to the ground (i.e., trochanter level with the knee).

2. The bar could be carried in a high bar (on the trapezius) or low bar (on posterior of deltoids) position.

3. No steps could be taken during the lift. Feet were to remain in full contact with the ground.

4. Lifting aids could be used (e.g., belts, knee raps, etc.)." (Fry & Kraemer, 1991, p. 127).

- Bench Press

"1. The bar had to touch the chest before being pressed to the arm's fully extended position. No pause at the chest was required, but the bar could not be bounced off the chest.

2. The feet had to remain flat throughout the lift, and the buttocks had to maintain contact with the bench.

3. Each arm was to be extended evenly.

4. No lifting aids such as bench press shirts or elbow wraps were allowed." (Fry & Kraemer, 1991, p. 127).

- 10-Yard Dash & 40-Yard Dash

"The sprint tests were conducted simultaneously following an approximately 10-minute team warm-up that consisted of stretching and four 20-yard sprints at ½, ¾, ¾, and ¾ speeds. Athletes started from a 3-point stance, and timers started their stopwatches when the athlete made the first movement to sprint. One timer recorded when the athlete crossed the 10-yd line. Two other timers recorded when the athlete crossed the 40-yd line, and the mean of the 40-yd times was calculated. Three trials were performed, with an approximately 10-minute rest between trials. The fastest 10- and 40-yd trials were designated as criterion speed scores." (Stuempfle, K. J., Katch, K. I., & Petrie, D. F., 2003, p. 240).

- 20-Yard Shuttle

  "The athlete began in a 3-point stance from a position in the middle of the timing gate. The electronic timer started with the athlete's first movement. The task required the athlete to run to either the right or the left for 4.56 m and touch a line with his foot, reverse direction and run 9.1 m, touch the opposite line with his foot, and run back through the timing gate that recorded the elapsed time. The fastest of 3 trials was recorded." (Sawyer, Ostarello, Suess, & Dempsey, 2002, p. 612).

- Vertical Jump

  "The VJ was performed on a wooden platform using a Vertec apparatus. A 2-footed takeoff was used with no approach steps permitted. The score was

> determined by measuring the difference between a fully extended standing reach and a maximal VJ reach. The best of 3 trials was recorded." (Sawyer, Ostarello, Suess, & Dempsey, 2002, p. 612).

The process by which players are selectively recruited can greatly influence the likelihood of success for a team (Humara, 2000). However, despite considerable speculation, very little definitive empirical evidence exists to suggest specific combinations of physical attributes that best predict performance (Davis, Barnett, Kiger, Mirasola, & Young, 2004). Meta-analyses were conducted among studies that sought to predict performance on the football field using traditional physical performance prediction metrics.

*Meta-Analyses*

A meta-analysis is defined by DerSimonian and Laird as, "the statistical analysis of a collection of analytic results for the purpose of integrating the findings." (DerSimonian & Laird, 1986, p. 177). Any given study may be restrictive in scope, thus limiting a researcher's ability to derive unequivocal or generalizable conclusions, when considered on its own. However, if results from multiple comparable studies are combined, evidence may be strengthened (DerSimonian & Laird, 1986).

## Method

### Procedure

One challenge associated with meta-analyses relates to sample size variability. Studies that incorporate differing sample sizes also contain different levels of sampling error. Therefore, one must assign to results from each study a weight that reflects the relative value of the information contained (DerSimonian & Laird, 1986). A Fisher's Zr transformation was applied to correlation coefficient effect sizes so that assumptions of a normal distribution could be met: $ES_{Zr} = .5\ln\left[\frac{1+r}{1-r}\right]$. In order to remove small sample bias, a correction was applied to standardized mean difference effect sizes: $ES'_{sm} = ES_{sm}\left[1-\frac{3}{4N-9}\right]$.

In addition, inverse variance weights were calculated for all standardized mean difference effect sizes: $w = \frac{1}{se^2}$ and correlation effect sizes: $w = n-3$ along with the corresponding standard errors terms: $se = \sqrt{\frac{n_1+n_2}{n_1 n_2} + \frac{\overline{ES}_{sm}}{2(n_1+n_2)}}$ and $se = \sqrt{\frac{1}{n-3}}$ respectively. Finally, weighted mean effect sizes were calculated for each traditional physical performance prediction metric: $\overline{ES} = \frac{\sum(w \times ES)}{\sum w}$, the standard error of the mean: $se_{\overline{ES}} = \sqrt{\frac{1}{\sum w}}$, as well as a Z-test statistics for significance: $Z = \frac{\overline{ES}}{se_{\overline{ES}}}$.

**Measures**

Another challenge associated with meta-analyses involves deciding the extent to which disparate studies are comparable. Studies are often diverse with respect to design and/or research methodology and one must carefully draw from logically related studies when conducting meta-analyses (Armitage, 1984). Since the goal of these meta-analyses was simply to generate an average overall effect size for the relationship between performance and each of the traditional physical performance prediction metrics that appear most frequently in published literature, the meta-analyses found in this chapter did not exclude any level of play nor did they exclude any particular operationalization of on-field performance. These meta-analytic studies were all-inclusive so that a general understanding of the subject matter could be derived as a basis from which more in-depth investigations could spawn.

*See Appendix 2.1*

**Subjects**

Multiple criterion variables were observed in the published literature. The most frequent way in which researchers measured performance on the football field involved comparing starters to non-starters; six studies utilized this approach. Two studies used draft order as a criterion measure; two studies compared players from successful teams to players from less successful teams; and, two studies compared players from a Division 1 college football program to players from a lower division. A single study was identified that used each of the following outcome measures: drafted players compared to non-drafted players; coaches' rankings of players, Rivals.com Stars for players, and professional players compared to collegiate players. A total of five studies included professional football players, ten studies included college football players, and one study included high school football players.

*See Appendix 2.1*

**Research Hypotheses**

I hypothesized that all of the traditional physical performance prediction metrics would be significantly related to performance on the football field. In addition, I hypothesized that the strength of the weighted effect size relationship between each of the traditional physical performance prediction metrics and performance would correspond directly with the frequency with which the test appeared in published literature. That is to say that the relationship between vertical jump and performance was expected to be greatest, relative to the other tests, since vertical jump was listed most frequently in published literature (14 studies), followed by body weight (13 studies), and then bench press and 40-yard dash, etc.

## Results

All of the traditional physical performance prediction metrics most frequently cited in published literature were significant predictors of performance on the football field. Lean body mass was the best predictor of performance on the football field with an average weighted effect size of 0.34. Back squat was also a strong predictor of performance with an average weighted effect size of 0.29. Vertical jump was the best "Functional" predictor of performance on the football field with an average weighted effect size of 0.28. The 20-yard shuttle run was least predictive of performance on the football field with an average weighted effect size of only 0.10.

**Table 2.1 Meta-Analyses Results**

| Test | Effect Size | Standard Error | $R^2$ | N | P-Value |
|---|---|---|---|---|---|
| Height | 0.11 | .021 | .01 | 6,151 | <.001* |

| | | | | | |
|---|---|---|---|---|---|
| Body Weight | 0.21 | .020 | .05 | 7,325 | <.001* |
| Body Fat % | 0.25 | .031 | .06 | 1,910 | <.001* |
| Lean Body Mass | 0.34 | .034 | .11 | 1,160 | <.001* |
| Clean | 0.28 | .027 | .08 | 4,334 | <.001* |
| Back Squat | 0.29 | .023 | .09 | 4,334 | <.001* |
| Bench Press | 0.24 | .016 | .06 | 8,051 | <.001* |
| 10-Yard Dash | 0.21 | .026 | .05 | 1,555 | <.001* |
| 40-Yard Dash | 0.22 | .015 | .05 | 9,659 | <.001* |
| 20-Yard Shuttle | 0.10 | .023 | .01 | 3,554 | <.001* |
| Vertical Jump | 0.28 | .016 | .08 | 8,865 | <.001* |

**indicates p-value <.05*

**Discussion**

Even though all of the traditional physical performance prediction metrics most frequently cited in published literature were significant predictors of performance on the football field when considered as part of a meta-analysis, it is important to note the sample size associated with each significance test. The meta-analysis for lean body mass included 1,160 subjects, which was the fewest of any of the meta-analyses. That means that statistical power for even the weakest test would have rendered a weighted mean effect size less than 0.1 statistically significant.

The meta-analysis for the 40-yard dash included 9,659 subjects; this was the most of any of the meta-analyses found in this chapter and roughly eight times the number of subjects who were included in the meta-analysis for lean body mass. For all intents and purposes, any relationship whatever between the 40-yard dash and performance would have been enough to reach the threshold for statistical significance. It is not saying much to report that each of the traditional physical performance prediction metrics most frequently cited in published literature is a significant predictor of performance on the football field. Nor does it seem appropriate to conclude from these results that current thought processes and procedures with respect to assessment and selection in the game of football are adequate.

Inflated sample sizes associated with each test offer one primary advantage. These analyses are valuable precisely because whether or not a given test predicts performance on the field is irrelevant. The sample sizes are so large that all tests are significant predictors, so we focus our attention instead on the size of the effects. It is most meaningful to evaluate traditional physical performance prediction metrics for the extent to which variance is accounted for. When one considers that lean body mass was the best predictor of performance on the football field, but that the variable only accounted for roughly 11% of variance in on-field performance, the state of affairs is put into perspective.

Variance in on-field performance that was accounted for among traditional physical performance prediction metrics ranged as low as 1% (i.e., height and 20-yard shuttle). Eight of the eleven metrics accounted for between 5% and 10% percent of variance. The bench press, power clean or hang clean, and back squat accounted for 6%, 8%, and 9% of variance, respectively. Even though strength measures captured more variance on average than anthropometric and functional measures, all accounted for less than 10%.

The goal of these meta-analyses was simply to generate an average overall effect size for the relationship between performance and each of the traditional physical performance prediction metrics that appear most frequently in published literature so that a general understanding of the subject matter could be derived as a basis from which more in-depth investigations could spawn. According to DerSimonian & Laird (1986), when results from multiple studies are combined, evidence can be strengthened. We can be more confident in the exactitude of these tests based on these meta-analyses and associated sample sizes than results from any single study due to the law of averages.

To summarize, lean body mass was the best predictor of performance on the football field with an average weighted effect size of 0.34. However, this value was based on only 13 effect sizes and 1,160 subjects. Back squat was also a strong predictor of performance with an average weighted effect size of 0.29. This value was based on 55 effect sizes and over 4,000 subjects. Vertical jump was the best "Functional" predictor of performance on the football field with an average weighted effect size of 0.28. The 20-yard shuttle or pro-agility run was least predictive of performance on the football field with an average weighted effect size of only 0.10.

Results from the meta-analyses suggest that some of the tests on which coaches rely most, such as the 40-yard dash and bench press, account for only 5% of the variance in performance on the football field. Even the best test, lean body mass, can only account for 11% of the variance in performance. Considering the implications of appropriately assessing and selecting prospective players *(Chapter 1)*, the results from these meta-analyses are surprising. Frequently cited tests, such as the 20-yard shuttle, were discovered to be nearly useless when used to predict performance on the football field.

## APPENDIX 2.1: Meta-Analyses Articles

**Title:** Performance Factors, Psychological Assessment, Physical Characteristics, and Football Playing Ability

**Authors & Year:** Barker, M., Wyatt, T. J., Johnson, R. L., Stone, M. H., O'Bryant, H. S., Poe, C., & Kent, M., 1993

**Criterion Variable:** Starters vs. Non-Starters

**Predictor Variables:** Height, Body Weight, Body Fat Percentage, Lean Body Mass, Back Squat, Vertical Jump

**Level:** College Football

**N** = 59

**Title:** Physical and Performance Characteristics of NCAA Division I Football Players

**Authors & Year:** Berg, K., Latin, R. W., & Baechle, T., 1990

**Criterion Variable:** Successful Teams vs. Unsuccessful Teams

**Predictor Variables:** Height, Body Weight, Body Fat Percentage, Back Squat, Bench Press, 40-Yard Dash

**Level:** College Football

**N** = 880

**Title:** Comparisons of Size, Strength, Speed, and Power in NCAA Division 1-A Football Players

**Authors & Year:** Black, W. & Roundy, E, 1994

**Criterion Variable:** Starters vs. Non-Starters

**Predictor Variables:** Body Weight, Bench Press, Back Squat, 40-Yard Dash, Vertical Jump

**Level:** College Football

**N** = 1618

**Title:** Physical and Performance Characteristics of Community College Football Players

**Authors & Year:** Dos Remedios, R. & Holland, G., 1992

**Criterion Variable:** Division 1 Football Players vs. Community College Football Players

**Predictor Variables:** Height, Body Weight, Back Squat, Bench Press, 40-Yard Dash, Vertical Jump

**Level:** College Football

**N =** 1490

**Title:** Physical Performance Characteristics of American Collegiate Football Players

**Authors & Year:** Fry, A. C. & Kraemer, W. J., 1991

**Criterion Variable:** Starters vs. Non-Starters

**Predictor Variables:** Hang Clean, Back Squat, Bench Press, 40-Yard Dash, Vertical Jump

**Level:** College Football

**N =** 775

**Title:** Comparisons of Selected Physical Fitness and Performance Variables between NCAA Division I and II Football Players

**Authors & Year:** Garstecki, M. A., Latin, R. W., & Cuppett, M. M., 2004

**Criterion Variable:** Division 1 Football Players vs. Division 2 Football Players

**Predictor Variables:** Height, Body Weight, Body Fat Percentage, Lean Body Mass, Power Clean, Back Squat, Bench Press, 40-Yard Dash, Vertical Jump

**Level of Play:** College Football

**N =** 572

**Title:** Combine Performance Descriptors and Predictors of Recruit Ranking for the Top High School Football Recruits from 2001 to 2009: Differences between Position Groups

**Author & Year:** Ghigiarelli, J., 2011

**Criterion Variable:** Rivals Stars

**Predictor Variables:** Height, Body Weight, 40-Yard Dash, 20-Yard Shuttle, Vertical Jump

**Level:** High School Football

**N** = 2560

**Title:** Physical and Performance Characteristics of Japanese Division 1 College Football Players

**Authors & Year:** Iguchi, J., Yamada, Y., Ando, S., Fujisawa, Y., Hojo, T., Nishimura, K., Kuzuhara, K., Yuasa, Y., & Ichihashi, N, 2011

**Criterion Variable:** Successful Teams vs. Unsuccessful Teams

**Predictor Variables:** Height, Body Weight, Body Fat Percentage, Lean Body Mass, Back Squat, Bench Press, 40-Yard Dash

**Level:** College Football

**N** = 115

**Title:** Evaluating the Importance of Strength, Power and Performance Tests in an NCAA Division I Football Program

**Author & Year:** Johnson, J., 2001

**Criterion Variable:** Starters vs. Non-Starters

**Predictor Variables:** Body Weight, Power Clean, Back Squat, Bench Press, 40-Yard Dash, Vertical Jump

**Level:** College Football Players

**N** = 452

**Title:** Morphological Profiles for First-Year National Collegiate Athletic Association Division I Football Players

**Authors & Year:** Kaiser, G. E., Womack, J. W., Green, J. S., Pollard, B., Miller, G. S., & Crouse, S. F., 2008

**Criterion Variable:** Professional Football Players vs. College Football Players

**Predictor Variables:** Height, Body Weight, Body Fat Percentage

**Level:** Professional Football and College Football

**N** = 65

**Title:** The NFL Combine: Does it Predict Performance in the National Football League?

**Authors & Year:** Kuzmits, F. E. & Adams, A. J., 2008

**Criterion Variable:** Draft Order

**Predictor Variables:** Bench Press, 10-Yard Dash, 40-Yard Dash, 20-Yard Shuttle, Vertical Jump

**Level:** Professional Football

**N** = 306

**Title:** The National Football League (NFL) Combine: Does Normalized Data Better Predict Performance in the NFL Draft

**Author & Year:** Robbins, D. W., 2010

**Criterion Variable:** Draft Order

**Predictor Variables:** Bench Press, 10-Yard Dash, 40-Yard Dash, 20-Yard Shuttle, Vertical Jump

**Level:** Professional Football

**N** = 1155

**Title:** Relationship between Football Playing Ability and Selected Performance Measures

**Authors & Year:** Sawyer, D. T., Ostarello, J. Z., Suess, E. A., & Dempsey, M., 2002

**Criterion Variable:** Coaches Ranking

**Predictor Variables:** Height, Body Weight, Power Clean, Back Squat, Bench Press, 10-Yard Dash, 20-Yard Shuttle, Vertical Jump

**Level:** College Football

**N** = 40

**Title:** Exercise Performance of Professional Football Players

**Authors & Year:** Shields, C. L., Whitney, F. E., & Zomar, V. D., 1984

**Criterion Variable:** Starters vs. Non-Starters

**Predictor Variables:** Height, Body Weight, Body Fat Percentage, Lean Body Mass, Bench Press

**Level:** Professional Football

**N =** 167

**Title:** The National Football League Combine: Performance Differences Between Drafted and Nondrafted Players Entering the 2004 and 2005 Drafts

**Authors & Year:** Sierer, S. P., Battaglini, C. L., Mihalik, J. P., Shields, E. W., & Tomasini, N. T., 2008

**Criterion Variable:** Drafted vs. Non-Drafted Players

**Predictor Variables:** Height, Body Weight, Bench Press, 40-yard dash, 20-yard shuttle, Vertical Jump

**Level:** Professional Football

**N =** 321

**Title:** Body Composition Relates Poorly to Performance Tests in NCAA Division III Football Players

**Authors & Year:** Stuempfle, K. J., Katch, K. I., & Petrie, D. F., 2003

**Criterion Variable:** Starters vs. Non-Starters

**Predictor Variables:** Height, Body Weight, Body Fat Percentage, Lean Body Mass, Bench Press, 10-Yard Dash, 40-Yard Dash, 20-Yard Shuttle, Vertical Jump

**Level:** College Football

**N =** 77

# Chapter 3

# APPLIED PRACTITIONERS: NFL SCOUTING PERSPECTIVES

*The National Football League*

The National Football League (NFL) is the highest level of football in the United States (Sierer et al., 2008). The league is composed of 32 teams and the average NFL franchise is worth over $1 billion (Badenhausen et al., 2010)! A team's ability to win is a function of the individual skill levels of players that make up the team (Brown & Jewell, 2004). The value of an NFL franchise is generally understood to increase with team success. Therefore, precise evaluation of prospective players has significant financial implications for NFL teams (Robbins, 2010).

