Science of Strength and Conditioning Series

NSCA's Guide to TESTS AND ASSESSMENTS

NSCA™
National Strength and Conditioning Association

Todd Miller

NSCA's Guide to Tests and Assessments

National Strength and Conditioning Association

运动力量训练　　NSCA's Guide to Tests and Assessments

Todd Miller, PhD, CSCS*D
George Washington University
EDITOR

Human Kinetics

Library of Congress Cataloging-in-Publication Data

National Strength & Conditioning Association (U.S.)
NSCA's guide to tests and assessments / National Strength and Conditioning Association ; Todd Miller, editor.
p. ; cm. -- (Science of strength and conditioning series)
Includes bibliographical references and index.
ISBN-13: 978-0-7360-8368-3 (hard cover)
ISBN-10: 0-7360-8368-5 (hard cover)
I. Miller, Todd, 1967- II. Title. III. Series: Science of strength and conditioning series.
[DNLM: 1. Athletic Performance--physiology--Guideline. 2. Physical Fitness--physiology--Guideline. 3. Exercise--physiology--Guideline. 4. Exercise Test--methods--Guideline. 5. Physical Examination--Guideline. 6. Sports--physiology--Guideline. QT 260]

613.7--dc23

2011038488

ISBN-10: 0-7360-8368-5
ISBN-13: 978-0-7360-8368-3

The web addresses cited in this text were current as of August 2011, unless otherwise noted.

Developmental Editor: Kevin Matz; **Assistant Editors:** Steven Calderwood and Bethany J. Bentley; **Copyeditor:** Patsy Fortney; **Indexer:** Betty Frizzell; **Permissions Manager:** Dalene Reeder; **Graphic Designer:** Nancy Rasmus; **Graphic Artist:** Joe Buck; **Cover Designer:** Keith Blomberg; **Photographs (interior):** Neil Bernstein, all photos © Human Kinetics unless otherwise noted; **Photo Asset Manager:** Laura Fitch; **Visual Production Assistant:** Joyce Brumfield; **Photo Production Manager:** Jason Allen; **Art Manager:** Kelly Hendren; **Associate Art Manager:** Alan L. Wilborn; **Art Style Development:** Jennifer Gibas; **Illustrations:** © Human Kinetics; **Printer:** Sheridan Books

We thank the National Strength and Conditioning Association in Colorado Springs, Colorado, for assistance in providing the location for the photo shoot for this book.

Printed in the United States of America 10 9 8 7 6 5 4 3 2

The paper in this book is certified under a sustainable forestry program.

Human Kinetics
Website: www.HumanKinetics.com

United States: Human Kinetics
P.O. Box 5076
Champaign, IL 61825-5076
800-747-4457
e-mail: humank@hkusa.com

Canada: Human Kinetics
475 Devonshire Road Unit 100
Windsor, ON N8Y 2L5
800-465-7301 (in Canada only)
e-mail: info@hkcanada.com

Europe: Human Kinetics
107 Bradford Road
Stanningley
Leeds LS28 6AT, United Kingdom
+44 (0) 113 255 5665
e-mail: hk@hkeurope.com

Australia: Human Kinetics
57A Price Avenue
Lower Mitcham, South Australia 5062
08 8372 0999
e-mail: info@hkaustralia.com

New Zealand: Human Kinetics
P.O. Box 80
Torrens Park, South Australia 5062
0800 222 062
e-mail: info@hknewzealand.com

E4846

Science of Strength and Conditioning Series

NSCA's Guide to Sport and Exercise Nutrition

NSCA's Guide to Tests and Assessments

NSCA's Guide to Program Design

National Strength and Conditioning Association

Human Kinetics

Preface

If you can't measure it, you can't control it. One of my mentors repeated this "quality axiom" to me on a daily basis during my years as a graduate student, and this fundamental message has become ingrained in my approach to training. As strength and conditioning professionals, our primary goal is to design and implement programs that result in optimal athletic performance. At first glance, this appears to be a simple task. By following the principles of specificity, overload, and progression, we can design conditioning and resistance training programs that increase fitness and athletic performance.

Unfortunately, while our programs may bring about improved performance for athletes and clients, it is impossible to know whether these adaptations are optimal without incorporating some well-conceived testing and measurement schemes into a regimen. Indeed, it is common for a trainer to claim that his or her program works, but the design of strength and conditioning programs is not simply about improving performance. It is about safely improving performance *to the greatest degree possible* for a specific individual with a specific set of goals. Achieving this optimal level of improvement is simply not possible without a strategy for tracking changes in performance over time.

