A Quantitative Analysis of the Effect of Resistance Training on Strength Test Score Variability

Ross R. Vickers, Jr.
Amanda C. Barnard
Linda K. Hervig

Naval Health Research Center
140 Sylvester Rd.
San Diego, California 92106-3521

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government. Approved for public release; distribution is unlimited.

This research has been conducted in compliance with all applicable federal regulations governing the protection of human subjects in research.
A Quantitative Analysis of the Effect of Resistance Training on Strength Test Score Variability

Ross R. Vickers Jr.
Amanda C. Barnard
Linda K. Hervig

Naval Health Research Center
140 Sylvester Road
San Diego, CA 92106-3521

e-mail: ross.vickers@med.navy.mil

Report No. 10-07 was supported by the Office of Naval Research, Arlington, VA, under Work Unit No. 60704. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government. Approved for public release; distribution is unlimited. Human subjects participated in this study after giving their free and informed consent. This research has been conducted in compliance with all applicable federal regulations governing the protection of human subjects in research.
Summary

This report extended an initial qualitative demonstration that test score variability increases during resistance training. Quantitative methods were applied to individual strength test data from 46 published studies. Analyses were limited to the four strength tests that were most often administered to experimental and control groups in the same study: bench press, leg press, biceps curl, and squat. A total of 97 contrasts of pretraining variation with posttraining variation were available for analysis because some studies administered more than one test and/or administered tests to more than one experimental group. Conducting separate analyses for each strength test eliminated statistical problems associated with having correlated observations. Resistance training increased test score variation on each of the four strength tests. Increased variation in test scores indicate a specific training programs are more effective for some individuals than others. This observation could be a point of departure for research to identify specific participant characteristics to guide decisions when matching individuals to training programs.
Resistance training increases the variability of strength test scores. A preliminary test of this assertion indicated that test score variability increased for 72.0% of tests administered to training groups compared with 49.5% of tests administered to control groups (Vickers, Barnard, & Hervig, manuscript in review). The odds ratio comparing training groups to control groups was statistically significant (odds ratio = 2.62, \( p < .001 \)). The trend applied equally to men and women. The trend was weaker as training experience increased—disappearing all together for highly experienced resistance trainers.

Increased test score variability implies that some individuals benefit more from a given training program than others (Bryk & Raudenbush, 1988). Participants who benefit little from a given program may achieve much better results in an alternative program. If so, training benefits would be maximized by assigning each individual to the program that will produce the greatest benefits for him or her.

Individual differences in response to training programs presumably arise from the interplay of program characteristics with participant characteristics. In educational research, this interplay is known as an aptitude-treatment interaction (ATI). For the present purposes, ATI will refer to attribute-treatment interaction. This change has been introduced to emphasize that the individual’s relevant characteristics are not necessarily limited to variables that would be thought of as aptitudes. If ATIs occur in resistance training, the interactions must be identified to match people to programs.

The existence of increased variability in test scores should be firmly established before searching for ATIs. The preliminary study of this topic was promising, but it had limitations. The preliminary study relied on qualitative comparisons of pre- and posttraining test score variation. It also treated all training groups and all control groups as though they came from different studies. When a study involved multiple strength measures, the preliminary study treated each measure as an independent observation. This follow-on study employed methods developed in educational research (Raudenbush, 1988) to address limitations of the earlier work. The analyses provide a quantitative test of the hypothesis that test score variability increases during resistance training.

**Methods**

**Data Sources**

Data came from 46 studies identified in an earlier meta-analysis of resistance training (Vickers et al., manuscript in review). These studies were a subset of the 196 studies that contributed to the earlier work. The subset was selected by applying two criteria. First, the research design had to contrast one or more experimental resistance training groups with a control group that did not train. Second, the strength tests administered in the study had to include the bench press, leg press, biceps curl, and/or squat tests.

**Strength Test Selection**

The four strength tests examined in this report, bench press, leg press, biceps curl, and squat, were selected because at least 10 comparisons were possible for each of these
tests. Restricting the analysis to tests represented by a relatively large number of estimates of the change in variability was expected to ensure reasonable statistical power for hypothesis tests.