Teams typically carry only 53 players on an active roster (Ghigiarelli, 2011). Each year, NFL teams have an opportunity to select and acquire new players from a pool of college athletes during the rookie draft (McGee & Burkett, 2003). Players prepare diligently to demonstrate their physical skills because a strong performance at the NFL Combine (described below) can maximize a player's draft status, advancing him within the same round or even between rounds, and securing a coveted contract with associated financial rewards (McGee & Burkett, 2003; Kuzmitz & Adams, 2008; Plothow, 2010).

*NFL Combine*

The NFL Combine is a week-long event, held each spring in preparation for the rookie draft, at which all 32 NFL franchises can thoroughly evaluate eligible prospects (Robbins, 2010). Only the very best college players are invited to attend; approximately 3% of all NCAA Division 1 football players receive an invitation (Kuzmitz & Adams, 2008). Players gather at a neutral site in Indianapolis and participate in a battery of tests amidst NFL scouts, front office executives and draft experts (Robbins, 2010). Tests conducted at the Combine are intended to exhibit attributes that contribute to a player's ability to be successful in the NFL (Plothow, 2010).

The Combine comprises multiple phases of evaluation including position specific drill work, cognitive testing, injury susceptibility analysis, drug screening and interviewing. Among the most important assessments are tests of physical ability (Robbins, 2010). Importantly, testing procedures are standardized in order to "level the playing field" for all athletes (Robbins, 2010). During the Combine, prospects participate in the following physical tests: bench press, 10-, 20-, and 40-yard dash, 20-yard shuttle run, 60-yard shuttle run, 3-cone drill, vertical jump, and broad jump (McGee & Burkett, 2003):

- 225-lb Bench Press to Fatigue Test

  "The 225-lb bench press to fatigue test is the only test to measure upper-body muscular strength in the testing battery of the combine. Athletes are instructed to complete as many bench press repetitions with 225 lb as possible. A countable repetition is defined as lowering the weight just touching the chest, followed by a brief pause and then an upward push to return the weight to the starting position with arms fully extended." (McGee, K. J. & Burkett, L. N., 2003, p. 7).

- 40-Yd Dash With 10- and 20-Yd Split Times

  "The 10-, 20-, and 40-yd dash tests anaerobic power, acceleration, and speed. Electronic timing devices are placed at the starting line and the 10-, 20-, and 40-yd lines. Time is recorded at all 3 distances to one-hundredth of a second. When the athlete is in proper position, he sprints as fast as he can from the starting line through the string placed at the 40-yd-dash marker that signifies the completion of the sprint." (McGee, K. J. & Burkett, L. N., 2003, p. 8).

- 20-Yd Shuttle Proagility Run Test

  "The 20-yd shuttle, also known as the proagility run test, measures the anaerobic power, the ability to increase and decrease speed rapidly, and the ability to change direction quickly. To perform the 20-yd shuttle, an athlete straddles the 15-yd line, runs to his right and touches the 20-yd line. Then, he quickly changes direction, sprints past the 15-yd line, and touches the 10-yd line. Again he quickly changes direction and finishes by sprinting through the 15-yd line. Athletes perform the test twice, once in each direction. The average time is recorded for each direction." (McGee, K. J. & Burkett, L. N., 2003, p. 8).

- 60-Yd Shuttle

"The 60-yd shuttle is a measure of speed, flexibility, body control, and a small level of endurance. The shuttle is a basic out-and-back running test from the goal line to the 5-, 10-, and 15-yd lines. Time is recorded from the athlete's initial movement until completion of the shuttle to the nearest one-hundredth of a second." (McGee, K. J. & Burkett, L. N., 2003, p. 8).

- 3-Cone Drill

  "The 3-cone drill is a measurement of agility, change in direction, and power. To perform the 3-cone drill, 3 cones are positioned in an upside-down "L" formation. The athlete starts at cone 1 from a 3-point stance behind the starting line. On his own volition, the athlete sprints as fast as possible and touches cone 2, which is 5 yd directly in front of him, and immediately returns to cone 1. Without stopping, the athlete changes directions, corners cone 2, and sprints directly to cone 3, which is 5 yd lateral to cone 2, on the athlete's right-hand side. The athlete circles cone 3 to his left, then returns to the first cone by cornering cone 2 and sprinting at full speed past cone 1, which marks the finish line. Time is recorded to the nearest one-hundredth of a second." (McGee, K. J. & Burkett, L. N., 2003, p. 8).

- Vertical Jump

  "The vertical jump is a measure of leg strength and anaerobic power. The Vertec is used to assess the vertical jump. The athlete positions himself directly underneath the vanes of the Vertec, allowing him to touch the vanes by jumping straight up without any lateral adjustments. The athlete lowers his center of gravity in a counter movement and explosively jumps straight up in the air off of both feet. The goal of the athlete is to hit the highest vane possible with 1 hand. The athlete's vertical jump is measured by subtracting the height of the athlete's standing reach from the height of the highest vane hit." (McGee, K. J. & Burkett, L. N., 2003, p. 8).

- Standing Broad Jump

  "The standing broad jump is used as a measure of leg strength and power. The athletes are instructed to assume a position with their toes behind taped line marked "zero inches." When an athlete is set, he jumps horizontally, taking off from both feet on his own command. The distance jumped is recorded from the start line to the point of heel contact or the closest body part measured to the nearest inch." (McGee, K. J. & Burkett, L. N., 2003, p. 8).

Significant time and resources are spent to assess player potential at the Combine (Ghigiarelli, 2011). Scores from physical performance tests are frequently used to predict future success in the game of football (McGee & Burkett, 2003). Draft experts attempt to select players who will be most productive and many tests conducted at the NFL Combine influence such selections (Plothow, 2010).

*Literature Review*

McGee and Burkett (2003) employed linear regression analyses to predict the draft status of prospects using physical test scores collected at the NFL Combine. Subjects included 326 players who entered the 2000 NFL draft. However, only data from players who were actually drafted were used to generate regression equations. Players were grouped according to 7 positions (i.e., quarterbacks, wide receivers, running backs, offensive line, defensive line, defensive backs, and linebackers). The round in which a player was drafted was used as the criterion variable. Independent variables included height, body weight, 10-yard dash, 40-yard dash, bench press, 20-yard shuttle, 3-cone drill, vertical jump, and broad jump. Not all positions completed each test.

Results suggested that multiple regression models, which include physical test scores collected at the NFL Combine, can account for 100% of the variance among wide receivers, running backs and defensive backs when predicting the round in which player would be drafted. For quarterbacks, 84% of the variance was accounted for; offensive lineman, 70%; and defensive linemen, 59%. Variance accounted for among linebackers was considerably less at only 22%. The authors concluded that the draft status of running backs, wide receivers, and defensive backs can be accurately predicted using tests conducted at the Combine, and that such tests generate fair estimates for other positions as well. In addition, players drafted in the first and second rounds were discovered to outperform players drafted in the sixth and seventh rounds in the broad jump, vertical jump, and 3-cone drill when evaluated for tests of significance.

Kuzmits and Adams (2008) utilized correlation analyses to investigate the relationship between NFL combine test results and draft status. Subjects were quarterbacks, running backs, and wide receivers drafted from 1999 – 2004. Combine tests included 10-, 20-, and 40-yard dashes, bench press, vertical jump, broad jump, 20- and 60-yard shuttles, and 3-cone drill. For quarterbacks, both vertical jump and broad jump were significantly related to draft order. Sprint times (i.e., 10-, 20-, and 40-yard dashes) were predictive of draft order among running backs.

Sierer et al., (2008) compared combine test scores of drafted players to those who were not drafted during 2004 and 2005 drafts. Subjects included 321 players who were classified as either a skill player (i.e., wide receiver, cornerback, free safety, strong safety, or running back), big skill player (i.e., fullback, linebacker, tight end, or defensive end), or linemen (i.e., center, offensive guard, offensive tackle, or defensive tackle). Combine test measures that were included in analyses were height, body weight, 40-yard dash, bench press, vertical jump, broad jump, pro agility run, and 3-cone drill.

Among skill players, drafted players performed significantly better than non-drafted players in the 40-yard dash, vertical jump, pro-agility run, and 3-cone drill. For big skill players, drafted players performed better at the 40-yard dash and 3-cone drill. Among linemen, drafted players performed better at the 40-yard dash, bench press, and 3-cone drill.

Robbins (2010) correlated normalized data (i.e., controlling for body weight), in addition to raw data, with draft order represented within each position. Subjects included 1,155 players who were drafted between 2005 – 2009 and classified by position (i.e., center, cornerback, defensive end, defensive tackle, free safety, fullback, inside linebacker, kicker, offensive guard, offensive tackle, outside linebacker, punter, quarterback, running back, strong safety, tight end, or wide receiver). Combine tests included 10-, 20-, and 40-yard dashes, vertical jump, broad jump, 20-yard shuttle, 3-cone drill, and bench press.

The primary finding was that normalized data were not more highly correlated with draft order than were raw data. Twenty-nine raw data test scores were significantly correlated with draft order. While some evidence was found to suggest that normalized data for tests involving change of direction (e.g., 20-yard shuttle or 3-cone drill) may correlate better with draft order than raw data, in general, raw data were comparable to, or better than, normalized data sets when used to predict draft status.

Plothow (2010) utilized regression analyses to predict draft number. Subjects included quarterbacks, running backs (i.e., fullbacks and halfbacks), receivers (i.e., excluding tight ends), and offensive linemen who were drafted between 2000 and 2007. Combine tests included height, weight, bench press, 10-, 20-, and 40-yard dashes, vertical jump, broad jump, and 3-cone drill, depending on the position.

Only height was included in analyses for quarterbacks, but it was not found to be a significant predictor of draft status. For running backs, 10-yard dash, 3-cone drill, and 20-yard shuttle run were all jointly significant predictors of draft number. However, the 20-yard shuttle run was negatively related to draft status. The 40-yard dash was by far the strongest predictor of draft number among receivers. For offensive linemen, body weight, 40-yard dash, and vertical jump were somewhat predictive of draft number.

Variations in research design elements might explain the variant results. Differing sample sizes were assessed across studies. Criterion variables varied and included a player's draft number (e.g., Kuzmits & Adams, 2008; Plothow, 2010), draft number within a position (e.g., Robbins, 2010), the round in which a player was drafted (e.g., McGee & Burkett, 2003), and whether or not a player was drafted at all (e.g., Sierer et al., 2008). In some studies players were classified by position (e.g., Robbins, 2010; McGee & Burkett, 2003; Kuzmits & Adams, 2008) whereas in other studies, positions were grouped together (e.g., Sierer et al., 2008; Plothow, 2010). For some investigations, raw data were used (e.g., Kuzmits & Adams, 2008), but for others, normalized data were evaluated (Robbins, 2010). Statistical analyses also varied (e.g., correlation, analysis of variance, or multiple regression analyses).

Results have clearly varied considerably. Some have found a speedy 20-yard shuttle run to be predictive of draft status, whereas others have discovered the 20-yard shuttle run to be inversely related to draft outcomes. Some have found jumps (i.e., vertical jump and broad jump) and 3-cone drill to be strong predictors of NFL draft status. Others have found sprint speed, over 10-, 20-, and 40-yards, to be a good indicator for certain positions.

Results for the 40-yard dash might be most thought provoking of all. Many coaches and fans alike evaluate players almost exclusively based on their 40-yard dash time. Indeed, results from multiple studies suggest that a player's 40-yard dash time is a statistically significant predictor of draft status. Interestingly however, some of the statistically significant relationships observed in published literature have been discovered among populations of linemen but not for other positions, even though linemen rarely ever run 40-yards during play.

Normalized data do not seem to improve predictability of draft status beyond raw score data, except maybe for tests involving change of direction. Some researchers have cautioned that, even though occasional statistically significant relationships emerge between physical test scores and draft outcomes, given the vast number of statistical tests conducted in most studies, one would expect to find some relationships by chance alone. This problem equates to alpha inflation (i.e., increased likelihood of committing a type 1 statistical error due to the number of pairwise comparisons) and researchers might benefit from Bonferroni corrections when evaluating tests for statistical significance.

The relationship between physical test scores and on-field performance was addressed in Chapter 2. This chapter is intended to supplement those findings with insights from applied practitioners who are considered exceptionally skilled at selecting talent in the game of football, namely NFL scouts. Consistent with this approach, the literature review emphasized the behavior of drafters rather than draftees. NFL scouting perspectives are considered but not necessarily the future success of players on the field.

NFL scouts likely consider multiple sources of information (e.g., results from injury susceptibility analyses, structured interviews, etc) (Plothow, 2010). Tests of physical ability are regarded among the most important assessments (Robbins, 2010). Clearly more work must be done in order for researchers to understand which physical performance metrics NFL scouts consider most important when drafting players. In an effort to better understand the physical characteristics that predict success on the football field, input provided by NFL scouts was retrospectively evaluated.

## Method

**Procedure**

In 2009, during regularly scheduled visits to the University of Nebraska, NFL scouts were asked by the school's head strength and conditioning coach for football to complete a questionnaire so that more could be learned about what it takes to be successful in the NFL. Each scout was given an indefinite amount of time to complete the questionnaire. Scouts were told that the information might be analyzed and shared, and were instructed to leave their team's name off the questionnaire or simply to not complete the questionnaire if they had any concerns.

**Measures**

The written questionnaire contained the question, "On which tests do you rely most when determining which players to draft?" The question was presented in a fill-in-the-blank / open-ended format so that scouts would not feel restricted and could respond in a subjective manner. That is to say, scouts were welcomed to answer the question with as many and whichever test scores they considered important. In addition, the total number of wins each team accumulated over the subsequent 3 seasons (i.e. 2009 – 2011) was recorded. A 3-year period was selected specifically to coincide with the approximate average tenure of an NFL head coach, during which the general approach to assessing and selecting players could be expected to remain somewhat consistent.

**Subjects**

Responses from each of the 32 National Football League franchises were solicited. Specifically, a representative from each team's scouting department was asked to complete a questionnaire when present at the institution. In order for a scout's response to be included in data analyses, the response had to be definitive (e.g., the scout had to have listed 40-yard dash rather than "speed") and the scout had to have indicated his team affiliation so that responses could be correlated with team success. A total of 25 teams responded directly to the inquiry.

### Research Hypotheses

I predicted that the vertical jump would be the most frequently cited combine test, and that winning teams would be more likely to consider it, since the vertical jump was discovered to be the best predictor of performance relative to other combine tests that were included in meta-analyses conducted in Chapter 2. Lean body mass, hang clean, and back squat were better predictors of performance than the vertical jump, according to the meta-analyses in Chapter 2, but these tests are not included at the NFL Combine. I hypothesized that the 20-yard shuttle run would be the least frequently cited test since results from the Chapter 2 meta-analyses and the literature review are mixed at best for that particular test.

## Results

Results are summarized in Table 3.1. The test that was most frequently cited by NFL scouts as being relied upon when determining which players to draft was the vertical jump (10 teams), followed by the 40-yard dash (9 teams) and 20-yard shuttle (8 teams). Least frequently cited was the 20-yard dash (only one team).

**Table 3.1 NFL Scouting Perspectives**

| Test | Number of Teams | Correlation | p-value |
|---|---|---|---|
| Vertical Jump | 10 | -.27 | >.05 |
| 40-Yard Dash | 9 | -.12 | >.05 |
| 20-Yard Shuttle Run | 8 | -.06 | >.05 |
| 10-Yard Dash | 6 | -.34 | >.05 |
| Broad Jump | 5 | -.22 | >.05 |
| Body Weight | 4 | .06 | >.05 |
| 3-Cone Drill | 4 | -.09 | >.05 |
| Bench Press | 4 | -.34 | >.05 |
| Height | 3 | -.09 | >.05 |
| Hang Clean or Power Clean | 2 | -.25 | >.05 |
| 20-Yard Dash | 1 | -.22 | >.05 |
| Number of Tests Listed | | -.56 | <.05* |

**indicates p-value <.05*

Of note, power clean or hang clean was not a test conducted at the NFL Combine. Scouts presumably consulted collegiate strength and conditioning coaches for estimates. See Chapter 2 for a description of power clean.

Some responses included anthropometric tests. Four teams listed body weight as a consideration and three teams listed height. Few teams listed strength measures; four teams listed the bench press test, and two teams listed the hang clean or power clean.

Correlation analyses were conducted to determine the relationship between relying upon a given test and winning. Specifically, whether or not a team had listed a test score was correlated with the total number of wins the team accumulated over the 3 subsequent seasons. Results suggest that there was no relationship between listing any of the tests and winning.

Lastly, a correlation analysis was conducted in order to determine whether the number of test scores a team had listed as tests that were relied upon when determining which players to draft was predictive of the number of games a team won over the subsequent 3 seasons. The number of test scores selected was associated with the total number of wins a team accumulated. Teams that listed fewer tests tended to win more games, $r(25) = -.56$, $p < .05$.

**Discussion**

The average NFL franchise is worth over $1 billion and the value of an NFL team is understood to increase with team success. With so much on the line, one might expect some consensus among NFL scouts regarding the way in which physical test scores can best be used to predict performance on the football field, especially considering the significant time and resources spent to assess player potential at the NFL Combine. On the other hand, the margin of difference between an average or below average team and a great team at the NFL level may be miniscule and teams might seek to maintain any competitive advantage they can muster by keeping their approach to assessment and selection confidential. Regardless of the perspective, tremendous disparity was revealed among NFL franchises with respect to emphases placed upon different metrics. This finding was consistent with the literature review (i.e., different research teams have come to different conclusions when predicting NFL draft status based on NFL Combine test scores).

