THE RELIABILITY OF THE 12-ITEM GRIT SCALE AMONG CROSSFIT PARTICIPANTS

Michael Cazayoux¹,
Alan Bishop²,
James Navalta³,
Chad Harris⁴,
Kent Adams⁵,
Mark DeBeliso¹

¹Department of Kinesiology and Outdoor Education, Southern Utah University, Cedar City, UT, USA
²Athletics Department, University of Houston, Houston, TX, USA
³Kinesiology and Nutrition Sciences, University of Nevada Las Vegas, Las Vegas, NV, USA
⁴Metropolitan State University of Denver, Denver, CO, USA
⁵Kinesiology Department, California State University, Monterey Bay, Seaside, CA, USA

Abstract:
Crossfit participation has experienced exponential growth. Like other sports, it may be that successful participation in Crossfit would rely upon a construct known as grit. Grit is considered to be a combination of perseverance of effort (PE) and consistency of interest (CI) and has been assessed in other populations with the 12-point Grit Scale. The ability to reliably assess grit among Crossfit participants may allow participants and coaches to monitor the constructs of PE and CI in order to further develop Crossfit participation skills. Hence, the purpose of this study was to examine the reliability of the 12-point Grit Scale among adult Crossfit participants. The 12-point Grit Scale was electronically administered twice to 25 adult Crossfit participants (female:11, male:14) separated by at least two weeks. The results of the two administrations of the 12-point Grit Scale were then compared with a battery of statistical analyses in order to assess the reliability of the 12-point Grit Scale. The test-retest 12-point Grit Scale scores were: 49.8±5.3 and 50.2±5.4. The reliability statistics along with 90% confidence limits were as follows: interclass reliability coefficient was \( r=0.87 \) (UL:0.94, LL:0.73), intraclass reliability coefficient was ICC=0.91 (UL:0.96, LL:0.83), Standard Error of Measurement (SEm)=1.6 (UL:2.2, LL:1.3), ΔMeans 12-point Grit Scores= -0.3±2.3 (UL:0.6, LL:-1.2), and typical error CV% =3.4 (UL:4.8, LL:2.7). The Bland-Altman plot suggested agreement between the two administrations of the 12-point Grit Scale with no evidence of
heteroscedasticity. The results of the 12-point Grit Scale and subscales ranged from moderately-high to excellent reliability among adult Crossfit participants.

**Keywords:** grit, Crossfit, perseverance, consistency, reliability

1. Introduction

Crossfit® was founded in 2000 and has become one of the most popular fitness regimens across the planet (Rische, 2011). Crossfit is defined as “constantly varied functional movements performed at high intensity” (What is Crossfit, 2018). Crossfit® has altered the worldwide fitness industry having made general physical preparedness (GPP) and functional fitness the standard for people to achieve and maintain peak levels of fitness. Crossfit® introduced the sport of fitness in 2007 known as the “Crossfit Games”. The initial Crossfit Games commenced with 70 athletes at a ranch in California (Finding The Fittest On Earth, 2018). Since the initial gathering of the Crossfit Games participation has grown at an exponential rate. More than 380,000 participants registered for the 2017 Crossfit Open (2017 Open Recap, 2018). Paralleling the rapid growth of participation in the sport is the cash prize payout, which grew from $25K in 2010 to $2.2 million in 2017 (CrossFit Games Prize Purse Grows, 2018).

The inclusivity of Crossfit has contributed to the growth in participation and cult-like community (Dawson, 2017). In a Crossfit class, it would be typical to see a cross-section of the general population (ex. mother of three, a firefighter, a nurse, an elite athlete) all working out alongside one another. Greg Glassman the Crossfit® co-founder states: “The needs of Olympic athletes and our grandparents differ by degree, not kind,” which is a foundational tenet of the Crossfit training philosophy (What is Crossfit, 2018).

Crossfit targets development of a host of fitness domains (Figure 1) which include accuracy, agility, balance, coordination, cardiovascular-respiratory endurance, flexibility, power, speed, stamina, and strength (“What is Crossfit”, 2018). Participants who aspire to compete in the Crossfit Games must also develop competencies in weightlifting, gymnastics, and strongman.