Historically, testing and measurement for the exercise sciences have been heavily slanted toward a clinical population and have been focused mainly through the lens of disease and disease prevention. Much less attention has been given to testing for athletic performance, and this is reflected in the paucity of literature on the topic. Tests for power, speed, agility, and mobility (all topics addressed in this text) lean heavily toward athletic performance and are rarely used in clinical settings. This book serves as a resource for coaches, trainers, students, and athletes of all skill levels and addresses the importance of testing and measurement for athletic performance.

The text begins by laying the foundation of testing and data analysis and the methods of interpreting results and drawing conclusions. The chapters that follow include tests from the rudimentary (such as body composition and blood pressure measurement) to the more complex, such as lactate threshold testing and aerobic power. While all of these tests vary in complexity, this variability is not indicative of their degree of importance. For example, measuring body composition is a relatively simple task, yet its implications in athletic performance are incredibly profound. It is clear that excess fat can be deleterious to performance in sports that rely on speed, acceleration, and rapid changes in direction. Despite this, coaches will often spend long hours on speed training but pay little attention to measuring or improving body composition. We hope that this text not only serves as an instructional tool for the mechanics of conducting specific tests but that it also helps coaches determine which tests are appropriate for specific populations. For example, a test of aerobic power may be inappropriate for a thrower, whose performance relies primarily on strength and power. Conversely, a coach of a distance runner would benefit little from conducting agility testing on athletes. Therefore, you should not assume that you need to read this text cover to cover, nor should you assume all tests are appropriate for all athletes.

As the field of strength and conditioning becomes increasingly sophisticated, so should the approach by which training programs are designed, implemented, and tested. A training program that lacks some type of progress tracking is grossly incomplete, yet it remains startlingly common among trainers of today. We are confident that this text will provide a solid foundation by which you can develop and implement your own testing and measurement programs, ultimately allowing you to grow as a coach and maximize the performance of your athletes.

Contents

1

Tests, Data Analysis, and Conclusions

Matthew R. Rhea, PhD, CSCS*D, and Mark D. Peterson, PhD, CSCS*D

Effective exercise prescription begins with an analysis to determine the needs of the client. Referred to as a needs analysis (National Strength and Conditioning Association 2000), this process involves determining the client's lifestyle and the demands of the sport, as well as identifying current and previous injuries and limitations, overall training experience, and the existing level of fitness and skill across a variety of fitness and athletic components. Without such data from which to provide baseline and follow-up evaluations, trainers and strength and conditioning professionals are inclined to design and implement cookie-cutter exercise programs created not for the individual, but for a large group of potential exercisers.

Conducting tests and assessing the collected data provides objective information regarding the strengths and weaknesses in a client's physiological and functional capacities. When done correctly, this process enables an exercise professional to develop the most effective and appropriate training program for the client. However, the process involves far more than simply collecting data. Gathering the appropriate data, analyzing it correctly, and presenting the information in a succinct and accurate manner are all important for the effective use of testing in a fitness or sport arena.

Sport Performance and Testing

Tests are conducted for a variety of reasons depending on the situation. Following are some examples in a professional setting:

- Identifying physiological strengths and weaknesses
- Ranking people for selection purposes
- Predicting future performances
- Evaluating the effectiveness of a training program or trial
- Tracking performance over time
- Assigning and manipulating training dosages (e.g., intensities, loads, volumes)

Exercise professionals can evaluate data to examine the overall effectiveness of an exercise routine. Specifically, strength testing data collected every month can be used to examine changes over time and to give an objective picture of the overall effectiveness of the strength training plan. If increases in strength are less than desirable, alterations may be made to enhance fitness adaptation during the subsequent training cycle.

Personal trainers may use test data to demonstrate and present improvements to clients and help them gain an understanding of the overall picture of the alterations in fitness brought about by their exercise programs. Alternatively, physical therapists might consult test data to determine appropriate rehabilitation progression timelines. When used properly, test data can help exercise professionals reach and maintain a higher level of practice.