The three most frequently cited test scores in this study were the vertical jump, 40-yard dash, and 20-yard shuttle. The vertical jump is a general measure of jumping ability. The 40-yard dash is a measure of sprint speed, and the 20-yard shuttle is a general measure of agility. The fact that each of these tests represents a different aspect of performance may suggest an attempt by several teams to capture a well-rounded and "complete" depiction of players. According to this approach, only the best test of each attribute, which has been deemed important (i.e., jumping ability, sprinting ability, and agility), is considered rather than referencing multiple tests of the same attribute (e.g., vertical jump and broad jump which are both measures of jumping ability, or 20-yard shuttle and 3-cone drill which are both measures of agility). The best measures of multiple aspects of performance may be combined to indicate, overall, how effective a player can be on the football field.

Both measures of strength that were listed (i.e., the bench press and power clean or hang clean) appeared in the bottom half in terms of the number of teams that consider these tests important. This may suggest that scouts tend to deemphasize measures of strength. One reason for this may be that few tests of strength are conducted at the NFL Combine (i.e., only the bench press as a measure of upper-body strength). While the questionnaire was formatted so that scouts could freely write in any and all test scores they consider, many scouts may not have considered tests performed outside of the NFL Combine context. Obtaining direct measures of lower-body strength, for example, requires a scout to request information from collegiate strength and conditioning coaches.

Of note, no scouts listed the back squat, which is often considered, along with power clean or hang clean and bench press, to be a core strength training movement and test to be considered when predicting on-field potential. This may be because the depth to which a player squats when completing a repetition can have a dramatic effect on the amount of weight one can handle. When a player squats to a full-squat position, with femur parallel to the ground, he will invariably be able to squat less weight than if he were to complete only a partial squat by stopping short of the full-squat position. Inherent error exists in the precision of measurement when NFL scouts rely upon college coaches for data as some strength and conditioning coaches require players to squat to full-squat position whereas others may not. This conundrum is less prevalent when reporting weights for the power clean or hang clean as the extent to which one has completed a "true" repetition is less ambiguous, but a lack of standardization still exists.

Another explanation for the lack of emphasis on measures of strength is the fact that strong players are likely to demonstrate strength, relative to their body weight, when completing functional tests as well. Since functional tests of speed, jumping ability, etc. are also indicative of strength, evaluation of strength measures may be viewed as redundant. To put it differently, players may be fast because they are strong, but players are never strong because they are fast. Unlike strength tests, functional tests (e.g., 10-yard dash or vertical jump) mimic movements that occur for many positions during play, and therefore may be preferred.

It is important to note that consideration of *none* of the tests was predictive of team success in the NFL. Whether or not a team had listed any given test as an important consideration for determining which players to draft was not related to the number of wins a team accumulated. While the 10-yard dash and bench press test approached significance, the direction of the relationship was opposite of what might have been expected. In fact, the direction of relationship (if there had been a statistically significant relationship) between whether or not a test score had been listed and winning was in the “wrong” direction for each of the tests except for body weight!

One explanation for the quirkiness of these results may be related to range restriction. By definition, players who are invited to the NFL Combine are superior in physical prowess. Players may actually be invited to the NFL Combine based on anticipated physical test scores. By all accounts, players who are invited to the Combine are the best the game has to offer and the fact that everyone is so good may limit the range of scores for any given test. When an occasional statistically significant relationship emerges, it may be that the test in question simply happened to distribute in such a way to offer enough range in scores for an effect to emerge. If few NFL Combine test scores distribute according to a meaningful range of scores, then NFL scouts may simply use Combine tests as low-end cutoffs to screen players (e.g., "Is the player fast enough?"), or to confirm a physical quality in a prospect that the scouts had previously observed while watching game film.

The only statistically significant relationship observed within the datasets was the negative relationship between the number of tests a team reported considering and the number of wins the team accumulated over the subsequent 3 years. Based on this relationship, it was concluded that either: 1) relying on too many tests may confuse the process and detract from honing in on what it takes to be successful in the NFL, or even worse, 2) relying on any formal tests at all when making assessment and selection decisions results in less success. The second perspective was corroborated by feedback received from some of the most successful teams in the NFL.

When asked, “On which tests do you rely most when determining which players to draft,” scouts from some of the most successful teams simply said, “We don’t look at test scores.” With some prompting (e.g., “But if you had to choose one, which would you choose?”), the answer was still very often, “We really don’t look at any test scores; we look at game film to make our decisions.” The three most successful teams surveyed listed no tests at all, and eight of the top 13 teams listed two or fewer tests. It seems that, when teams consider many different test scores, they may simply be guessing. Such teams may not know which attributes are indicative of future success in the NFL, and their lack of precision is expressed in drafting efforts, which manifest themselves in a weaker overall team performance on the football field.

It is also worthwhile to reconsider the direction of the cause-effect relationship between selection techniques and overall team performance, which has thus far been assumed. Maybe poor assessment and selection decisions do not cause poorer team performances. It is possible that the number of tests a team reported considering does not necessarily lead to problems winning games, but that listing multiple test scores is indicative of deeper seated problems within the organization such as a lack of clear and precise direction: correlation does not necessarily ensure causation.

It stands to reason that a snowball effect could occur. For example, a team that is having success, even if for reasons unrelated to assessment and selection techniques, may be more likely to stick to a very specific approach that has worked for them and converge on a small list of indicators when drafting players, whereas a team that has struggled as of late may find itself trying many combinations of predictor variables without any understanding of how different variables relate to on-field performance or a scientific method to discover as much. In the words of one NFL scout, "A great deal of our profession is based on conventional wisdom with little empirical evidence to support our notions."

It is important to note that this study emphasized draft status and NFL scouting perspectives, but did not assess future on-field performance among NFL Combine attendees specifically. That is to say that this study investigated *what* scouts think, but did not evaluate the accuracy of those perspectives directly (although team success over the subsequent three seasons could be considered an indirect measure). The relationship between physical test scores and on-field performance was addressed in Chapter 2; this chapter is intended to supplement those findings with insights from applied practitioners who are considered exceptionally skilled at selecting talent in the game of football, namely NFL scouts.

Possible limitations of this study include the generalities that were forced upon NFL scouts as they answered the surveys. Many scouts think in terms of specifics for respective positions. The physical attributes that are likely to culminate to produce a successful offensive lineman are likely much different than those for a successful quarterback or running back. In addition, the previously discussed potential desire of NFL scouts to maintain any competitive advantage they can manage may have encouraged NFL scouts to contribute disingenuous responses. While cordial of folks to comply with the survey request, no data would be less harmful to the analyses than misleading data.

Also problematic in this study are the inherent limitations associated with linking survey responses from individual scouts to the number of wins a team accumulated over the subsequent 3 years. It is somewhat presumptuous to assume this relationship when so many other factors contribute to team success (e.g., coaching, strength of schedule, injuries, etc.). Further, scouting departments for any given franchise consist of multi-level, jurisdictional constituents; multiple scouts work together within a hierarchy to cover different geographic regions. It is likely an oversimplification to assume that the voice of one individual scout can represent the views of all other scouts within the organization and the franchise at large. Nonetheless, results from this study contribute to our body of knowledge in so far as the complexities of applying physical performance test scores to predict future performance on the football field are extended, even at the highest level of football.

# Chapter 4

# SUBJECT MATTER EXPERT:

# DR. TOM OSBORNE'S INTEGRATIVE APPROACH

*College Football Recruiting Landscape*

College football coaches recruit throughout the year. Home visits, during which coaches can get to know prospective players and their families, are a critical aspect of the recruiting process. High school practice and game attendance allows coaches to observe prospective players in action. Coaches may also rely on guidance counselors and teachers for input (Spieler, 2005). The NCAA limits direct contact between prospective student-athletes and collegiate coaches during certain periods (NCAA, 2010). Other means of correspondence, for example questionnaires (e.g., height, weight, 40-yard dash time, SAT score, etc.), may be utilized during such periods (NCAA, 2010).

Recruiting efforts culminate the first week of February on "Letter of Intent Day," which marks the first opportunity for recruits to formally accept a scholarship. That particular day, and the college football recruiting landscape in general, has become such a hot topic that self-proclaimed "experts" in private industry promote their rankings nationwide (e.g., Borderwars.com, Superprep.com, Poolrecruiting.com, etc.) and head football coaches hold nationally televised press conferences to make announcements (Langelett, 2003). Strong public interest reaffirms the assumption that strong recruiting classes are the basis for successful teams.

Two types of errors may occur during the recruiting process. A type 1 error occurs when a coach admits a player into his football program who fails to contribute sufficiently during the course of his career. Providing such an athlete a scholarship results in monetary loss associated with financial support of the athlete, limited scholarship expenditure as institutions are limited to 85 football scholarships, as well as squandered revenue that might have been accumulated had the coach offered the spot to a different player. A type 2 error occurs when a coach opts to refrain from recruiting or rejects an athlete who would have contributed in a positive way to the overall team product. Potentially most disconcerting about the prospects of a type 2 error is the possibility that an overlooked player might compete for a rival school within the same conference (Spieler et al., 2007).

*Dr. Tom Osborne's Perspective*

Dr. Tom Osborne is considered among many to be the greatest college football coach of all time (Osborne, 2009). Osborne (1999) specifically identified 5 factors that contribute to a successful college football program:

1. Good facilities
2. Tradition
3. Coaching
4. A large population base from which to recruit
5. Particularly good weather during the recruiting season

While the Nebraska Football Program has always had exceptional facilities, a strong tradition, and excellent coaches, the state of Nebraska has a population of only

1.85 million people and adjacent states are not densely populated. Furthermore, cold weather and snow are common during the months when most recruiting visits occur (i.e., December, January, and February).

Needless to say, the Nebraska Football program does not fare well with respect to two of the five factors (i.e., a large population base from which to recruit and particularly good weather during the recruiting season). Both of these factors are directly related to recruiting. Consequently, Nebraska recruiting classes were rarely ranked among the top 10 or 20 nationally during Osborne's tenure. Nevertheless, Nebraska football *teams* ranked among the top 20 nationally for thirty consecutive years (beginning in 1969), and 21 times even finished in the top 10 (Osborne, 1999).

The consistent success of Nebraska Football during the Osborne era can be attributed in part to his approach to assessing, selecting, and developing players. Players were often far more productive at Nebraska than had been expected based on recruiting class rankings (Osborne, 1999). Osborne agreed to present his perspective on record in order to contribute to this dissertation.

## Method

### Procedure

A series of questions were sent via email to University of Nebraska Athletic Director Dr. Tom Osborne. He promptly agreed to discuss his perspective during an interview the following week. On 5/4/2012, Osborne generously participated in a question and answer session that lasted approximately 60 minutes. The interview was structured, but informal in nature, in order to elicit extensive feedback regarding any and all related topics.

Careful notes were taken as Osborne outlined his perspective. The interview flowed seamlessly from one topic to the next, and at times, circled back so that Osborne could address a concept that he had previously left out. This process enhanced the discussion and resulted in a more complete understanding of the relationships between topics.

In this dissertation, Osborne's feedback was reported in context, albeit paraphrased at times for clarity and coherence. Whenever possible, direct quotations were used. The goal of both the interview and presentation of the material was to capture as much information about recruiting as possible as access to such a prominent figure, with as extensive and diverse a background as Osborne, was presumed to be extremely fortunate.

**Measures**

The questions that were sent to Osborne were structured to direct his attention to 5 categories in particular: Recruiting Processes, Talent, Evaluation, Physical Attributes, and Psychological Attributes. This particular format, and the specific questions contained within each section, was selected in an effort to expand upon previous findings (Chapters 2 & 3). In light of Osborne's experience as a psychologist, special attention was given to sport psychology, and questions were phrased to promote explicit comparisons between physical attributes and psychological attributes of players. Considering Osborne's background as both a football coach and psychologist, he may well be the world's most qualified subject matter expert to make such comparisons.

**Subject**

Dr. Tom Osborne was head football coach at the University of Nebraska for 25 years. During that period, the Nebraska Football Program accomplished unparalleled consistency in NCAA annals. Osborne never had a losing season. In fact, each team won at least 9 games and went to a bowl game. Three Nebraska football teams won the National Championship during his last 4 seasons as head coach.

In 1997, Dr. Tom Osborne retired from coaching. He was promptly inducted into the National Football Foundation and the College Football Hall of Fame. Osborne's 84% winning percentage as a head coach ranks among the best in the history of college football and he was named the greatest college football coach of all time by an ESPN poll (Osborne, 2009).

**Research Hypotheses**

I hypothesized that Osborne would present a unique and enlightened approach to recruiting based on both physical attributes and psychological attributes of players. I also hypothesized that Osborne's perspective would advance my understanding of the subject matter beyond what was learned from meta-analyses of published literature (Chapter 2) and NFL scouting perspectives (Chapter 3).

## Results

*Recruiting Processes*

1. How important is recruiting to a college football team?

Many people are aware that good coaching is extremely important to a college football team. However, having talented athletes is at least as important as good coaching, if not more important. A great coaching staff may have a difficult time consistently winning football games if their players lack talent. Recruiting is extremely important to every college football program.

2. Who has the final say on recruiting matters?

The head coach has the ultimate say. But, within successful programs, responsibilities are delegated so that coordinators and position coaches also play a significant role. Different coaches on the staff have different regions of the country that they are responsible for recruiting and a head coach may visit particularly talented players anywhere. Every coach has a role and recruiting efforts are often joint efforts.

3. Please outline the recruiting process.

Division I football programs typically send cards out to high school football coaches all over the country. High school coaches indicate which of their players may be talented enough to play college football and return the cards to the respective collegiate programs. A college coach may then request game film from the high school coach in order to observe a particularly promising player during his sophomore and junior seasons.

Spring months are spent visiting various high schools around the country. When official visits to campus are arranged, the head coach, position coach, and regional recruiter (if different from the position coach) typically meet with the player and his family. Different coaches sell their programs in different ways (e.g., facilities, academics, coaching, etc.). If the prospect seems to be a good fit for the program, representatives from the coaching staff continue to make contact with the player approximately once per week until signing day at which point the high school player may sign a national letter of intent to attend the institution of his choice.

4. Do most college football programs understand the importance of recruiting?

It has been said that college football coaches experience two equally important seasons each year. One season when games are played and another season filled with recruiting activities. Both can be equally time- and labor-intensive, and both can have serious implications in terms of the ongoing success of the program. I think everyone understands the importance of recruiting. Most programs are very conscious of it and continually work to gain an advantage in the area of recruiting.

5. How is the college football recruiting landscape different now than when you were coaching?

Recruiting services and general media attention has greatly changed the landscape of college football recruiting. Today, high school players are bombarded with inquiries from private companies and there is a lot of pressure to entertain social media. Prospects may not feel like taking additional phone calls from college coaches after being consistently hassled, even though a coach’s perspective matters much more than any recruiting service.

*Talent*

1. What does it take to play football at the highest level?

Players who go on to play in the NFL tend to be pretty talented. The NFL is interested in drafting players who can play right away. That is a different approach than recruiting at the college level. At the collegiate level, coaches can search for players with potential and develop them within their programs.

2. How well do NFL scouts evaluate talent?

NFL scouts are only average, at best, at evaluating talent. Some are pretty good, but others are not. The NFL Combine is an opportunity for scouts to take a look at everyone and they try to conduct testing that will reveal a player's potential. But, there are flaws in the process. For example, the Wonderlic Test, which some NFL scouts rely upon to assess football ability, was designed for a completely different purpose and probably shouldn't be used to evaluate football players.

3. Which of your teams was most talented?

The Nebraska Football team that won the 1995 National Championship was the most talented team that I ever coached. Our coaching staff didn't necessarily realize how good they were going to be when we were recruiting them, but they developed into a really exceptional group. Most all of the players on our starting defense also started at one time or another in the NFL, and about 2/3 of the starting offense. At some positions, even backups had extended NFL careers. Not only were we talented, but we had exceptional team chemistry and toughness as well.

4. What are some indications that a team is talented?

Teams that are big, fast, and strong at many positions are typically considered talented. Speed and physicality of play are very evident when observing game film of a talented team. That style of play is very imposing on opponents. But, just because a team is extremely talented doesn't mean that they will go all the way.

5. What are some indications that a team has talent but lacks other essential characteristics necessary to be a great team?

Penalties, for example off-sides penalties or late-hits, are indicative of an undisciplined team that may struggle to win even with talent. Players talking back to coaches can indicate that a team lacks cohesiveness. There is also an element of tenacity and toughness that great players and teams must possess. How does a team perform in the $4^{th}$ quarter when they are behind?

*Evaluation*

1. How did you learn to evaluate players?

I have spent a majority of my life in various capacities within the game of football so my perspective has probably been shaped by characteristics of the game. Unfortunately, there is no tutorial or coach's manual on the topic of evaluating players. Most coaches evaluate their most successful players over the years and look for similar characteristics in recruits. I suppose the opposite may also be true for players who didn't pan out.

2. Why are you able to understand players better than most?

I have some background in psychology and that probably influences the way I treat people. I'm not sure how well I read people, but I believe that most people do not intentionally disrupt productivity or the greater good. There is typically a reason that people act out (e.g., act lazy, disrespectful, etc.). I think it has been helpful for me to interpret behavior from that perspective.

3. How has being a football coach helped you understand people better?

During any given football season, various ups and downs are experienced. Injuries are sustained, players shift roles, and important games are won and lost. The extraordinary highs and lows associated with the game of football expose more primitive traits. A lot can be learned about anyone who is part of a football team enduring the trials and tribulations of a season.

4. How has your educational psychology background helped you understand people better?

I try to ask questions rather than criticize. If you come at it from that angle, it changes the way you treat people and you are able to learn much more. Players can tell that you really care about them and they work much harder.