While an athlete’s physical characteristics are of utmost importance, in order to compete in the Crossfit games, psychological traits likely play a critical role. Research by Gould et al. (2002) demonstrated that elite athletes have higher self-confidence, concentration, arousal control, motivation, and commitment. These same characteristics are likely to be important factors for achievement in the Crossfit Games. Likewise, Grit would appear to be an important construct as related to high achievement in Crossfit Games. Grit is a combination of persistence of effort (PE) and consistency of interest (Duckworth et al. 2007). If grit could be validly and reliably assessed among Crossfit participants, coaches and athletes could use that information to monitor the constructs of PE and CI in order to further develop Crossfit participation skills.

Tedesqui and Young (2017) provided evidence that PE was associated with “deliberate practice and higher skill group membership” while CI was associated with “less
thoughts of quitting/switching out of sport” in a sample of adult athletes. Tedesqui and Young (2017) assessed grit with the 12-point Grit Scale that was developed by Duckworth et al. 2007. However, sufficient evidence of the test-retest reliability of the 12-point Grit Scale is lacking. Duckworth et al. 2009 reported a test-retest reliability coefficient of r=0.68 in a sample of middle/high school students using the 8-point Grit Scale. Various researchers have suggested that reporting a Pearson correlation coefficient (r) alone is not sufficient to describe the test-retest reliability of an instrument (Bland & Altman, 1986; Hopkins, 2013). Given the recent concerns regarding reproducibility of research (Open Science Collaboration, 2015) and the potential importance of grit to Crossfit participants, we felt it relevant to perform an in-depth examination of the test-retest reliability of the 12-point Grit Scale among Crossfit participants.

2. Methods

2.1 Participants
Potential novice/recreational participants were contacted at a local Crossfit training facility in Texas, USA regarding their interest in participating in the study. Advanced games level participants were contacted regarding participating in the study through personal connections with the lead investigator. Advanced games level participants lived in the US, Canada, and Australia. The participants (female and male) ranged in age from 18-65 years and experience levels of 6 months-10 years and provided informed consent to participate. The study was reviewed and approved by a University Institutional Review Board for the ethical use of humans in research.

2.2 Instrument
The 12-point Grit Scale was developed and validated by Duckworth et al. (2007) to assess grit and the associated sub constructs of PE and CI for “adolescents and adults pursuing goals in a variety of domains”. The 12-point Grit scale has 6 questions related to PE and CI. The CI and PE subscales are comprised of scores from questions (2,3,5,7,8,11) and (1,4,6,9,10,12) respectively. Subscale scores are based on a 1-5 Likert scale. Total of the points for the 12 questions are added together and divided by 12. A score of 5 is the maximum score (i.e. extremely gritty), while 1 is the lowest score (i.e. not at all gritty). However, the analysis in the current study did not divide the total scores by 12 in order to establish a more precise analysis of the reliability of the 12-point Grit Scale and associated subscales.

2.3 Procedures
Participants that expressed interest (n=85) were emailed a link to SurveyMonkey.com. SurveyMonkey is an online survey company that provided the platform for collecting the 12-point Grit Scale information in the current study. Participants responded to the first administration of the 12-point Grit Scale in a time frame of 1-3 weeks. One week subsequent to the completion of the first administration of the 12-point Grit Scale a
second solicitation was sent regarding the second administration of the 12-point Grit Scale. Participants completed the second administration of the 12-point Grit Scale in a 1-3 week time frame. Subsequent to the completion of the second administration of the 12-point Grit Scale the data was arranged in an MS Excel spreadsheet for statistical analysis.

2.4 Statistical Analysis
Multiple statistical approaches were conducted to assess the test-retest reliability of the 12-point Grit Scale. The following procedures were carried out to compare the two sets of 12-point Grit Scale, PE and CI scores. A scatter plot was examined for the appearance of linearity; a Bland-Altman plot was constructed in order to examine error uniformity and limits of agreement (LOA). The interclass (r) and intraclass reliability coefficients (ICC), the mean difference between trials, and the standard error of measurement (SEm) were calculated as well as 90% upper and lower limits (UL, LL). Finally, the coefficient of variation percent (CV%, UL, LL) was also calculated to inspect the typical error from the log-transformed trial data. Statistical analysis was carried out in Microsoft Excel, with a spreadsheet prepared by Hopkins [2007].

3. Results
A total of 26 Crossfit participants responded to two electronic solicitations to complete the 12-point Grit Scale (two trials or test-retest). After reviewing the data we felt that one survey result was erroneously scored and as such the associated 12-point Grit Scale scores were discarded. Hence, the analysis of data was based on response from 25 Crossfit participants (female:11, male:14).