A separate data analysis was carried out for each strength test. Separate analyses meant that each experimental group contributed only one observation to any given analysis, so the findings were based on independent experimental observations.

**Analysis Procedures**

The analysis followed Raudenbush’s (1988) procedures as illustrated by Bryk and Raudenbush (1988). A natural logarithm transformation of the standard deviations provided a measure that was approximately normally distributed with a known variance (Raudenbush & Bryk, 2002, p. 219). A correction for bias was added to the transformed variable. The corrected transformed standard deviation became the dependent variable in the data analyses.

Two analyses were conducted. One analysis simply compared the posttraining standard deviation of each experimental group with the posttraining standard deviation for the control group from that study. The second analysis compared the difference between the posttraining and pretraining standard deviation for the experimental group with the corresponding difference for the control group. Following Bryk and Raudenbush (1988), these analyses are referred to as the “Post” and “Gain” analyses.

The Gain analysis allowed for the correlation of pretraining test scores with posttraining test scores. This second analysis required estimates of the pre-post correlations. The estimates were not available from the primary studies that contributed to this paper, so estimates derived from the analysis of data available from several resistance training studies were used (Appendix B of Vickers et al., manuscript in review). The estimated correlations were $r = .90$ for the bench press, $r = .82$ for the leg press, $r = .77$ for biceps curl, and $r = .83$ for squats. The bench press, leg press, and biceps curl estimates were based on empirical evidence for these specific tests. Correlation estimates were not available for the squat, so the weighted average correlation for the other three tests was used for this measure.

The gain analyses will be more sensitive to experimental-control differences (Raudenbush, 1988) for the same reason that a correlated $t$ test is more sensitive than a simple between-groups $t$ test. The results for both analyses have been reported to provide the reader the opportunity to evaluate the importance of the correlation estimates.
Table 1

Tests for Increased Variability of Strength Test Scores

<table>
<thead>
<tr>
<th>Test</th>
<th>k</th>
<th>Post/Pre SD Ratio</th>
<th>δ^a</th>
<th>Z^c</th>
<th>Sig. d</th>
<th>δ^b</th>
<th>Z^c</th>
<th>Sig. d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bench Press</td>
<td>27</td>
<td>1.10</td>
<td>1.03</td>
<td>1.73</td>
<td>.021</td>
<td>2.14</td>
<td>.008</td>
<td></td>
</tr>
<tr>
<td>Leg Press</td>
<td>27</td>
<td>1.24</td>
<td>.99</td>
<td>2.20</td>
<td>.007</td>
<td>4.48</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Biceps Curl</td>
<td>27</td>
<td>1.19</td>
<td>1.04</td>
<td>4.90</td>
<td>.000</td>
<td>3.41</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Squats</td>
<td>16</td>
<td>1.13</td>
<td>.97</td>
<td>1.26</td>
<td>.052</td>
<td>2.42</td>
<td>.004</td>
<td></td>
</tr>
</tbody>
</table>

^aExperimental-control difference for Post analysis. ^bExperimental-control difference for Gain analysis. ^cZ = Method of adding weighted Zs (Rosenthal, 1978). ^dOne-tailed test of the hypothesis that variation was greater after training.

Results

Test score variability increased during resistance training (see Table 1). The increase ranged from 10% for the bench press to 24% for the leg press. The corresponding figures for control groups ranged from -3% to 4%. The control group’s median value was 1.01, a result that suggests for practical purposes variability did not increase for control groups.

As expected, the gain analysis provided stronger support for the hypothesis. Gain analysis Z values for the four strength tests ranged from Z = 2.14 to Z = 4.48, so the null hypothesis of no difference between the experimental and control conditions was rejected for each test. The positive signs of the average Z scores indicated that the standard deviation of the tests scores was larger after training than before.

The statistical leverage provided by repeated measures designs was not essential to the null hypothesis tests. Post analysis Z values indicated statistically significant trends except for the squat test. The trend for the squat test approached significance (p < .052).

Discussion

This quantitative analysis strongly supported the claim that resistance training increases strength test score variability. The effect was evident for each strength test considered here. When combined with the findings from the earlier qualitative study, the findings reinforced the view that resistance training increases test score variability across a variety of strength tests and training populations.