5. Do you believe that using data collection processes or intuition is a better approach to selecting players?

I think that a combination of data collection and intuition must be used to fully analyze a player's potential. Some physical characteristics of a player can be measured such as a 40-yard dash time or vertical jump. On the other hand, some attributes cannot be readily measured. For example, there is no scale for how respectful players are or how tough they will be. To some extent one can observe the way players respond to their parents during an official visit or ask guidance counselors about the character of players. You can watch game film to get some sense for how tough a player is. But ultimately, intuitive judgments will have to be made about some of the most important psychological characteristics of a player.

*Physical Attributes*

1. What are the most important physical characteristics to evaluate when judging one's potential as a football player?

It depends on the position. Height is important for linemen, but it's less important for other positions. Height isn't very important at all for running backs. Weight matters most for linemen. Agility and acceleration are important for all positions, and top speed is important for all positions except linemen. Upper-body strength is important for many positions, but not as important for quarterbacks, defensive backs, and wide receivers. A kicker really doesn't need much beyond leg strength.

2. How did you measure physical characteristics of players?

As part of the recruiting process, test times and weights are reported (e.g., 40-yard dash, bench press, etc.). But, you have to be careful; just because a player reports a great 40-yard dash time, someone somewhere else might be timing it in a different way. It's hard to compare players sometimes. It may be that a player once ran a 4.4 second 40 yard dash but typically covers that distance in between 4.7 – 4.8 seconds.

Boyd Epley was a successful strength and conditioning coach. He and his staff developed the Performance Index to manage physical indicators. The Performance Index is basically an objective scale that allows you to evaluate players at each position. It hits on all of the main performance indicators and is probably the best way to interpret test scores for physical ability. The Performance Index is also a tremendous motivator. Our players really competed to get their names on the wall or to be able to train on a special platform if they scored a certain number of Performance Index points.

3. Are physical characteristics of players generally attributable more to genetics or the environment in which one was raised?

There seems to be a strong genetic component to physical characteristics of a player. Some players are naturally more fast-twitch than others. Many programs recruit the siblings of successful athletes. However, the pattern doesn't always hold.

4. Are there any physical characteristics of players that you consider immediate disqualifiers?

Poor footwork is an immediate red flag. Players who lack agility also tend to struggle. You also look for good speed at many positions.

5. How did you develop physical characteristics of players once they were in your program?

Players could be expected to mature quite a bit physically once they joined our program. A good strength and conditioning program really helps with that. Most are aware of Husker Power and the influence Boyd Epley and his strength and conditioning staff had on our program.

*Psychological Attributes*

1. What were the most important psychological characteristics to evaluate when judging one's potential as a football player?

In all people, I value integrity. This involves telling the truth and keeping promises. It's easy to understand how an organization that consists of people with integrity can trust one another. If a player is given an opportunity to cheat, how will he handle it? Toughness and resiliency are also particularly important attributes in football players due to the ups and downs associated with a football season. Self-discipline is important as well.

2. How did you measure psychological characteristics of players?

Game film can help you understand how well a player performs under pressure. And, does he take plays off? But unfortunately, like I said, a lot of psychological evaluation for coaches is intuitive.

Years ago we worked with psychologists from the Gallup Organization who were part of the positive psychology movement to try to develop a psychological profiling tool. Several characteristics have been identified as positive psychological characteristics. I believe those qualities can be leveraged to improve performance.

The initiative ultimately fell by the wayside because we didn't want to be the only school asking recruits to remain on the phone for an extra 20 minutes in order to answer questions. But, I think this is one place in particular that coaches can really improve. I think there is a real opportunity to improve our understanding of players by studying psychology. It's something that needs to be done.

3. Are psychological characteristics of players generally attributable more to genetics or the environment in which one was raised?

Psychological traits are probably a little bit of both. The group of friends that player spends his time with can have a big influence on him. It's important to associate with a positive group of peers.

4. Are there any psychological characteristics a player might exhibit that you consider an immediate disqualifier?

If a player has trouble with authority, it is a red flag. If a player has ever been kicked off a team or been in trouble with the law, it may be an indication that he has trouble dealing with authority. Failed drug tests or academic fraud are also a concern.

5. How did you develop psychological characteristics of players once they were in your program?

The culture of your program has a big influence on your players. Players want to know that coaches truly have their best interest at heart. Each recruiting class will have 5 or 6 players who do not make it for one reason or another (e.g., grades, drugs, trouble with the law, etc.). Typically these players don't make it for psychological reasons; it isn't because they are not talented enough. If there were a way to better evaluate the psychology of players and make the number of players lost from a given class only 2 or 3 instead of 6, it would make quite a difference over a 4 or 5 year period.

**Discussion**

Dr. Tom Osborne had extraordinary success as a college football coach. He won 3 National Championships with a program that had some inherent recruiting disadvantages. Although Osborne had experience evaluating physical characteristics of great players, he suggested that psychological factors are equally important. Osborne's background in educational psychology gave him a unique, more complete perspective with respect to what it takes to be successful in the game of football.

Osborne argued that recruiting talented players may be even more important than coaching. He suggested that most programs recognize the importance of recruiting and delegate responsibilities to all members of the coaching staff in order to manage the formidable process of gathering information, etc. He also acknowledged the way in which the current recruiting landscape is different from what he experienced, primarily due to media attention.

Osborne felt that a distinction must be made between selecting players to play in the NFL and selecting players to play at the collegiate level. NFL players are expected to play immediately whereas collegiate players can be expected to develop after joining a program. Osborne suggested that some NFL scouts may not be particularly skilled at identifying talent and cautioned that the NFL Combine process may be flawed in some respects.

Osborne didn't seem to agree with NFL scouts (*Chapter 3*) regarding the importance of the vertical jump test, although he did mention agility (20-yard shuttle) and speed (40-yard dash) as good indicators of physical ability at a number of positions. He also mentioned height and weight as factors worthy of consideration when evaluating linemen. He felt that the importance of different physical test scores varied depending on a player's position.

When evaluating physical characteristics of players, Osborne cautioned against relying too heavily on test scores. He did however encourage using the Performance Index, as developed by Nebraska Strength and Conditioning Coach Boyd Epley, when traditional physical performance prediction metrics need to be interpreted. He acknowledged a genetic component to physical attributes, red flagged bad footwork, and emphasized the importance of a good strength and conditioning program so that players can develop physically once joining a program.

Osborne values integrity in all people. He also values toughness, especially in football players. Toughness is necessary to endure the immense challenges of a football season. While admitting that evaluating psychological attributes tends to be a somewhat intuitive process, Osborne made attempts as a head coach to scientifically study positive psychological constructs in football players in order to better capture the essence of what it takes to be successful.

For any given recruiting class, 5 or 6 players will not make it due to psychological reasons. According to Osborne, psychological traits are just as important to develop as physical characteristics. He acknowledged that great teams have toughness and tenacity to persevere and prevail. Osborne commented that even the most talented team he ever coached (i.e., the 1995 National Champion team) had an element of toughness, in addition to talent, which allowed them to thrive.

# Chapter 5

# PERFORMANCE INDEX:

# AN ENHANCED EVALUATION PROCESS

*Husker Power History*

Boyd Epley is a pioneer figure in the history of organized strength and conditioning for football. He was a pole-vaulter at the University of Nebraska in the late 1960s but suffered a career ending back injury. While rehabilitating himself, he spent time working with injured football players. Epley drew on his experience with bodybuilding, Olympic weightlifting, and power lifting to develop a training program that came to be known as Husker Power.

The Husker Power program emphasized multi-joint exercises using free weights rather than weight training machines, which often restrict motion. Epley initially trained a small group of injured athletes in a weight room at Memorial Stadium that was only 424 square feet; there was room for only 12 players to train at any given time. There was only one bench press, with one barbell, and 390 lbs. of weight.

Dr. Tom Osborne was an assistant football coach at Nebraska at the time and noticed that injured football players who had spent time training with Epley tended to return to the playing field in better condition than they had been in prior to sustaining injuries. Players appeared to improve both in terms of size, strength and speed. This was surprising considering conventional wisdom held that weightlifting made players stiff, slow and clumsy.

Osborne took a chance on intuition and recommended to Head Coach and Athletic Director, Bob Devaney, that the University of Nebraska hire Epley full-time so that he could coordinate workouts for the entire football team. On August 15, 1969, Devaney made Nebraska the first school in the Big 8 Conference to hire an official strength coach (Epley), adding the now infamous statement, "If anyone gets slower, you're fired." The weight room was increased in size to 1,344 square feet, and then to 2,844 square feet to accommodate the Husker Power program.

Devaney's teams went 42-2 and won the 1970 and 1971 National Championships in the first four years after adding Epley to their coaching staff. Epley went on to found the National Strength and Conditioning Association (NSCA) in 1978. He was awarded the first ever National Strength Coach of the Year Award in 1980, and the first ever Lifetime Achievement Award by the NSCA. In March 2001, American Football Monthly magazine tabbed him the godfather of Strength and Conditioning. Such a title was fitting considering that 49 of Epley's assistants went on to work for other programs during his 34 years as Head Strength and Conditioning Coach at Nebraska.

Today, Division 1 college football weight rooms may range in size up to 100,000 square feet or greater, and strength and conditioning is an essential aspect of any football program (Arthur & Bailey, 1998). Although many of Epley’s Husker Power core principles have become common practice among strength and conditioning coaches around the country, one major aspect of the Husker Power program has remained largely unrecognized. The Performance Index is believed by many to have had a significant impact on the Nebraska Football Program.

Beginning in the 1970s, Epley began a rigorous process to objectively measure the physical performance of Nebraska football players using various test scores. These tests included the 10-yard dash, 40-yard dash, 20-yard shuttle, vertical jump, hang clean, back squat, and bench press. Ironically, most of the tests on which Epley relied for information are still in use today *(Chapters 2 & 3*) despite the fact that the reasons for why many of them came to be used is largely an historical accident and merely a matter of initial convenience.

*Performance Index*

Epley worked with Mike Arthur and Dr. Chris Eskridge to develop the Performance Index as a decathlon-type scoring system that rewards individuals points on a 1 – 1000 scale based on physical test scores (e.g., a player runs a 10 yard dash in 1.7 seconds and receives 510 points, or a player hang cleans 320 pounds and receives 470 points). The point system uses a standard normal distribution and is based on the performance of over 30,000 athletes from many different sports, over 25 years, who performed each respective test. A world-record-level performance corresponds with a score of 1000 points. A top-level Division 1 athlete would be expected to score 700 points on a given test. A score of 500 points is considered an above average landmark for which all players should strive, and a score of 440 points represents the average score among Division 1 athletes for each of the tests.

It stands to reason that, as a player improves, continued improvement of the same magnitude becomes exponentially more difficult to accomplish. Furthermore, a player who performs slightly better than an above average athlete is more impressive than a player who performs slightly better than a below average athlete. Therefore a power curve was used that rewards points on an increasing basis the closer an athlete gets to a world record performance (e.g., only 1 additional point is rewarded to a player for improving .2 seconds in the 40-yard dash if his time improves from 6.0 to 5.8 seconds, whereas 150 additional points are rewarded a player for improving his time from 4.6 to 4.4 seconds). Most coaches and athletes alike can intuitively understand this concept and can relate with experience.

The Performance Index controls for body weight. A 300 pound player wouldn't be expected to run as fast as another who weighs only 185 pounds. Conversely, one wouldn't be expected to lift as much weight weighing 185 pounds as another weighing 300 pounds. In the unlikely event that a 300 pound player can run as fast as a 185 pound player, the larger player is considered much more impressive and effective on the football field than the smaller player, and thus ought to be rewarded more points. A separate curve was developed for each 5-pound weight class based on bench-mark performances for each of the specific weight classes.

To extend this concept, the Performance Index rewards performance on strength and functional tests relative to body weight so that heavier players must perform better than lighter players in order to receive the same number of points on tests of strength and lighter players must perform better than heavier players in order to receive the same number of points on functional tests. This orientation allows for objective comparisons, not only among players of the same size, but among players of any size. Such a feature is essential for predicting on-field performance among players at different positions (e.g., to compare the potential of offensive linemen to running backs).

Using historical data as bench marks, scales were externally adjusted so that 500 points would be equivalent, relative to the distribution, among the scoring systems for each physical test. The Performance Index is viewed much like an ACT-test for general academic aptitude in the sense that it is a general pound-for-pound measurement of physical performance ability. A Performance Index exists for lean body mass, hang clean, back squat, bench press, jerk, snatch, 10-yard dash, 40-yard dash, 20-yard shuttle, and vertical jump. Despite the fact that most of the tests that Epley used to collect information on players were decided upon as a matter of convenience, many of the same tests are still in use today. In the words of Dr. Tom Osborne, “The Performance Index hits on all of the main performance indicators and is probably the best way to interpret test scores for physical ability.”

# Method

## Procedure

This study was facilitated, over 2-consecutive years, by strength and conditioning coaches at the University of Nebraska. They agreed to provide Performance Index scores for all physical tests conducted as part of regularly scheduled summer performance testing as well as on-field performance evaluations for each player at the conclusion of the season, which immediately followed summer testing. Data collection processes were identical from the first year to the second.

On day 1 of summer training (early June), players reported throughout the day to complete body composition analyses, which were conducted by the team's nutritionist.

On day 2 (the following day), players arrived for testing around 6:30am. After providing saliva samples beginning at 7:00am (*Chapter 6*), the players began to warm up at 7:10am. The warm up consisted of 10 sprint-mechanic drills conducted over the course of 20 yards and took approximately 10 minutes.

After completing the warm up, players were informed of testing procedures and then completed 2 attempts at both the 10-yard dash and 20-yard shuttle run. All players had finished this phase of testing by 7:30am, at which point they provided more saliva samples (*Chapter 6*). Players then proceeded to the weight room where they were tested in the vertical jump. Day 2 testing for all players was complete by 9:00 a.m.

During the team's summer training months (i.e., June and July), players were introduced to the hang clean, back squat, and bench press exercise in the weight room. Players had an opportunity to improve their technique and received instruction from members of the strength and conditioning staff as they performed each exercise.

Throughout the summer, players were given an opportunity to earn a “bump” (i.e., a recorded increase in one-repetition max weight), for each exercise (i.e., hang clean, back squat, and bench press) when completing "max-effort sets". A max-effort set required players to perform as many repetitions for a given exercise as possible using a prescribed weight. Players earned a "bump" based on assumptions that an athlete can perform 80% of their one-repetition max weight 8 times, 85% of their one-repetition max weight 5 times, and 90% of their one-repetition max weight 3 times. For example, if a player performed 5 repetitions of the hang clean with 225 pounds during a max-effort set, then his one-repetition max weight for the hang clean would be estimated to be 265 pounds (85% of 265 pounds = 225 pounds). When athletes performed 2, 4, 6, 7, or more than 8 repetitions, an estimate was made based on the 3-, 5-, and 8-repetition reference points. For example, if a player performed 4 repetitions of the hang clean with 225 pounds during a max-effort set, then his one-repetition max weight for the hang clean would be estimated to be 257 pounds (87.5% of 257 pounds = 225 pounds).

Depending on the number of repetitions that a player performed during a max-effort set, his one-repetition max was reconfigured. All players increased their one-repetition max weights throughout the summer, and final test scores for strength measures (i.e., hang clean, back squat, and bench press) were not recorded until the last week of the summer training phase. The one-repetition max weight, for each strength training exercise (i.e., hang clean, back squat, and bench press), that each player had recorded after the final week of summer training was used in data analyses.

## Measures

### *Anthropometric Measure*

- Lean Body Mass

Lean body mass (i.e., muscle, bone, ligament, tendon, etc.) is a measure of body mass that excludes fat mass. The team's nutritionist measured the bodyweight of each player using a standard scale. Then, he used skin calipers to conduct a 7-site (i.e., chest, tricep, subscapular, axilla, suprailiac, abdomen, and thigh) skin fold test in order to calculate the body fat percentage of each player. Each player's fat mass was calculated by multiplying his body fat percentage by his total body weight, and lean body mass was calculated by subtracting a player's fat mass from his total body weight. Lean body mass was recorded to the nearest pound. Members of the strength and conditioning staff converted each player's lean body mass measurement into a Performance Index score.

*Strength Measures*

- Hang Clean

The hang clean is a measure of lower-body explosiveness. Players began with their feet hip-width apart and held an Olympic bar in the hang position. Using one continuous motion, players initiated movement into a power position by slightly bending their knees as they pushed their hips back and shoulders in front of the bar so that the bar slid down their legs to a position just above the knees. From the power position, players aggressively jump-shrugged so that the bar accelerated to a vertical height just under the chin. Then players quickly dropped under the bar, into a quarter front squat position, so that the bar could be caught on top of the anterior deltoids. Squat cleans (i.e., a player drops down into a deep squat position in order to catch) were not desired. A repetition was considered valid if the bar successfully attained the proper catch/rack position. Lifting straps were used by most players. Members of the strength and conditioning staff converted each player's hang clean one-repetition max weight into a Performance Index score.

- Back Squat

The back squat is a measure of lower-body strength. Athletes began with their feet shoulder-width apart and held a standard bar on top of their posterior trapezious muscles with their hands gripping the bar, close but comfortable, outside of the shoulders. Using one continuous motion, athletes initiated descent by pushing their hips back and bending their knees until they reached the full-squat position with their femurs parallel to the ground while keeping the chest erect. From the full-squat position, athletes ascended, keeping their chest erect, until they had returned to the starting position. A repetition was considered valid if an athlete attained the full-squat position during the descent, and then returned to an upright position with legs fully extended. Feet were to have remained in full contact with the ground throughout the entire range of motion. Lifting aids were not used (e.g., belts, knee wraps, etc.). Members of the strength and conditioning staff converted each player's back squat one-repetition max weight into a Performance Index score.

- Bench Press

The bench press is a measure of upper-body strength. Athletes began in a supine position on a standard bench press bench with their feet flat on the floor. Hands were placed comfortably on a standard bar and players were given a "lift off" in order to get to a starting position with arms locked out directly above the chest. Players lowered the bar until it touched the chest, and then pressed to return the bar to the starting position. No pause at the chest was required, but the bar was not to have been bounced off the chest.