Screening Tests

The first step in selecting components to include in the test battery is to determine the physiological components to be evaluated. Specific to the needs analysis, a preliminary assessment should include several additional tests to determine the client's exercise readiness. Depending on the client, this step requires a careful examination of potential sources of physical complications; this might involve a cardiovascular screening or an assessment of joint and posture mobility or integrity. Regardless of the client's age and training history, this pre-activity screening is a vital step in the needs analysis, and is necessary for identifying potential health risks of engaging in exercise, prior to the start of a program. Clearly, tests conducted to identify health risks are somewhat different from those used to simply gauge and monitor basic fitness. Nonetheless, these tests are all needed for creating effective programming and ensure client safety.

After the completion of a health risk appraisal, testing for current fitness is likely warranted. For personal trainers, this process is relatively straightforward, involving a thorough review of the client's health history and current health risk, as well as exercise and fitness goals. For strength and conditioning professionals, this step requires an intimate understanding of not only the tests needed for evaluating athlete preparedness, but also the fitness and performance benchmarks for that athlete to aspire to, to compete successfully in a given athletic endeavor.

To complete the testing process in a time- and energy-efficient manner, fitness professionals need to ensure that the tests are valid—that is, that they measure what they are intended to measure. A strength test should measure force production, whereas an endurance test should measure the ability to repeatedly exert force. From the many tests that have been developed and validated to measure specific health and fitness components, fitness professionals must select the most appropriate and valid one for a given client. They need to keep in mind that certain tests have been validated only for specific populations and may not be appropriate for people who are not in that classification. Therefore, caution must be used when selecting tests, because producing invalid results is very easy to do.

Tests must not only measure what they are supposed to measure, but also measure it consistently. Reliable tests result in consistent measures with a low opportunity for error. When using external raters or observers to measure performance measures, examiners should consider the reliability of each observer. To compare future test results to baseline results, fitness professionals must either ensure that the same observer conducts both tests or that multiple observers provide the same measure for a given performance. To verify consistency among raters, examiners can have all raters assess the same performance; this might reveal differences in ratings, or the extent of such differences.

Although many tests have been shown to be valid and reliable in a clinical or laboratory setting, some are not feasible in many work environments. Financial resources, time and space, as well as qualified staff to oversee testing are all factors that may determine the practicality of a specific test. However, examiners should consider alternatives for testing, because there are often multiple options for determining specific fitness and performance characteristics.

Validity, reliability, and feasibility should be the foremost considerations in test selection. Professionals who take all of these variables into account will get better, more useful information throughout their careers.

Data Evaluation and Statistical Analysis

Data collection represents only half of the overall process of testing and assessment. Once testing has been completed, data evaluation and interpretation must be conducted. Many fitness professionals are very good at conducting tests and storing information. However, where they frequently fall short is in the actual evaluation of the information they have collected, as well as in the subsequent use of those findings to inform their exercise prescriptions. Without an objective examination of the data, the full value of exercise testing cannot be realized.

Applied Statistics

Many fitness professionals view statistics as complex, useless mathematical equations. Although many complex equations and statistical procedures exist, and some of them do lack professional applications, applied statistics can offer an objective means for evaluating data. Developing a functional knowledge of statistics may require an investment of time and effort; however, the ability to perform even the most basic applied statistical analyses will greatly add to the fitness professional's toolbox of skills.

In statistics, very little emphasis is placed on one piece of data (e.g., one client's vertical jump score or one athlete's bench press 1RM). Instead, statistical evaluations focus on group dynamics. For instance, if one person in a group of 10 decreased performance after participating in an organized training program, but the other nine participants increased performance, we would not want to judge the program as ineffective simply because one person in the group did not experience a positive response. Yet, in the world of exercise prescription and programming, we must consider the individual responses. Although one treatment may work well for a large population, it may not be the most effective for a given person. However, if only one client improves, or if one client improves much more than the others do, promoting the training program as effective based on that one person is inappropriate (although many fitness professionals do this). Care is needed when applying statistical evaluations in the real world.

Probability Versus Magnitude

Two characteristics of collected data must be considered and understood when performing statistical evaluations. The first is the probability of the results. Probability represents the reproducibility of the findings and is presented as a probability value (α or the p value). This value can range from 0.0 to 1.0 in the research literature, and is often reported as $p \geq$ or ≤ 0.05. Further, this value represents the chance that the findings of the analysis were obtained erroneously. If the p value is equal to 0.05, then there is only a 5% risk of error and a 95% chance that the same findings would be achieved if the conditions were repeated.