Program characteristics and changes in testing conditions or methods cannot explain the increased test score variation. These potential influences on test scores are constant within each study. ATIs are the remaining explanation for the increased variability of posttraining scores. The present analyses did not identify the relevant attributes of the individual trainees. Further study would be needed to determine whether those attributes are physiological (e.g., size), psychological (e.g., motivation), or both.
The Gain and Post analyses produced slightly different results. One reason for this is that the analyses rely on different definitions of the difference between the training and control group standard deviations. The Gain difference is:

$$\Delta_{\text{Gain}} = [\delta_{\text{post(train)}} - \delta_{\text{pre(train)}}] - [\delta_{\text{post(con)}} - \delta_{\text{pre(con)}}]$$

where $\delta$ is the bias-adjusted logarithmic transformation of the standard deviation and the subscripts indicate the measurement occasion (posttraining and pretraining) and the experimental condition (training and control). The Post difference is:

$$\Delta_{\text{Post}} = \delta_{\text{post(train)}} - \delta_{\text{post(con)}}$$

Clearly, $\Delta_{\text{Gain}} = \Delta_{\text{Post}}$ if, and only if, $\delta_{\text{pre(train)}} = \delta_{\text{pre(con)}}$. The latter equality will not hold in most cases. A second reason for the differences in the significance tests is that the variability of $\Delta_{\text{Gain}}$ will be less than that of $\Delta_{\text{Post}}$. This inequality arises because the variance estimate for $\Delta_{\text{Gain}}$ takes account of the correlation of pretraining test scores with posttraining test scores while the variance estimate for $\Delta_{\text{Post}}$ does not. The results of both analyses were reported to demonstrate that using estimates of the pre-/posttraining test score correlations was not critical for the primary study finding.

ATIs are a factor in resistance training. This is the logical inference from the evidence for increased test score variation following resistance training. Treatment factors are constant within a given study, so participant attributes must be the source of the differential response to training. It was not possible to search for critical attributes in this study because aggregate data were being analyzed. Still, the available evidence suggests productive lines of inquiry for identifying attributes that effect the training response. The training effect on test score variability has tended to weaken as the training experience of program participants increased (Vickers et.al, manuscript in review). Attributes that differ between trained and untrained individuals may influence the training response within studies. Even if this is not the case, the stronger effect of training on test score variability among untrained individuals indicates that studies of this population should be the most productive place to begin the search for relevant attributes.

The analyses in this paper corrected some of the earlier study limitations. The analyses provided a quantitative test of the study hypothesis and did so with observations that were largely independent within a given analysis. The observations were not completely independent within an analysis because the same control group sometimes was compared with more than one experimental group from the same study. In addition, some studies contributed data for more than one of the strength tests. For these reasons, the statistical significance tests are only approximate. Despite this caveat, the overall trends in the evidence covered in this report and the earlier qualitative investigation is strong enough to state with some confidence that resistance training increases the variability of strength test scores.

In summary, this report extended an initial qualitative demonstration that test score variability increases during resistance training. Quantitative methods were applied to individual strength tests to obtain results based on independent observations. The quantitative analyses reinforced the initial qualitative findings. Resistance training
increased test score variation on each of four strength tests—bench press, biceps curl, leg press, and squats. The increased variation in test scores indicates that a specific training program is more effective for some individuals than others. It may be possible to use this observation as a point of departure for identifying program participants’ characteristics that influence their response to training. Identifying those characteristics would be a step toward guidelines to match training programs to participants based on participants’ attributes.
References

*Indicates study that provided data for the meta-analysis.


This report extended an initial qualitative demonstration that test score variability increases during resistance training. Quantitative methods were applied to individual strength test data from 46 published studies. Analyses were limited to the four strength tests that were most often administered to experimental and control groups in the same study: bench press, leg press, biceps curl, and squat. A total of 97 pre- and post-training contrasts were available, this is because some studies administered more than one test and/or administered tests to more than one experimental group. Resistance training increased test score variation on each of the four strength tests. Increased variation in test scores indicate a specific training programs are more effective for some individuals than others. This observation could be a point of departure for research to identify specific participant characteristics to guide decisions when matching individuals to training programs.