A repetition was considered valid if the bar touched the chest before it was pressed and the arms were fully extended to complete the motion. Feet were to have remained flat throughout the lift, and the buttocks were to maintain contact with the bench. Lifting aids were not used (e.g., wrist wraps, etc.). Members of the strength and conditioning staff converted each player's bench press one-repetition max weight into a Performance Index score.

*Functional Measures*

- 10-Yard Dash

The 10-yard dash is a measure of acceleration. A strength and conditioning coach held an electronic start switch and an electronic timing device was placed at finish line (i.e., a light beam apparatus), which was 10 yards from the start. Starting from a three-point stance, players commenced a 10 yard sprint on their own volition, at which point the strength and conditioning coach released the start switch to begin timing. The time was automatically stopped when players finished through the light beam at the finish line. Each player was given 2 opportunities to run the test, and the fastest time was recorded to the nearest one-hundredth of a second. Members of the strength and conditioning staff converted each player's 10-yard dash time into a Performance Index score.

- 20-Yard Shuttle

The 20-yard shuttle is a measure of agility. A strength and conditioning coach held a stop watch at the 15-yard line and began timing when a player initiated movement. Players began in a three-point stance, straddling the 15-yard line, with their right hand touching the ground. On their own volition, players ran to their right to touch the 20-yard line with their right hand, then sprinted back past the 15-yard line and touched the 10-yard line with their left hand. Finally, they sprinted back (in the initial direction) through the 15-yard line, where they had started, to finish. The strength and conditioning coach stopped the stop watch when the player passed through the 15-yard line to finish. Each player was given 2 opportunities to run the test, and the fastest time was recorded to the nearest one-hundredth of a second. Members of the strength and conditioning staff converted each player's 20-yard shuttle time into a Performance Index score.

- Vertical Jump

The vertical jump is a measure of jumping ability. A Vertec was used for this measurement. Players began with their feet hip-width apart and flat on the ground. They reached as high as possible with their dominant hand so that a standing reach could be recorded based on the highest Vertec vane that could be reached. To measure vertical jump, players stood flat-footed in front of the Vertec and, in a countermovement fashion (i.e., players quickly dipped into a quarter-squat position before jumping), jumped as high as possible, reaching with their dominant hand to swat the highest vane they could reach. Vertical jump height was calculated by subtracting players' standing reach height from the height of the highest vane moved. Each player was given 3 jumps, and the highest vertical jump was recorded to the nearest half inch. Members of the strength and conditioning staff converted each player's vertical jump height into a Performance Index score.

*Composite Measure*

- Total Performance Index Score

Performance Index scores for each test (i.e., lean body mass, hang clean, back squat, bench press, 10-yard dash, 20-yard shuttle, and vertical jump) were added together to generate a Total Performance Index Score.

*On-field Performance Measure*

Players were judged by their coaching staff to have significantly contributed to the program during the season immediately following summer testing, or were deemed to have not significantly contributed, and were labeled accordingly. This on-field performance measure was deliberately left open-ended and subjective as suggested by members of the coaching staff. A coach's subjective measure of significant contribution can refer to starting in a game if one is talented enough to receive such an opportunity, significantly contributing as a backup or reserve, or even performing well as a scout team player in preparation for the week's game.

Ultimately, the coaching staff reasoned, the most important measure of a football player is whether he significantly contributes on the field, and that can occur in many forms. Very few freshman football players actually play in games. However, many are significant contributors to the program. The subjective rating of on-field performance utilized in this study allows freshmen, among others, to be categorized and better captures the essence of what coaches truly care about than some other more objective measures of on-field performance.

**Subjects**

A total of 47 freshman football players at the University of Nebraska, over a 2-year period, volunteered to participate in this study and completed all required tasks. All players were 19 year old males. Their average height was 73 inches and the average weight of players in the sample was 233 pounds. A total of 29 players indicated that they were Caucasian, 17 said they were African American, and 1 player identified with a different ethnicity than Caucasian or African American.

Recruitment took place at the team's training location (Memorial Stadium in Lincoln, NE). All participants gave informed consent to participate in the study prior to their first day of data collection. Participants received no compensation.

**Research Hypotheses**

I hypothesized that players who were significant contributors on the football field during the season following summer testing would also have earned more Performance Index points for all tests of physical ability and the Total Performance Index Score than players who did not contribute significantly. In order to test these hypotheses, between-groups ANOVA analyses were conducted to compare the mean number of Performance Index points accumulated by each group of players (i.e., contributors vs. non-contributors) for each of the physical performance tests conducted by the team's strength and condition staff during summer testing as well as the Total Performance Index Score.

## Results

A total of 18 players were judged by the coaching staff to have significantly contributed on the football field during the season. The coaching staff indicated that 29 of the players had not significantly contributed on the football field during the season. There was a significant difference between contributing and non-contributing players in the 10-yard dash and vertical jump. Differences in the 20-yard shuttle, bench press, and Total Performance Index Score approached statistical significance. There was no difference between players who had contributed and players who had not contributed on the field in lean body mass, hang clean, or back squat Performance Index scores.

**Table 5.1 Performance Index Scores. Data are mean (standard deviation).**

| | **Contributing (n = 18)** | **Non-Contributing (n = 29)** | **p-value** |
|---|---|---|---|
| *Anthropometric Metric* | | | |
| Lean Body Mass | 275.01 (20.52) | 260.25 (25.7) | p = .759 |
| *Strength Metrics* | | | |
| Hang Clean | 337.26 (75.27) | 334.67 (62.36) | p = .899 |

| | | | |
|---|---|---|---|
| Back Squat | 387.69 (51.44) | 377.83 (43.31) | p = .484 |
| Bench Press | 444.13 (52.73) | 413.58 (59.19) | p = .080** |
| *Functional Metrics* | | | |
| 10-Yard Dash | 490.38 (66.82) | 428.13 (76.77) | p = .007* |
| 20-Yard Shuttle | 349.48 (93.34) | 307.27 (86.94) | p = .123 |
| Vertical Jump | 442.33 (66.42) | 379.24 (60.98) | p = .002* |
| *Composite Measure* | | | |
| Total Index Score | 2745.69 (427.95) | 2525.77 (371.48) | p = .069** |

** indicates p-value <.05*

***indicates p-value <.1*

## Discussion

The 10-yard dash and vertical jump were the only two physical tests that proved to be predictive of performance on the football field for this sample of players. This is surprising since each test conducted as part of this study is considered by the team's coaches to be an effective discriminator. Players who contributed significantly on the football field exhibited a significantly greater mean Performance Index vertical jump score and a faster mean Performance Index 10-yard dash score than players who did not significantly contribute on the field.

Both the 10-yard dash and vertical jump occur frequently on the field of play. Players at many positions "practice" each test almost every time they step onto the field, and have done so ever since they began playing football. In addition, both tests are simple and straightforward; neither test is very technically demanding. Less evident learning curves associated with these tests may allow for the index scores to better represent one's physical ability than some of the other tests, which don't occur frequently on the field of play, or are more complex.

Mean differences between groups of contributing and non-contributing players in Performance Index points for the 20-yard shuttle and bench press approached statistical significance but did not quite meet the threshold. This was likely a result of limited power associated with including only 47 subjects. Statistical power for these analyses was between 0.5 - 0.6 (i.e., well below the desired standard of 0.8). Had 85 athletes been included in this study, a significant difference between groups for both tests would have been discovered given the same effect sizes.

Relative to the other functional tests evaluated in this study (i.e., the 10-yard dash and vertical jump), the 20-yard shuttle is the most technically sophisticated. Performing well seems to be a learned skill and players prepare diligently, especially in preparation for the NFL Combine, to master the test. NFL Combine attendants can be observed performing the 20-yard shuttle according to a wide variety of methods, especially with respect to covering the first 5 yards of the test (e.g., pushing off of the outside leg to begin vs. using a crossover step, or using a slide turn into the first cut vs. not, etc.). Many strength and conditioning coaches literally teach a specific number of steps between cuts, etc.

Players tend to significantly improve in the 20-yard shuttle as their legs become stronger during the course of their career. Great leg strength is needed, relative to one's bodyweight, to decelerate and to change directions quickly. The average number of 20-

yard shuttle Performance Index points scored in this study was only 323 compared to 459 for the 10-yard dash and 410 for the vertical jump. Since 440 Performance Index points represents an average score for a Division 1 athlete on each test, this population of players would be expected to improve more in the 20-yard shuttle than in the 10-yard dash and vertical jump.

Given these characteristics, it is not a surprise that the 20-yard shuttle is less predictive of performance on the football field among freshman football players than the 10-yard dash and the vertical jump. Increased error associated with using this measure to assess agility among freshman players likely reduces the utility of this test. An incoming freshman's score in the 20-yard shuttle reflects much more than his athleticism (e.g., familiarity with the test).

Relative to the hang clean and back squat, the bench press is a much less technically sophisticated strength training movement. The exercise isolates smaller muscle groups in the arms and the body is stabilized by the bench. Further, the motion tends to be more unidirectional than the hang clean and back squat as the bar tends to remain in one plane during the path of ascent. Similar to the 10-yard dash and the vertical jump, the movement is simple and straightforward.

Since the bench required to perform the bench press exercise is less expensive to purchase and takes up less room in a weight room than power racks and platforms used to complete the hang clean and back squat, high schools are more likely to have a bench press. Most athletes have learned the bench press exercise during their high school careers as many high school coaches are more familiar with the bench press than the hang clean or back squat and feel more comfortable teaching the less complicated movement. Freshman football players arrive on college campuses with somewhat equal experience and familiarity with the bench press exercise so test scores "truly" reflect upper body strength, and only upper body strength.

Mean hang clean and back squat index scores were not significantly different between groups of players who contributed and those who did not. Like many Olympic lifts (e.g., snatch or jerk), the hang clean is a neuromuscularly complex strength training exercise to perform. High school coaches are often uncomfortable teaching the hang clean to their athletes due to unfamiliarity. Therefore, many football players arrive on college campuses with no previous exposure to the hang clean. Such players then spend years improving their technique in order to become proficient in the movement.

While not as complicated to perform as the hang clean, the back squat presents other challenges when used as a measure of lower-body strength among freshmen athletes. Worse than having no exposure to an exercise at all, is having been taught improper technique. Many high school athletes develop bad habits when performing the back squat (e.g., going up on the toes during the descent phase or not descending to the full-squat position). Collegiate strength and conditioning coaches often spend the entire first year of training correcting bad habits for athletes who have been taught poor technique.

In extreme circumstances, student-athletes may develop severe muscle imbalances. For example, leaning forward on the toes during the descent phase of the back squat may increase strain in the quadriceps and lower back muscles. Chronic injuries from performing back squats with excessive weight using improper technique can occur (e.g., bulging disc) before an athlete even arrives at a university. Injured athletes may be extremely limited in terms of how much weight they will ever be able to use or possibly be unable to perform the back squat exercise at all.

Lean body mass is considered by the developers of the Performance Index to be among the best predictors of physical ability. In this study, lean body mass scores were not predictive of performance on the football field, but these results may be attributable to range restriction. Few freshman players have had an opportunity to develop the muscle mass exhibited by older, more physically mature players, which probably accounts for differences in on-field performance. Players in this study averaged only 265 Performance Index points for lean body mass and there was very little variation in scores. If 440 points represents the mean score for Division 1 athletes, then athletes in this study have a lot of room for improvement. Expectedly, the mean number of lean body mass Performance Index points for contributing and non-contributing players were almost identical.

General findings suggest that functional tests are superior to strength tests when distinguishing freshman football players who significantly contribute on the field during their first season of play from players who do not. In addition, less technically sophisticated neuromuscular movements better discriminate than those that are more complex. Both functional and strength tests probably work better than anthropometric tests, although only one anthropometric test was used in this study.

Lower-body explosiveness, relative to one's bodyweight, would allow a player to score well on both the 10-yard dash and vertical jump. One can conclude from the results of this study that lower-body explosiveness is among the most important qualities to consider when predicting performance on the football field. In a sense, the 10-yard dash and vertical jump are the purest of tests that made up the battery used in this study. These tests represent rote athleticism and ought to be included in any battery of physical tests being used to predict performance in the game of football.

One limitation of this study was the restricted number of physical test variables. The strength and conditioning staff at the University of Nebraska do not currently conduct the 40-yard dash as part of summer physical performance testing in order to avoid the unnecessary risk of soft-tissue injuries (e.g., a pulled hamstring). Such an injury to a key player could be quite detrimental to a team in the months immediately preceding the season. Consequently, there was no measure of top-end speed in this study from which to predict performance on the field.

When interpreting the results from this study, it is important to carefully consider the measure of on-field performance that was used. Players were judged by members of the coaching staff to have significantly contributed on the field during the season immediately following summer testing or not to have significantly contributed. This measure tells nothing about the long-term success of players in the program. Most coaches recognize that athletes change a great deal during the course of their college career and care much more about the overall contribution of a player during his entire career than what he may, or may not, be able to accomplish during his first season of play.

The Nebraska Football Program is one of the most successful in the history of college football and much of the success has been attributed to the Husker Power Strength and Conditioning Program. One major aspect of Husker Power, namely the Performance Index, has remained largely unrecognized despite the fact that it is believed by many to have had a significant impact. Current strength and conditioning coaches should consider reemploying this method of evaluation since, in the words of Dr. Tom Osborne, “The Performance Index hits on all of the main performance indicators and is probably the best way to interpret test scores for physical ability.”

# Chapter 6

# TOUGHNESS:

# MODEL COMPARISONS SUGGEST PREDICTIVE UTILITY OF BIOMARKERS

*Mental Toughness*

Athletes, coaches, and sport psychologists seem to agree that mental toughness is important because of its apparent relationship with successful performance (Crust, 2007). Countless manuscripts have been written on the topic, especially texts that purport to help one develop mental toughness (e.g., Bull, Albinson, & Shambrook, 1996; Goldberg, 1998; Loehr, 1986, 1995). However, despite considerable interest, there has been no consensus on the definition, conceptualization, and precise way in which to develop mental toughness (Crust, 2007).

Recently, a rigorous effort was undertaken that has contributed immensely to our understanding of mental toughness in athletics. Within the framework of personal construct theory (Kelly, 1955), Jones et al. (2002, 2007) studied mental toughness in elite athletes (i.e., gold medalists and world champions) from a variety of sports using qualitative methods. Jones et al. (2007, p. 247) defined mental toughness as “having the natural or developed psychological edge that enables you to, generally, cope better than your opponents with the many demands (competition, training, lifestyle) that sport places on a performer and, specifically, be more consistent and better than your opponents in remaining determined, focused, confident, and in control under pressure.” Jones et al.

(2007) outlined 4 separate mental toughness dimensions with 30 attributes: Attitude / Mindset (belief, focus), Training (using long-term goals as the source of motivation, controlling the environment, pushing yourself to the limit), Competition (handling pressure, belief, regulating performance, staying focused, awareness and control of thoughts and feelings, controlling the environment) and Post-Competition (handling failure, handling success).

This conceptualization of mental toughness has been endorsed by multiple research teams (e.g., Bull et al., 2005; Thelwell et al., 2005), but the paradigm is not without its limitations. Jones et al. (2007, p. 244) suggested that, since their definition of mental toughness "contains a dimension that relates to successful outcomes, mental toughness should be investigated in samples of athletes who have achieved ultimate success in their respective sports." Clearly an assumption has been made that elite athletes must possess extraordinary mental toughness in order to be extraordinarily successful. It is possible, however, that physical characteristics, even more than psychological attributes, enable some athletes to be successful.

Many researchers have focused entirely on elite athletes (e.g., Fourie & Potgieter, 2001; Jones et at., 2002; 2007; Thelwell et al., 2005). But, it may be an oversimplification to assume that *only* elite performers possess mental toughness. An alternative approach could emphasize mental toughness relative to one's ability. Bull et al. (2005) hinted at this concept with statements about having determination to make the very most of ability. Likewise, Loehr's (1995) definition of mental toughness includes the ability to consistently perform toward the upper range of one's talent and skill, regardless of the competitive circumstances.

A lack of objective measures might encourage the oversimplifications and generalizations that have plagued mental toughness research. Ironically, Jones et al. (2007) claimed to assume a relationship between mental toughness and successful outcomes precisely because "there is no validated measure of mental toughness" (p. 244). Qualitative methods have unquestionably generated rich descriptions about the way in which mental toughness can be developed, but there exists a need for quantitative measures, with more objective operationalizations, to enhance the field of mental toughness research.

Instruments have been developed to measure mental toughness (e.g., MT48 – Clough et al., 2002; PPI-A – Golby, Sheard, & van Wersch, 2007) and have been shown to be psychometrically valid and reliable. But, the culture of competitive athletics has well established traditions based on conventional wisdom, and rigid applied practitioners may be resistant to the use of self-report questionnaires based on hypothetical psychological constructs. Measures that correspond with previously conceived notions might be more appealing and gain more traction with coaches.

*Physiological Toughness*

A novel approach among some researchers has considered physical toughness a primary component of mental toughness. This conceptualization seems to make intuitive sense to many applied practitioners. In a seminal work, Dienstbier (1989) reviewed literature on human and non-human confrontations with stress and highlighted commonalities in physiological responding that were most likely to lead to successful outcomes. Specifically, Dienstbier coined the term “Toughness” to describe a specific hormonal response pattern that promotes success in myriad tasks. Dienstbier’s conjectures about the interrelationship between one’s physiological response to stress and performance are important to consider in the context of mental toughness (Crust, 2007).

Acute stress can be defined as the body’s homeostatic state being threatened or the perception of such threat. Such a state can cause physiological and psychological processes to become imbalanced, forcing systems to function outside of normal ranges. When an individual experiences physical stress or perceives psychological stress, a physiological stress response is initiated that is relative to the intensity, duration and type of stressor encountered (Charmandari, Tsigos, & Chrousos, 2005).

Stress affects human biological systems in two general ways. The first route is direct; hormones released by the biological stress response systems directly modify physiological functioning. The second route is indirect; stress hormones affect the brain, which may lead to changes in psychological states (e.g., emotions), which may result in modifications of behaviors (O’Connor, Jones, Conner, McMillan, & Ferguson, 2008).