The level of probability needed for reaching the predetermined significance level is set based on the amount of acceptable error. In medical research, in which decisions about drugs or treatment protocols carry life-altering consequences, less risk of error is allowed; α levels of .01 are generally used. In exercise science, in which differences in training programs or routines do not carry life-threatening consequences, it is common to accept levels of error at the 0.05 level. In either case, it is important to remember that the α level represents how many times we would expect different results if the study were repeated 100 times.

Probability is based on statistical power (most influenced by the number of people in the group being studied) and the variation in performance among the group. Although it is important to evaluate the reproducibility of a statistical analysis, it provides no measure of the actual magnitude of the change(s) in data. For instance, if 1,000 people were tested on the bench press 1RM, and then subsequently trained for three months to specifically improve bench press strength, examiners need to take into consideration the sample size and its influence on probability values to predict statistical difference, as well as the interpretation of the findings. If these people were retested and, as a group, demonstrated a 1-kilogram improvement in strength, the probability of generating the same results if the conditions were repeated would be high, because of the large number of participants in the group: perhaps even at $p < 0.01$. Assuming that this group was composed of members of the general population, even though a 1-kilogram improvement would be expected 99 times out of 100, and would therefore likely yield a statistically significant result, this improvement actually represents a very small increase in strength. Therefore, because it is not uncommon for average exercisers to see an increase of up to 15 kilograms in a three-month period, these findings would be clinically insignificant.

Ultimately, to describe and evaluate the magnitude of the improvement, we must rely on another calculation, usually the effect size (described later). The differences in these outputs must be understood because many errors have been made as a result of the incorrect assumption that probability is equal to magnitude.

Descriptive Analysis

The first step in evaluating a data set simply provides an overview of the data. This is done by calculating descriptive values such as the mean, median, mode, range, and variance. The average score (mean), which is calculated by summing all scores and dividing by the number of scores, represents the average score. The median represents the middle score and can be found by arranging scores in ascending order and finding the middle score. This represents the 50th percentile score, signifying that half of the scores fall above this score and half fall below it. The mode is the most frequently occurring score. These three measures of central tendency provide ample information for interpreting how a given subject's score compares to those of the group.

Measures of central tendency are often used to create normative data, calculated from tests conducted in a very large group. For instance, if we tested a group of 10,000 firefighters to see how many push-ups and sit-ups they can perform, and then calculated the average for the group (e.g., 50 in one minute), we could state that the norm for a firefighter is 50. We could then test other firefighters to see how they compared to the normative score measured in a larger group of firefighters. Although normative data provide a good comparison to peers, they provide no information regarding individuals' ability to perform a certain task. Do firefighters need to be able to perform 50 push-ups in one minute to do the job safely and effectively? Do they need to be able to do 100 push-ups in one minute? Measures of central tendency simply describe the typical performance of a group; they do not necessarily represent the optimal level of performance.

Another important consideration in comparing an individual's score to those of the group is the evaluation of the variability in the scores. As one example, a data range may be used (i.e., the high score minus the low score) to see how much of a spread exists among all scores. Another common measure of variability is the standard deviation, which is a calculation of how closely the data set clusters around the mean. In a normal distribution, in which scores are evenly distributed above and below the mean, 68.26% of all scores will fall within ±1 standard deviation from the mean, 95.44% will fall within ±2 standard deviations, and 99.74% will fall within ±3 standard deviations. Examining a single score, and subsequently determining how many standard deviations above or below the mean it falls, offers a greater perspective on the quality of that score.

Relationships Among Performance Variables

The ability to examine the relationship among variables is often of interest to exercise professionals. The way one variable changes in relation to another can provide valuable information. For instance, as cardiorespiratory fitness increases, the risk of heart disease decreases. This relationship has led to a greater focus on cardiorespiratory fitness promotion and the development of effective exercise programs.

Although a relationship among variables can be useful information, it is important to realize that such relationships do not represent a cause-and-effect situation. For instance, a strong correlation exists between shoe size and IQ: people with larger shoe sizes also tend to have higher IQ scores. However, having big feet does not impart intelligence. This relationship simply reflects the fact that as people grow and mature, they gain knowledge.

The correlation coefficient can be used to determine a linear relationship among variables. Consider the hypothetical relationship between body weight and vertical jump in the 10 subjects presented in table 1.1 and figure 1.1.

As body weight increases, gravitational force increases, making it more difficult to propel the body vertically. Therefore, people who are heavier are at a disadvantage. In the data provided, the correlation coefficient is $r = -0.85$. This calculation is made in the following way:

$$r = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{\sqrt{\left[n(\Sigma X^2) - (\Sigma X)^2\right]\left[n(\Sigma Y^2) - (\Sigma Y)^2\right]}}$$

in which
n = number of subjects,
X = variable 1, and
Y = variable 2.

The strength and direction of the relationship are important considerations when evaluating data. The strength of the correlation is determined

TABLE 1.1 Weight and Vertical Jump Data

Subject	Weight	Vertical jump
1	225	31
2	289	18
3	186	42
4	190	25
5	245	30
6	265	21
7	300	18
8	175	36
9	180	33
10	290	21

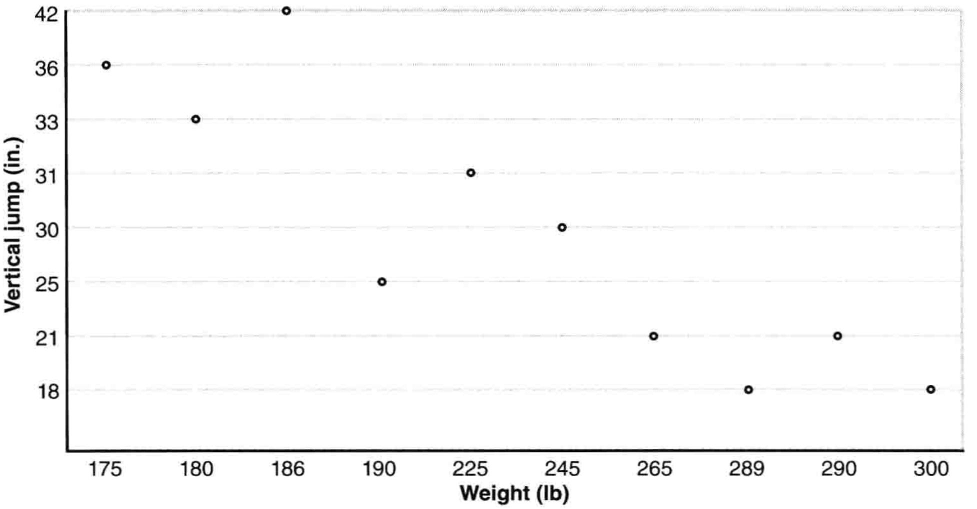

FIGURE 1.1 Plotting of weight and vertical jump data points.

TABLE 1.2 Scale for the Strength of a Correlation

Zero	0.0
Low	0.0–0.3
High	0.3–0.7
Perfect	1.0

by comparing the value to a scale ranging from 0 to 1.0 (see table 1.2; Morrow et al. 2000). The most frequently used statistic to evaluate the relationship between two variables on interval or ratio scales (e.g., the relationship between body weight in kilograms and vertical jump height in inches) is the Pearson product moment correlation coefficient. Calculation of the correlation coefficient relies on variance in both sets of data being analyzed for covariation (i.e., the degree to which the two variables change together). When two variables covary, they may be correlated to each other positively or negatively, which is indicated by a +/– designation. For the purposes of data interpretation, a larger correlation value (i.e., closer to 1.0 or –1.0) represents a stronger underlying association. Higher magnitudes of one variable occurring with higher magnitudes of another, and lower magnitudes on both variables, is a demonstration of a positive correlation. Conversely, two variables may covary inversely or oppositely, such as with a negative correlation (i.e., the higher magnitudes of one variable correspond with the lower magnitudes of the other, and vice versa). Thus, the relationship between weight and vertical jump from the sample data represents a strong negative correlation.

Differences Among Performance Variables

Determining differences among performance variables is often an important use of data collection and analysis. A variety of ways are available to objectively determine whether differences exist and to examine the magnitude of the actual difference. The technique used depends on the circumstances of testing. Examples of times when determining differences among measures might be desirable include a coach wanting to know whether athletes' strength levels are increasing, a physical therapist comparing two treatment strategies to see which is more effective, a trainer comparing changes in jump performance following a plyometric training program, and a researcher wanting to know the difference in performance level between major and minor league baseball players.

If the same group is compared, with measures taken before and after an intervention, a paired-samples t-test or a repeated measures analysis of variance would be used to examine changes. If different groups are placed on separate interventions, an independent-samples t-test or analysis of variance can be used. The statistical analysis evaluates the difference in