The autonomic nervous system is responsible for regulating the release of catecholamines like epinephrine and norepinephrine. This system is known for its fast-acting, but short-lived effects. Sympathetic nerves originating in the spinal cord are stimulated by the neurotransmitter acetylcholine. Organs such as the heart, skeletal muscles and gut are innervated by sympathetic fibers; when activated, the fibers directly release epinephrine and norepinephrine at target sites. Sympathetic fibers also innervate the adrenal medulla, which can secrete epinephrine and norepinephrine into circulation and amplify the widespread effects of the autonomic stress response (Habib, Gold, &

Chrousos, 2001). Epinephrine contributes to arousal in many ways, but stimulating the release of glucose and facilitating utilization for energy is a primary function.

Another principle system responsible for the widespread effects of stress in the human body is the hypothalamus-pituitary-adrenal (HPA) axis, a neuroendocrine system that regulates secretion of the glucocorticoid cortisol (Tsigos & Chrousos, 2002). This system is known to be slow-acting, but long-lasting. First, neurons of the hypothalamus synthesize and release corticotrophin-releasing hormone (CRH), which stimulates the anterior pituitary to secrete adrenocorticotropic hormone (ACTH) into the blood stream. Circulating ACTH then stimulates the adrenal cortex to synthesize and release glucocorticoids into general circulation. The hormones cortisol and corticosterone are the principle glucocorticoids in humans.

Circulating glucocorticoids have widespread effects on target tissues, such as those in the skeletal muscles, cardiovascular system, immune system, and brain (Munck, Guyre, & Holbrook, 1984). The actions of glucocorticoids temporarily have catabolic, anti-growth and anti-reproductive effects, all as a part of the body's effort to increase blood glucose levels to help the individual power through stressors (Charmandari, et al., 2005). Like the autonomic nervous system, cortisol stimulates energy. However, its contribution to energy comes at some costs, such as immune system suppression and toxic effects to the hippocampus.

The autonomic and HPA-axis systems are complementary. There are many sites within the autonomic and HPA-axis response systems that interact (Charmandari et al., 2005). But, although the two systems coordinate to produce a stress response, they do not necessarily act symmetrically (Granger, Kivlighan, El-Sheikh, Gordis, & Stroud, 2007). The dominance of one system versus the other depends on the time-point during stress exposure as well as the context and type of stressor.

Folkman and Lazarus (1985) suggested that the term *stress* implies "a relationship between the person and the environment that is appraised by the person as relevant to his or her well-being and in which the person's resources are taxed or exceeded" (p. 152). Folkman and Lazarus also distinguished appraisal components of *challenge:* involving "potential for growth" vs. *threat:* perceived as involving "potential for harm or loss" (p. 152). Interestingly, one's physiological response to stress seems to depend directly on one's appraisal of a situation (Dienstbier, 1989).

Karolinska researchers (e.g., Frankenhaeuser, 1979) and others have shown that one's appraisal of a situation as *challenging* leads to arousal of the autonomic nervous system, whereas perceptions of *threat* lead to a combination of autonomic and HPA-axis arousal. Other researchers have shown that, when stressors are perceived as uncontrollable or unpredictable, HPA-axis reactivity is increased (e.g., Henry & Ely, 1975; Henry, 1997; Dickerson, Gruenewald, & Kemeny, 2004). Beltzer (2008) conducted a sport-specific investigation involving college-aged hockey players. Autonomic activity and HPA-axis responding during a practice condition were compared to a game condition response. Players demonstrated consistent autonomic responding to both conditions, but increased HPA-axis response to the game condition. This response pattern was attributed to the added components of uncontrollability, unpredictability, and social evaluation associated with the game condition.

In relation to performance, early studies involving animals suggested a positive adaptive value associated with strong and fast autonomic arousal (Pfeifer, 1976). HPA-axis activity, by contrast, depressed neural functioning (Barnes, 1986), especially in the brain where marked hippocampal cell deterioration was evident (Meaney et al., 1987). Increased catecholamine levels have also been predictive of successful performance in humans for a variety of tasks. Scandinavian researchers have consistently found catecholamine increases to be positively related to performance, even in complex tasks. By contrast, high levels of cortisol appear to correspond with poorer performance (e.g., Ursin, Baade, & Levine, 1978; Vaernes, Ursin, Darragh, & Lambe, 1982).

Further support for the distinction between autonomic responding and HPA-axis responding has been revealed by factor analyses of responses to complex situations. Ursin et al. (1978) measured serum and urinary catecholamine and cortisol levels in relation to performance among Norwegian Army paratroopers. Exemplary performance in written evaluations and jumping corresponded with the catecholamine factor along with, on high-activity days, blood glucose levels. The cortisol factor, on the other hand, was correlated with poor performance in jumps from the training tower and the aircraft. Separate catecholamine and cortisol factors were also identified in recruits engaged in basic training (Rose, Poe, & Mason,1967) and among United States Navy recruits training in a swimming pool (Vaemes, Ursin, Darragh, & Lambe, 1982).

Results from both human and non-human studies suggest that high levels of cortisol reactivity are indicative of a lack of appropriate responses (Dienstbier, 1989). A *toughening effect* is manifested in neuroendocrine system modifications whereby one's capacity to generate norepinephrine and epinephrine is enhanced. A toughened individual's increased capacity to secrete catecholamines leads to delay and/or suppression of HPA-axis responses. An ideal pattern of cortisol arousal requires an ability to suppress the cortisol response (Dienstbier, 1989).

Multi-source measures are important to establish a more accurate measurement of mental toughness. Evaluating mental toughness in relation to stress biomarkers (e.g., cortisol) might help us better understand the way in which mental toughness operates. Specifically, it could be hypothesized that mental toughness is characterized by an ability to respond to stressors more effectively with less cortisol reactivity (Crust, 2007).

## Method

### Procedure

This study was made possible, over 2-consecutive years, by strength and conditioning coaches at the University of Nebraska. Members of the strength and conditioning staff facilitated saliva sample collection during summer physical performance testing and provided on-field performance evaluations at the conclusion of the season that immediately followed summer testing for each player. Data collection processes were identical from the first year to the second.

On day 1 of the team's regularly scheduled summer testing, players were instructed to abstain from eating, drinking (with the exception of water), or brushing their teeth, the following morning prior to providing saliva samples. On day 2 of summer testing, players arrived at team facilities around 6:30 a.m. Players provided saliva samples beginning at 7:00 a.m. (Time 1). To do so, each player obtained 2 microcentrifuge tubes (2 mL) along with a 2-inch straw and drooled saliva through the straw to fill each tube. After all players had provided Time 1 samples, they commenced warm up activities at 7:10 a.m. The warm up consisted of 10 sprint-mechanic drills conducted over the course of 20 yards and took approximately 10 minutes.

After completing the warm up, players were informed of all testing procedures for the day. Each player then completed 2 attempts at both the 10-yard dash and 20-yard shuttle run (*Chapter 5*). All players finished this phase of testing by 7:30, and then promptly provided another saliva sample (Time 2) in the same fashion as Time 1. Saliva samples were stored in a secure freezer at 0 degrees C until laboratory analyses were conducted by the University of Nebraska-Omaha Endocrine BioServices Assay Laboratory using an enzyme immunoassay (EIA) process under the direction of Dr. Jeffery French.

**Measures**

*Toughness*

- Cortisol Reactivity

As a long established indicator of HPA-axis responding, salivary cortisol measures are easily obtained and analyzed (Kirschbaum & Hellhammer, 1989), and reliably estimate blood cortisol levels (Aardal Eriksson, Karlberg, & Holm, 1998). Saliva was analyzed for cortisol using an enzyme immunoassay process. A 96-well plate was coated the day before analyses with a cortisol antibody and incubated overnight. The next morning, the plates were washed and a phosphate buffer solution was added to each well.

Saliva samples were thawed the morning of analyses. Each saliva sample was plated in duplicate. Once each well was treated with a steroid conjugate, the plates were incubated for three hours. After the incubation period was complete, the plates were washed and cortisol substrate was added to each well.

Absorbance was measured at 410 nm using a standard laboratory plate reader. The intra-assay coefficient of variation was calculated by comparing the duplicate samples, and the mean concentration was calculated by averaging the duplicate samples. Samples were re-assayed if the coefficient of variation exceeded 20%, or if the calculated concentration of cortisol exceeded 1200pg/50uL. Highly concentrated samples were re-assayed at a 1:2 dilution. Results were expressed in pg/mL. Salivary cortisol reactivity was calculated using the following formula: (Time 2 – Time 1) / Time 1.

*On-field Performance Measure*

Players were judged by their coaching staff to have significantly contributed to the program during the season immediately following summer testing, or were deemed to have not significantly contributed, and were labeled accordingly. This on-field performance measure was deliberately left open-ended and subjective as suggested by members of the coaching staff. A coach's subjective measure of significant contribution can refer to starting in a game if one is talented enough to receive such an opportunity, significantly contributing as a backup or reserve, or even performing well as a scout team player in preparation for the week's game.

Ultimately, the coaching staff reasoned, the most important measure of a football player is whether he significantly contributes on the field, and that can occur in many forms. Very few freshman football players actually play in games. However, many are significant contributors to the program. The subjective rating of on-field performance utilized in this study allows freshmen, among others, to be categorized and better captures the essence of what coaches truly care about than some other more objective measures of on-field performance.

**Subjects**

A total of 47 freshman football players at the University of Nebraska, over a 2-year period, volunteered to participate in this study and completed all required tasks. All players were 19 year old males. Their average height was 73 inches and the average weight of players in the sample was 233 pounds. A total of 29 players indicated that they were Caucasian, 17 said they were African American, and 1 player identified with a different ethnicity than Caucasian or African American.

Recruitment took place at the team's training location (Memorial Stadium in Lincoln, NE). All participants gave informed consent to participate in the study prior to their first day of data collection. Participants received no compensation.

**Research Hypothesis**

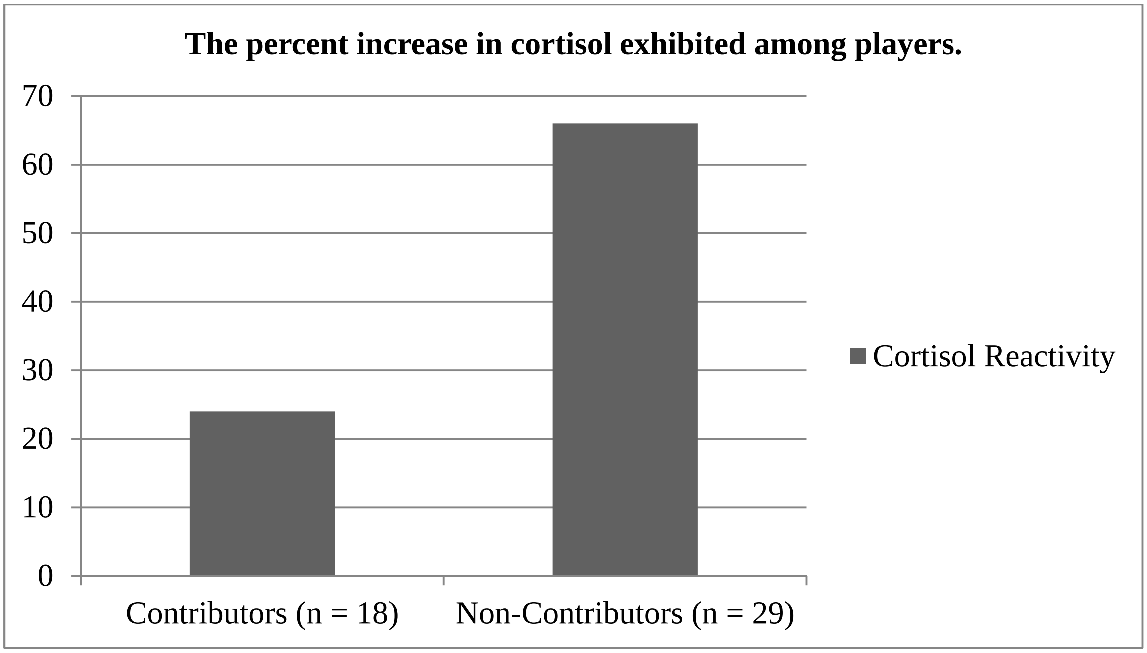
I hypothesized that players who were significant contributors on the football field would also have exhibited significantly less mean cortisol reactivity during summer testing than players who did not contribute significantly. I also hypothesized that a model, which included cortisol reactivity in addition to traditional physical performance prediction metrics (*Chapter 5*), would account for more variance in on-field contribution than a model that included only traditional physical performance prediction metrics.

## Results

Eighteen players were judged by the coaching staff to have significantly contributed on the football field during the season. The coaching staff indicated that 29 of the players had not significantly contributed on the football field during the season. In order to test the hypothesis that players who were significant contributors on the football field would also have exhibited significantly less mean cortisol reactivity during summer testing than players who did not contribute significantly, a between-groups ANOVA analysis was conducted to compare the mean cortisol reactivity of each group of players (i.e., contributors vs. non-contributors).

Contributing players displayed a mean cortisol reactivity of 24% (std = .58) (i.e., the percent increase in salivary cortisol from the beginning to the end of the testing session) whereas non-contributing players had a mean cortisol reactivity of 66% (std = .91). This difference approached statistical significance, $F(1,45) = 3.099$, $p = .085$, $Mse = .638$.

**Figure 6.1 Toughness N = 47**




*Model Comparisons*

A nested model comparison using the $R^2$-change $F$-test was applied to determine whether significantly more variance could be accounted for by the full model than by the reduced model. The multiple regression model that included only traditional physical performance prediction metrics (i.e., lean body mass, hang clean, back squat, bench press, 10-yard dash, 20-yard shuttle, and vertical jump) produced $R^2 = .28$, $F(7, 39) = 2.209$, $p = .054$. Only the vertical jump test made a statistically significant contribution to the reduced model. The vertical jump test had a positive regression weight in this model suggesting that players who jumped higher tended to contribute more on the field, after controlling for all other variables in the model.

The multiple regression model that included cortisol reactivity, in addition to traditional physical performance prediction metrics, produced $R^2 = .39$, $F(8, 38) = 2.970$, $p = .011$. As hypothesized, this model performed significantly better than the reduced model, $R^2$-change = .385, $F(8, 38) = 2.970$, $p = .017$. Cortisol reactivity and the vertical jump test both made a statistically significant contribution to the full model. Cortisol reactivity had a negative regression weight in this model meaning that players who exhibited less cortisol reactivity tended to contribute more on the field, after controlling for all other variables in the model. The vertical jump test had a positive regression weight in this model suggesting that players who jumped higher also tended to contribute more on the field, after controlling for all other variables in the model.

*See Table 6.1*

## Discussion

When it is adaptive within their respective environment, a majority of individuals within a population express a particular phenotypic trait. Individuals who possess the most adaptive phenotypic traits are most likely to survive and to reproduce. Consequently, similar phenotypic characteristics will continue to occur within that population over time.

For example, aggressive behavior can be very adaptive for populations that fall victim to high predation rates. If such is the case, then aggressiveness will be naturally selected over time as the trait facilities survival (Boyce & Ellis, 2005; Ellis, Jackson, & Boyce, 2006). From an evolutionary perspective, when ancestral human beings encountered a threat, a *fight-or-flight* response was essential for survival. The evolutionary purpose of the physiological stress response system is to divert energy away from systems not immediately necessary for survival (e.g., the immune and digestive systems) and towards muscles and the brain (Sapolsky, 2004).

Individuals from a given population express similar phenotypic traits in response to shared environments, but there are individual differences as well. For example, there is variation among human beings in the expression of depressive symptoms related to stress. Genetic variation in the promoter region of the serotonin transporter (5-HTT) gene can lead to such differences. Individuals who possess the shorter (s) allele, rather than the homozygous longer (l) allele, tend to be more sensitive to environmental stressors and are more likely to develop long-term anxiety and depressive symptoms (Ellis et al., 2006).

Boyce and Ellis (2005) proposed the biological sensitivity to context (BSC) theory. BSC theory seems to offer the most comprehensive theoretical basis for interpreting variations in phenotypic outcomes among individuals from within the same population. According to this theory, variability in the way environmental and genetic factors interact results in the development of phenotypic outcomes. It is the gene/environment interactions that lead to an assortment of phenotypic outcomes expressed among individuals from a given population. An individual's stress response phenotype is ultimately the product of environmental influences on the individual's genetic make-up.

The concept of *allostasis* emphasizes phenotypic plasticity in response to environmental change; some individuals acclimate more easily than others to environmental changes (McEwen & Wingfield, 2003; Wingfield, 2005). Allostasis is an individual's ability to adjust and respond to environmental stressors while maintaining long-term physical and neural health (Wingfield, 2005). Conceptually similar to Boyce and Ellis' BSC theory, allostasis emphasizes the importance of individual differences in phenotypic outcomes (e.g., stress response patterns) in relation to environmental factors (Wingfield, 2005). Some individuals have stress response phenotypes that are more plastic, which allow them to acclimate to environmental stressors while others have difficulty acclimating to change. The latter individuals are more likely to develop dysfunction in the stress response systems.

Mean differences between groups of contributing and non-contributing players in cortisol reactivity approached statistical significance but did not quite meet the threshold. This was likely a result of limited power associated with including only 47 subjects in this study. Statistical power for this analysis was only 0.4. Had 120 athletes been included in this study, a statistically significant difference between groups would have been discovered given the same effect size. Of special note, the size of the relationship between cortisol reactivity and on-field contribution in this study was $eta^2 = .25$. This result suggests that cortisol reactivity may be a better predictor of on-field performance than many individual traditional physical performance prediction metrics (*Chapter 2 & Chapter 5*).

When cortisol reactivity was added to a model used to predict performance on the football field, which included traditional physical performance prediction metrics, the full model accounted for approximately 10% more variance (roughly a 35% increase) than the reduced model! Furthermore, cortisol reactivity was one of only two variables that significantly contributed to the model (along with vertical jump).

None of the traditional physical performance prediction metrics behaved substantially differently in the model when cortisol reactivity was added, which suggests that cortisol reactivity was uncorrelated with any traditional physical performance metrics. The correlation matrix confirmed as much (*Table 6.2*). While a great deal of multicollinearity was discovered among traditional physical performance prediction metrics (e.g., vertical jump was correlated with almost every other traditional physical performance prediction variable), cortisol reactivity seemed to capture a distinctive attribute.

The theory of Toughness emphasizes a reciprocal flow and correspondence of psychological systems with physiological systems, and a distinction between appraising circumstances as challenging (i.e., potentially taxing but likely to lead to positive outcomes) or threatening (i.e., unpredictable, uncontrollable, and likely to lead to adverse consequences). A mutually causal relationship has been observed between toughness and appraisals, and two physiological systems are particularly relevant: the autonomic nervous system and the HPA-axis.

Since increased HPA-axis responding is associated with circumstances appraised as threatening, one interpretation of the results is that players who are great on the football field are confident in their ability to generate strong test scores and view a testing session as challenging rather than threatening. At the very least, good football players may be confident enough in their on-field ability to not feel threatened by the prospects of posting some bad test scores should that happen. Either perspective would, in all likelihood, lead to exhibiting a "toughened" hormonal pattern in response to a testing session. This perspective equates toughness to genuine confidence in one's ability; it holds that reduced HPA-axis arousal in response to a testing session is primarily a result of confidence derived from previous success on the football field.

An alternative interpretation might hold that superior performance on the football field is a direct result of the propensity to respond to stressors with toughness (i.e., reduced cortisol reactivity). Players who exhibit a toughened hormonal response pattern reap neural benefits and are more likely to perform successfully on the football field. In all likelihood, a mutually causal relationship exists between performance and Toughness, such that successful performances are likely to lead to future appraisals of challenge (rather than threat), and therefore toughened patterns of hormonal responses, and conversely, toughness is likely to contribute directly to successful performance by means of improved neural functioning.

Regardless of the interpretation, toughness, as reflected by HPA-axis responding/cortisol reactivity, appears to offer insight into a unique aspect of performance on the football field that cannot be captured by any traditional physical performance prediction metrics. Players who were more likely to contribute on the football field also tended to demonstrate physiological toughness in response to a testing session. If cortisol reactivity, as a proxy measure for toughness, can be used to predict performance on the football field, then the metric should be included in performance prediction models, which have traditionally included only tests of physical ability.

## Table 6.1 Model Comparisons

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .620 | .385 | .255 | .424 | .385 | 2.971 | 8 | 38 | .011 |
| 2 | .535 | .287 | .159 | .451 | -.098 | 6.068 | 1 | 38 | .018 |

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 4.273 | 8 | .534 | 2.970 | .011 |
| | Residual | 6.833 | 38 | .180 | | |
| | Total | 11.106 | 46 | | | |
| 2 | Regression | 3.153 | 7 | .450 | 2.209 | .054 |
| | Residual | 7.953 | 39 | .204 | | |
| | Total | 11.106 | 46 | | | |

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | -.830 | .656 | | -1.266 | .213 |
| | Cortisol | -.203 | .082 | -.337 | -2.463 | .018 |
| | VJ | .003 | .001 | .455 | 2.286 | .028 |
| | 10 | .001 | .001 | .221 | 1.433 | .160 |
| | Shuttle | .009 | .035 | .042 | .248 | .805 |
| | Clean | -.002 | .002 | -.216 | -1.041 | .305 |
| | Squat | -.001 | .002 | -.090 | -.388 | .700 |
| | Bench | .001 | .002 | .064 | .328 | .745 |
| | Lean | -.006 | .015 | -.057 | -.378 | .708 |
| 2 | (Constant) | -.861 | .697 | | -1.236 | .224 |
| | VJ | .003 | .001 | .452 | 2.135 | .039 |
| | 10 | .001 | .001 | .175 | 1.078 | .288 |
| | Shuttle | -.007 | .036 | -.032 | -.180 | .858 |
| | Clean | -.001 | .002 | -.153 | -.700 | .488 |
| | Squat | -.001 | .003 | -.134 | -.549 | .586 |
| | Bench | .001 | .002 | .178 | .882 | .383 |
| | Lean | -.008 | .016 | -.082 | -.516 | .609 |

## Table 6.2 Correlation Matrix

| | Cont vs. Non-Cont | Cortisol Reactivity | Vertical Jump | 10-Yard Dash | 20-Yard Shuttle | Hang Clean | Back Squat | Bench Press | Lean Body Mass | Total Index |
|---|---|---|---|---|---|---|---|---|---|---|
| Cont vs. Non-Cont | 1 | -0.254 | 0.445* | 0.389 | 0.242 | 0.019 | 0.105 | 0.258 | 0.046 | 0.226 |
| Cortisol Reactivity | -0.254 | 1 | 0.04 | 0.156 | 0.16 | -0.136 | -0.092 | -0.192 | 0.02 | 0.022 |
| Vertical Jump | 0.445* | 0.04 | 1 | 0.504* | 0.624* | 0.467* | 0.546* | 0.567* | 0.378 | 0.718* |
| 10-Yard Dash | 0.389 | 0.156 | 0.504* | 1 | 0.396 | 0.08 | 0.185 | 0.281 | 0.17 | 0.52* |
| 20-Yard Shuttle | 0.242 | 0.16 | 0.624* | 0.396 | 1 | 0.398 | 0.4* | 0.376 | 0.357 | 0.677* |
| Hang Clean | 0.019 | -0.136 | 0.467* | 0.08 | 0.398 | 1 | 0.758* | 0.564* | 0.469* | 0.722* |
| Back Squat | 0.105 | -0.092 | 0.546* | 0.185 | 0.4* | 0.758* | 1 | 0.703* | 0.441* | 0.695* |
| Bench Press | 0.258 | -0.192 | 0.567* | 0.281 | 0.376 | 0.564* | 0.703* | 1 | 0.392 | 0.664* |
| Lean Body Mass | 0.046 | 0.02 | 0.378 | 0.17 | 0.357 | 0.469* | 0.441* | 0.392 | 1 | 0.815* |
| Total Index | 0.226 | 0.022 | 0.718* | 0.52* | 0.677* | 0.722* | 0.695* | 0.664* | 0.815* | 1 |

** indicates p-value <.05 after Bonferroni correction*

# Chapter 7

# CONCLUSION:

# TOUGHNESS PREDICTS PERFORMANCE IN COLLEGE FOOTBALL

A significant portion of almost all collegiate athletic department budgets depend on football (Zimbalist, 2001). In 2011, NCAA Division 1 public schools reported total revenues of over $7 billion (Schnaars et al., 2012). Brown (2010) updated previous estimates to report that a single premium college football player's marginal revenue product is over $1 million. NCAA athletics is big business and college football drives soaring revenues.

A football team's ability to generate revenue is a result of the team's total skill level, which is assumed to be a function of the individual skill levels of players that make up the team (Brown & Jewell, 2004). When a coach recruits a player to compete for his football program who fails to sufficiently contribute during the course of his career, a type 1 error has occurred. By contrast, when a coach does not admit an athlete who would have contributed in a positive way to the overall team product, the coach has committed a type 2 error (Spieler et al., 2007). Both kinds of errors can be extremely detrimental to a program. Considering the value of acquiring a premium player, an optimal prospective player assessment and selection process is paramount.

Success on the football field is conventionally attributed to physical characteristics of players (Hyllegard, Radlo, & Early, 2001). Division I football players are typically assessed via a battery of physical tests designed to address multiple aspects of performance; categories of tests often include anthropometric, strength, and functional measures. Although test scores are not direct measures of playing ability, one-repetition maximum strength tests (e.g., hang clean or back squat) and functional assessments (e.g., 40-yard dash or vertical jump) are believed to reflect the physical performance characteristics that represent a player's potential to perform on the field (Fry and Kraemer, 1991). Therefore, predictions about players' likelihood for success on the football field are frequently derived from physical performance test scores (McGee & Burkett, 2003).

The overall success a team can achieve is greatly affected by the process by which players are selected (Humara, 2000). However, despite considerable speculation, very little definitive empirical evidence exists to suggest specific combinations of physical attributes that best predict performance (Davis, Barnett, Kiger, Mirasola, & Young, 2004).

Chapter 2 evaluated traditional physical performance prediction metrics that have appeared most frequently in published literature for the extent to which they have accounted for variance in performance on the football field. A total of 16 publications were included in meta-analyses. Lean body mass was the best predictor of on-field performance with an average weighted effect size of 0.34. Back squat was also a strong predictor of performance with an average weighted effect size of 0.29. Vertical jump was the best "functional" predictor of performance on the football field with an average weighted effect size of 0.28. The 20-yard shuttle was least predictive of performance on the football field with an average weighted effect size of only 0.10.

While all of the traditional physical performance prediction metrics most frequently cited in published literature were significant predictors of performance on the football field when considered as part of meta-analyses, results suggest that some of the tests on which coaches tend to rely most, such as the 40-yard dash and bench press, typically account for only about 5% of the variance in on-field performance. Even the best predictor, lean body mass, only accounts for about 11% of the variance in performance. Considering the implications of appropriately assessing and selecting prospective players, the results from the meta-analyses are surprising. Frequently cited tests, such as the 20-yard shuttle, were discovered to be nearly useless when used to predict performance on the football field.

The National Football League (NFL) is the highest level of football in the United States and precise evaluation of prospective players has significant financial implications for all teams (Sierer et al., 2008; Robbins, 2010). Tests of physical ability are among the most important considerations when drafting players (Robbins, 2010). Therefore, significant resources are employed to assess player potential at the annual NFL Combine where players complete multiple physical tests: vertical jump, broad jump, 10-, 20-, and 40-yard dash, 20-yard shuttle run, 3-cone drill, bench press, and 60-yard shuttle run (Ghigiarelli, 2011; McGee & Burkett, 2003). Chapter 3 conveyed insights from NFL scouts, who are considered exceptionally skilled applied practitioners at selecting talent in the game of football.

NFL scouts were asked, “On which tests do you rely most when determining which players to draft?” Responses were tallied in order to identify the tests that are relied upon most by NFL scouts to predict player performance. In order to determine which tests the most successful teams tend to rely upon, the number of wins that each franchise accumulated during the subsequent 3 seasons was correlated with whether or not the team's scouting department had designated a specific performance test to be useful.

NFL franchises differed considerably with respect to emphases placed upon different metrics. The test that was most frequently cited by NFL scouts as a test relied upon when determining which players to draft was the vertical jump (10 teams). Also frequently cited were the 40-yard dash (9 teams) and 20-yard shuttle (8 teams). Least frequently cited was the 20-yard dash (only 1 team) and scouts tended to deemphasize measures of strength. It appeared that teams may attempt to identify a single test for each physical attribute rather than considering multiple tests of any particular characteristic.

Correlation analyses suggested that there was no relationship between listing any of the tests and winning. It is surprising that consideration of *none* of the tests was predictive of success in the NFL. The only statistically significant relationship observed within the dataset was a negative relationship between the number of tests a team reported considering and the number of wins the team accumulated over the subsequent 3 years. However, this relationship was not in the expected direction; teams that listed fewer tests tended to win more games.

Dr. Tom Osborne was extraordinarily successful as a Division 1 head football coach despite some recruiting disadvantages that were inherent to his program. Chapter 4 presented a structured interview during which Osborne described his integrative approach to recruiting. Osborne had a unique perspective to offer as he is considered an authority in both the world of football and psychology.

Osborne didn't seem to agree with NFL scouts in regard to the importance of the vertical jump test, but did mention agility and speed as good indicators of physical ability at a number of positions. He also mentioned height and weight as factors worthy of consideration when evaluating linemen. When interpreting any physical test scores, he cautioned against self-report data and highlighted the efforts of Boyd Epley and his strength staff for their work to develop the Performance Index as an objective scale that facilitates evaluation of physical test scores among football players who play different positions. Osborne considers the Performance Index valuable because it encompasses all of the primary physical performance metrics he considered when evaluating players. He suggested that the Performance Index may be a preferable method for assessing physical ability among players.

Chapter 5 introduced the Performance Index, which was developed as part of the Husker Power Strength and Conditioning Program at the University of Nebraska, as an optimal conversion of traditional physical performance prediction metrics. The Performance Index is a decathlon-type scoring system that rewards points on an increasing basis as athletes approach a world-class performance. Importantly, the calculation controls for body weight, and therefore allows for objective comparisons of players from different positions.

Physical performance test scores were recorded among a group of 47 Division 1 freshman football players and then converted into Performance Index scores and compared between groups of players who were judged by their coaching staff to have significantly contributed on the field during their first season of play or were judged to have not significantly contributed during their first season. The on-field criterion variable was deliberately left open-ended and subjective as suggested by members of the coaching staff. The coaching staff reasoned that the most important measure of a football player is whether he significantly contributes on the field, and that contribution can occur in many forms.

Variables included in these analyses were lean body mass, hang clean, back squat, bench press, 10-yard dash, 20-yard shuttle, vertical jump, and a total Performance Index score, which was simply the sum of all test scores. The 10-yard dash and vertical jump were the only two physical test scores that differed between contributors and non-contributors in this sample of freshman football players. Mean differences between groups of contributing and non-contributing players in Performance Index points for the 20-yard shuttle, bench press, and total Performance Index score approached statistical significance but did not quite meet the threshold, likely due to insufficient statistical power. Hang clean and back squat index scores were not significantly different between groups of players who contributed and those who did not. Lean body mass is considered by the developers of the Performance Index to be among the best predictors of physical ability. However, in this study, lean body mass scores were not predictive of performance on the football field. This result may be attributable to range restriction.

General findings suggested that functional tests are superior to strength tests when distinguishing freshman football players who significantly contribute on the field during their first season of play from players who do not. In addition, less technically sophisticated neuromuscular movements discriminated better than those that are more complex. Both functional and strength tests probably work better than anthropometric tests, although only one anthropometric test was included in this study.

A primary objective of this dissertation was to explore: 1) the traditional physical performance prediction metrics that are most predictive of performance on the football field, the way in which one can best incorporate such metrics during evaluation processes, and the extent to which such metrics can be relied upon when making decisions. Meta-analyses that included published literature suggested advantages to measuring lean body mass, back squat, and vertical jump, and a lack of utility associated with the 20-yard shuttle. NFL scouts tended to rely on the vertical jump, 40-yard dash, and 20-yard shuttle, while the 20-yard dash and strength measures were judged to be less useful.

Osborne hinted that speed (i.e., the 40-yard dash), agility (i.e., the 20-yard shuttle), and even height and weight at certain positions are relevant, and suggested that, when interpreting traditional physical performance prediction metrics, one ought to convert scores into Performance Index scores. When traditional physical performance prediction metrics among freshman football players were converted into Performance Index scores, the 10-yard dash and vertical jump were significant predictors of on-field performance, and the 20-yard shuttle, bench press, and total Performance Index score approached statistical significance. Hang clean, back squat, and lean body mass measurements were not predictive of on-field performance among the freshman players who were studied.

The various sources of information were contradictory and conveyed an incomplete representation. Average weighted effect sizes were small and variance accounted for in on-field performance was less than expected (Chapters 2 & 5). In addition, reliance on none of the tests of physical ability for selection purposes were predictive of team success in the NFL; in fact, the more tests a team considered when selecting players, the fewer games the team won (Chapter 3). According to Dr. Tom Osborne, traditional physical performance prediction metrics tell only part of the story and fail to completely capture the essence of what it takes to be successful in the game of football (Chapter 4).

As early as 1984, Shields discussed the existence of an immeasurable feature, different from size, strength, and endurance, which helps explain performance in the game of football. Humara (2002) said that, in addition to a players' past performance and bio (physical) data, practitioners should consider making greater use of psychological evaluations. In light of the lack of predictive validity associated with many of the NFL Combine tests, Kuzmits and Adams (2008) called for more rigorous psychological testing. Sports psychology asserts that psychological skills and mental attributes contribute to success in athletics (Laguna & Ravizza, 2003).

Osborne had vast experience evaluating players based on physical characteristics but suggested that psychological attributes were equally important to his selection process. He explained his unique perspective with emphasis on his psychology background. Osborne felt that the Nebraska Football Team that won the 1995 National Championship was the best team he ever coached due in large part to an element of toughness the team possessed and reiterated that there is an element of toughness that great players and teams must possess. He felt that mental toughness was a particularly important attribute for football players to possess due to the ups and downs associated with a football season, but acknowledged the fact that there may be no good objective measures for such psychological attributes. He indicated that coaches are often forced to rely on intuitive judgments of psychological characteristics among players and believed that the study of psychology is one area, in particular, that coaches could significantly improve upon in order to understand players better.

Osborne explained that each recruiting class consists of five or six players who do not make it, typically for psychological reasons. He reasoned that, if there were a better way to evaluate psychological characteristics of players and to reduce the number of players lost down from six to only two or three, it would make quite a difference over a four to five year period. Osborne's perspective prompted the second objective of this dissertation, to explore 2) the extent to which measuring the construct referred to as "toughness" can improve models used to predict performance on the football field.

Because of the apparent connection between the characteristic of mental toughness and successful performance in athletics, athletes, coaches, and sport psychologists seem to agree upon the importance of the attribute (Crust, 2007). However, despite considerable interest, there has been no consensus on the definition, conceptualization, and precise way in which to develop mental toughness (Crust, 2007). Some researchers have considered physical toughness an important component of mental toughness (Gucciardi et al., 2007). Dienstbier (1989) coined the term "Toughness" to describe a specific hormonal response pattern that promotes success in myriad tasks. Chapter 6 discussed a measure of toughness that could be used to evaluate prospective football players.

The theory of Toughness embodies a reciprocal flow between psychological and physiological systems and emphasizes a distinction between evaluating circumstances as being challenging (i.e., potentially taxing but likely to lead to positive outcomes) or threatening (i.e., unpredictable, uncontrollable, and likely to lead to adverse consequences). Interestingly, one's appraisal of a stressful situation seems to affect physiological processes (Dienstbier, 1989). A relationship has been observed between appraisals and the correspondence of two primary physiological stress response systems (i.e., the autonomic nervous system and the HPA-axis).

When one appraises a situation as *challenging*, arousal of the autonomic nervous system is exhibited. Conversely, when one appraises a situation as *threatening,* then joint autonomic and HPA-axis arousal is elicited (e.g., Frankenhaeuser, 1979). In relation to performance, Scandinavian researchers have consistently found that increased catecholamine release predicts successful performance, even in complex tasks. By contrast, high levels of cortisol appear to be predictive of poorer performance (e.g., Ursin, Baade, & Levine, 1978; Vaernes, Ursin, Darragh & Lambe, 1982).

Results from both human and non-human studies suggest that increased cortisol reactivity is counterproductive (Dienstbier, 1989). A *toughening effect* can modify one's neuroendocrine system and enhance the capacity to generate norepinephrine and epinephrine. A toughened individual's increased capacity to secrete catecholamines leads to delay and/or suppression of HPA-axis responses and an optimal pattern of cortisol arousal entails an ability to suppress cortisol reactivity (Dienstbier, 1989).

Mental toughness could be characterized by an ability to cope with stressors more effectively as a result of reduced cortisol reactivity (Crust, 2007). As a long established indicator of HPA-axis responding, salivary cortisol measures are easily obtained and analyzed (Kirschbaum & Hellhammer, 1989), and reliably estimate blood cortisol levels (Aardal Eriksson, Karlberg, & Holm, 1998). Salivary cortisol reactivity in response to a testing session was measured among freshman football players as an indication of HPA-axis activation, and by extension, “Toughness”.

Results suggested that mean cortisol reactivity in response to testing is less for players who significantly contributed on the football field than for players who did not, although the results were not statistically significant. Results from this study did not meet the threshold for statistical significance likely as a result of including too few research subjects. Of note, the size of the relationship between cortisol reactivity and on-field performance in this study was $eta^2 = .25$; this value is comparable to many of the best traditional physical performance prediction metrics in terms of predictive validity (Chapters 2 & 5).

When cortisol reactivity was added to a model used to predict performance on the football field that initially included only traditional physical performance prediction metrics, the full model accounted for approximately 10% more variance (roughly a 35% increase) than the reduced model based solely on traditional physical performance prediction metrics. Furthermore, cortisol reactivity was one of only 2 variables that significantly contributed to the model (along with vertical jump). These results suggest not only that cortisol reactivity may be a better predictor of on-field performance than many individual traditional physical performance prediction metrics, but also that cortisol reactivity may capture a distinctive attribute.

Toughness, as reflected by HPA-axis responding/cortisol reactivity, appears to offer insight into a unique aspect of player performance that cannot be captured by any of the physical test scores that are traditionally used to predict performance on the football field. Players who were more likely to contribute on the football field also tended to demonstrate toughness. If cortisol reactivity, as a measure of toughness, can be effectively used to predict performance on the football field, then the metric should be included in performance prediction models that have traditionally included only physical attributes of players.

Coaches are often considered experts with respect to identifying physical attributes that contribute to athletic success, but may lack the expertise to evaluate psychological factors involved (Humara, 2000). Coaches have often relied on informal judgments of psychological factors to determine a player's potential to succeed (Humara, 2000). Crust (2007) recommends that researchers should search for physiological correlates of psychological characteristics like mental toughness. Coaches accustomed to evaluating physical indices might be more inclined to consider psychological attributes that have physiological underpinnings. Coaches ought to consider measuring psychological attributes such as toughness in order to better understand the essence of what it takes to successfully compete in the game of football.

References

Aardal Eriksson, E., Karlberg, B. E., & Holm, A. C. (1998). Salivary cortisol- an alternative to serum cortisol determinations in dynamic function tests. *Clinical Chemistry and Laboratory Medicine*, *36*, 215-222.

Armitage, P. (1984). Controversies and achievements in clinical trials. *Controlled Clinical Trials*, *5*, 67-72.

Arnold, J. A., Coker, T. P., & Micheli, R. P. (1977). Anatomical and physiological characteristics to predict football ability at the university of arkansas. *The Journal of the Arkansas Medical Society 74*, 253-260.

Arthur M. R. & Bailey, B. (1998). *Complete conditioning for football.* Champaign, IL: Human Kenetics.

Barker, M., Wyatt, T. J., Johnson, R. L., Stone, M. H., O'Bryant, H. S., Poe, C., & Kent, M. (1993). Performance factors, psychological assessment, physical characteristics, and football playing ability. *Journal of Strength and Conditioning Research*, *7*, 224-233.

Barnes, D. M. (1986). Steroids may influence changes in mood. *Science*,*232*, 1344–1345.

Beltzer, E. K. (2008). *Physiological responses to naturally occurring stressors: The roles of context and dehydroepiandrosterone.* Lincoln, NE: University of Nebraska-Lincoln.

Berg, K., Latin, R., & Baechle, T. (1990). Physical and performance characteristics of ncaa division I football players. *Research Quarterly for Exercise and Sport*, *61*, 395-401.

Black, W. & Roundy, E. (1994). Comparisons of size, strength, speed, and power in ncaa division 1-A football players. *Journal of Strength and Conditioning Research, 8,* 80-85.

Boyce, W. T. & Ellis, B. J. (2005). Biological sensitivity to context: An evolutionary-developmental theory of the origins and functions of stress reactivity. *Development and Psychopathology, 17*, 271-301.

Brown, R. W. (1993). An estimate of the rent generated by a premium college football player. *Economic Inquiry, XXXI*, 671-684.

Brown, R. W. & Jewell, R. T. (2004). Measuring marginal revenue product of college athletics: Updated estimates. In R. Fort & J. Fizel (Eds.), *Economics of college sports.* Westport, CT: Praeger Publishers.

Brown, R.W. (2010). Research Note: Estimate of College Football Rents, *Journal of Sports Economics*, August 2010.

Bull, S. J., Albinson, J. G., & Shambrook, C. J. (1996). *The mental game plan: Getting psyched for sport.* Eastbourne, Sports Dynamics.

Charmandari, E., Tsigos, C., & Chrousos, G. (2005). Endocrinology of the stress response. *Annual Review of Physiology, 67*, 259-284.

Clough, P. J., Earle, K., & Sewell, D. (2002). Mental toughness: The concept and its measurement. In I. Cockerill (Ed.), *Solutions in sport psychology* (pp. 32-43). London: Thompson Publishing.

Crust, L. (2007). A review and conceptual re-examination of mental toughness: Implications for future researchers. Retrieved June 25, 2011, from http://eprints.lincoln.ac.uk/2532/1/A_review_and_conceptual_re-examination_of_mental_toughness.pdf

Davis, D. S., Barnette, B. J., Kiger, J. T., Mirasola, J. J., & Young, S. M. (2004). Physical characteristics that predict functional performance in division I college football players. *Journal of Strength and Conditioning Research, 18,* 115-120.

DerSimonian, R. & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7,* 177-188.

Dickerson, S. S., Gruenewald, T. L., & Kemeny, M. E. (2004). When the social self is threatened: Shame, physiology and health. *Journal of Personality, 72*, 1191-1216.

Dienstbier, R. A. (1989). Arousal and physiological toughness: Implications for mental and physical health. *Psychological Review, 96,* 84-100.

Dos Remedios, R. & Holland, G., (1992). Physical and performance characteristics of community college football players. *National Strength and Conditioning Association Journal, 14*, 9-12.

Ellis, B. J., Jackson, J. J., & Boyce, W. T. (2006). The stress response systems: Universality and adaptive individual differences. *Developmental Review, 26,* 175-212.

Folkman , S. & Lazarus, R. S. (1985). If it changes it must be a process: Study of emotion and coping during three stages of a college examination. *Journal of Personality and Social Psychology, 48*, 150-170.

Fourie, S. & Potgieter, J. R. (2001). The nature of mental toughness in sport. *South African Journal for Research in Sport, Physical Education and Recreation, 23,* 63-72.

Frankenhaeuser, M. (1979). Psychoneuroendocrine approaches to the study of emotion as related to stress and coping. In H. E. Howe, Jr. & R. A. Dienstbier (Eds.), *Nebraska Symposium on Motivation 1978; Human emotion* (Vol. 27, pp. 123-161). Lincoln: University of Nebraska Press.

Fry, A. & Kraemer, W. (1991). Physical performance characteristics of American collegiate football players. *Journal of Applied Sport Science Research, 5,* 126-138.

Garstecki, M. A., Latin, R. W., & Cuppett, M. M. (2004). Comparison of selected physical fitness and performance variables between ncaa division I and II football players. *Journal of Strrength and Conditioning Research, 18,* 292-297.

Ghigiarelli, J. J. (2011). Combine performance descriptors and predictors of recruit ranking for the top high school football recruits from 2001 to 2009: Differences between position groups. *Journal of Strength and Conditioning Research, 25,* 1193-1203.

Goldberg, A. S. (1998). *Sports slump busting: 10 steps to mental toughness and peak performance.* Champaign, IL: Human Kenetics.

Golby, J., Sheard, M., & van Wersch, A. (2007). Evaluating the factor structure of the psychological performance inventory. *Perceptual and Motor Skills, 105,* 309-325.

Granger, D. A., Kivlighan, K. T., El-Sheikh, M., Gordis, E. B., & Stroud, L. R. (2007). Salivary a-amylase in biobehavioral research: Recent developments and applications. *Annals of the New York Academy of Sciences, 1098,* 122-144.

Habib, K. E., Gold, P. W., & Chrousos, G. P. (2001). Neuroendocrinology of stress. *Endocrinology Metabolism Clinics of North America, 30,* 695-728.

Henry, J. P. & Ely, D. L. (1975). Physiology of emotional stress: Specific responses. *Journal of the South Carolina Medical Association, 75,* 501-509.

Henry, J. P. (1997). Psychological and physiological responses to stress: The right hemisphere and the hypothalamo-pituitary-adrenal axis, an inquiry into problems of human bonding. *Acta Physiologica Scandinavica, 640,* 10-25.

Humara, M. (2000). Personnel selection in athletic programs. *Athletic Insight, 2,* 1-7.

Hyllegard R., Radlo, S. J., & Early, D. (2001). Attribution of athletic expertise by college coaches. *Perceptual Motor Skills, 92,* 193-207.

Iguchi, J., Yamada, Y., Ando, S., Fujisawa, Y., Hojo, T., Nishimura, K., Kuzuhara, K., Yuasa, Y., & Ichihashi, N. (2011). Physical and performance characteristics of japanese division 1 collegiate football players. *Journal of Strength and Conditioning Research, 25,* 3368- 3377.

Johnson, J.B. (2001). Evaluating the importance of strength, power, and performance tests in an ncaa division I football program. Retrieved June 25, 2011, from http://scholar.lib.vt.edu/theses/available/etd-12052001-125914/unrestricted/johnsondiss.pdf

Jones, G., Hanton, S., & Connaughton, D. (2002). What is this thing called mental toughness? An investigation of elite sport performers. *Journal of Applied Sport Psychology, 14*, 205- 218.

Jones, G., Hanton, S., & Connaughton, D. (2007). A framework of mental toughness in the world's best performers. *The Sport Psychologist, 21,* 243-264.

Kaiser, G. E., Womack, J. W., Green, J. S., Pollard, B., Miller, G. S., & Crouse, S. F. (2008). Morphological profiles for first-year national collegiate athletic association division I football players. *Journal of Strength and Conditioning Research, 22,* 243-249.

Kirschbaum, C. & Hellhammer, D. H. (1989). Salivary cortisol in psychobiological research: An overview. *Neurpsychobiology, 22*, 150-169.

Kuzmits, E. F. & Adams, A. J. (2008). The NFL combine: Does it predict performance in the national football league? *Journal of Strength and Conditioning Research, 22,* 1721-1727.

Laguna, P. L. & Ravizza, K. (2003). Collegiate athletes' mental skill use and perceptions of success: An exploration of the practice and competition settings. *Journal of Applied Sports Psychology, 15,* 115-128.

Langelett, G. (2003). The relationship between recruiting and team performance in division 1a college football. *Journal of Sports Economics, 2,* 240 – 245.

Loehr, J. E. (1995). *Mental toughness training for sport: Achieving athletic excellence.* Lexington, MA: Stephen Greene.

McEwen, B. S. & Wingfield, J. C. (2003). The concept of allostasis in biology and biomedicine. *Hormones and Behavior, 43*, 2-15.

McGee, K. J. & Burkett, L. N. (2003). The national football league combine: A reliable predictor of draft status? *Journal of Strength and Conditioning Research, 17,* 6-11.

Meaney, M. J., Aitkens, D. H., Berkel, C., Bhatnagar, S., Sarrieau, A., & Sapolsky, R. M. (1987). *Post-natal handling attenuates age-related changes in the adrenocortical stress response and spatial memory deficits in the rat.* Paper presented at the 17th Annual Meeting of the Society of Neuroscience, New Orleans.

Munck, A., Guyre, P. M., & Holbrook, N. J. (1984). Physiological functions of glucocorticoids in stress and their relations to pharmacological actions. *Endocrine Reviews, 5*, 25-44.

National Collegiate Athletic Association, (2010). *2010 – 11 NCAA Division I Manual: Constitution Operating Bylaws, Administrative Bylaws*. Indianapolis: NCAA.

O'Connor, D. B., Jones, F., Conner, M., McMillan, B., & Ferguson, E. (2008). Effects of daily hassles and eating style on eating behavior. *Health Psychology, 27,* S20-S31.

Osborne, T. (1999). *Faith in the game.* New York: Random House Inc.

Osborne, T. (2009). *Beyond the final score.* Ventura, CA: Regal.

Pearson D. R. & Gehlson G. M. (1998). Athletic performance enhancement: A study with college football players. *Strength and Conditioning Journal,* 70 – 73.

Pfeifer, W. D. (1976). Modification of adrenal tyrosine hydroxylase activity in rats following manipulation in infancy. In E. Usdin, R. Kvetnansky,& I. J. Kopin (Eds.), *Catecholamines and stress* (pp. 265–270). Oxford, England: Pergamon.

Plothow, L. (2010). The nfl combine and draft: Correlation and implications for nfl success. Retrieved June 25, 2011, from http://news.byu.edu/releases/archive10/Apr/Paper.pdf

Robbins, D. W. (2010). The NFL combine: Does normalized data better predict performance in the NFL draft? *Journal of Strength and Conditioning Research, 24*, 2888-2899.

Rose, R. M., Poe, R. O., & Mason, J. W. (1967). Observations on the relationship between psychological state, 17-OHCS excretion, and epinephrine, norepinephrine, insulin, BEI, estrogen and androgen levels during basic training. *Psychosomatic Medicine*, *29*, 544.

Sapolsky, R. M. (2004). *Why Zebras Don't Get Ulcers* (3rd Ed.). Owl Books: New York.

Sawyer, D.T., Ostarello, J. Z., Suess, E.A., & Dempsey, M, (2002). Relationship between football playing ability and selected performance measures. *Journal of Strength and Conditioning Research, 16,* 611-616.

Schnaars, C., Upton, J., Mosemak, J., & DeRamus, K. (2012). NCAA college athletics department finances database. Retrieved August 7, 2012, from http://www.usatoday.com/sports/college/story/2012-05-14/ncaa-college-athletics-finances-database/54955804/1

Shields, C. L., Whitney, F. E., & Zomar, V. D. (1984). Exercise performance of professional football players. *The American Journal of Sports Medicine*, *12,* 455-459.

Sierer, S. P., Battaglini, C. L., Mihalik, J. P., Shields, E. W., & Tomasini, N. T. (2008). The national football league combine: Performance differences between drafted and nondrafted players entering the 2004 and 2005 drafts. *Journal of Strength and Conditioning Research, 22,* 6-12.

Spieler, M., Czech, D.R., Joyner, A. B., & Munkasy, B. (2007). Predicting athletic success: Factors contributing to the success of ncaa division I aa collegiate football players. *Athletic Insight*, *9,* 22-33.

Stuempfle, K. J., Katch, K. I., & Petrie, D. F., (2003). Body composition relates poorly to performance tests in ncaa division III football players. *Journal of Strength and Conditioning Research, 17,* 238-244.

Spieler, M., Czech, D. R., Joyner, A. B., Munkasy, B., Gentner, N., & Long, J. (2007). Predicting athletic success: Factors contributing to the success of NCAA Division I aa collegiate football players. *Athletic Insight, 9*, 22-33.

Stuempfle, K. J., Katch, F. I., & Petrie, D. F. (2003). Body composition relates poorly to performance tests in ncaa division III football players. *Journal of Strength and Conditioning Research, 17,* 238-244.

Thelwell, R., Weston, N., & Greenlees, I. (2005). Defining and understanding mental toughness within soccer. *Journal of Applied Sport Psychology, 17*, 326-332.

Tsigos, C. & Chrousos, G. P. (2002). Hypothalamic-pituitary-adrenal axis, neuroendocrine factors and stress. *Journal of Psychosomatic Research*, *53*, 865-871.

Ursin, H., Baade, E., & Levine, S. (Eds.). (1978). *Psychobiology of stress: A study of coping men.* New York: Academic Press.

Vaernes, R., Ursin, H., Darragh, A., & Lambe, R. (1982). Endocrine response patterns and psychological correlates. *Journal of Psychosomatic Research, 26,* 123–131.

Wieberg, S., Upton, J., & Berkowitz, S. (2012). Texas athletics overwhelm rivals in revenue and spending. Retrieved August 7, 2012, from http://usatoday30.usatoday.com/sports/college/story/2012-05-15/texas-athletics-spending-revenue/54960210/1

Wingfield, J. C. (2005). The concept of allostasis: Coping with a capricious environment. *Journal of Mammalogy*, *86*, 248-254.

Zatsiorsky, V. M. & Kraemer, W. J. (2006). *Science and practice of strength training* (2nd ed.). Champaign, IL: Human Kinetics.

Zimbalist, A. S. (2001). *Unpaid Professionals.* Princeton: Princeton University Press.