The mean and standard deviation for the 12-point Grit Scores and subscale scores are presented in Table 1. Neither the mean of 12-point Grit Scores nor subscale scores deviated by more than 0.5 of a point between trials.

Figure 2 is a scatter plot of the 12-point Grit Scale scores that suggests a strong linear relationship between the test-retest scores. Figure 3 is a Bland-Altman plot, comparing the average trial scores versus the difference in the scores as well as LOA. None of the data pairs exceeded the 95% LOA suggesting the 12-point Grit scale is reliable in the absence of bias or non-uniform error [Bland & Altman, 1986]. Further, no evidence of bias or non-uniform error was suggested by the scatter or the Bland-Altman plot.

Table 2, 3 and 4 provide the reliability statistics with 90% confidence limits (UL, LL) for the 12-point Grit Scale, subscale of PE, and subscale CI respectively.
4. Discussion

The purpose of this study was to examine the test-retest reliability of the 12-point Grit Scale and subscales of PE and CI among adult Crossfit participants. The various indices of reliability employed in this study suggest the 12-point Grit Scale and subscales of PE and CI ranged from moderate-high to excellent reliability.

Table 2, 3, and 4 shows the reliability statistics with 90% confidence limits (UL, LL) for the 12-point Grit Scale score, PE and CI. The ICC’s ranged from 0.76-0.91, which is considered as “good reliability” to “excellent reliability” [Koo & Li, 2016]. The interclass coefficients (r) ranged from 0.74-0.87 which is viewed as “moderate high” to “high” [Safrit & Wood, 1995]. The SEm ranged from 1.1-2.0 on the Grit scale (or 2.67-6.66 %). Data was Log-transformed as recommended by Hopkins [2013] in order to quantify typical error, noting that the 12-point Grit Scale, PE and CI scores did not suggest non-uniform error. The typical error (CV%) ranged from 3.4-11.6 percent.

Within the parameters of this study it is concluded that: the 12-point Grit Scale as well as the PE and CI subscales exhibit sufficient test-retest reliability such that Crossfit coaches and adult participants can monitor the constructs of PE and CI which may allow for the further development of Crossfit participation skills.

Conflict of Interest
No funding was received for this research. The authors have no conflict of interest related to this research.

References


Appendix

Table 1: Grit Trial Scores

<table>
<thead>
<tr>
<th></th>
<th>Total Score</th>
<th>Grit Sub-scale Perseverance Effort</th>
<th>Grit Sub-scale Consistency Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>12-Item Grit Scale</td>
<td>Score</td>
<td>Score</td>
</tr>
<tr>
<td>----</td>
<td>-----------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial 1</td>
<td>25</td>
<td>49.8±5.3</td>
<td>26.7±2.7</td>
</tr>
<tr>
<td>Trial 2</td>
<td>25</td>
<td>50.2±5.4</td>
<td>27.2±2.5</td>
</tr>
</tbody>
</table>

Note: 12-Item Grit and subscale scores (mean±SD). No significant difference between trials (p>0.05).

Table 2: 12-Item Grit Scale Reliability Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Upper Limit</th>
<th>Lower Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>∆ Means (trials)</td>
<td>-0.3±2.3</td>
<td>0.6</td>
</tr>
<tr>
<td>r</td>
<td>0.87</td>
<td>0.94</td>
</tr>
<tr>
<td>ICC</td>
<td>0.91</td>
<td>0.96</td>
</tr>
<tr>
<td>Typical Error (CV%)*</td>
<td>3.4</td>
<td>4.8</td>
</tr>
<tr>
<td>SEM</td>
<td>1.6</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Note: 90% Confidence limits for selected reliability statistics. *Typical error expressed as a CV% based on Log-transformed data. SEM- standard error of the measure. r- Pearson correlation coefficient. ICC- Intraclass correlation coefficient.

Table 3: Grit Sub-scale Perseverance Effort Reliability Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Upper Limit</th>
<th>Lower Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>∆ Means (trials)</td>
<td>0.5±1.6</td>
<td>1.2</td>
</tr>
<tr>
<td>r</td>
<td>0.78</td>
<td>0.90</td>
</tr>
<tr>
<td>ICC</td>
<td>0.83</td>
<td>0.91</td>
</tr>
<tr>
<td>Typical Error (CV%)*</td>
<td>4.3</td>
<td>6.0</td>
</tr>
<tr>
<td>SEM</td>
<td>1.1</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Note: 90% Confidence limits for selected reliability statistics. *Typical error expressed as a CV% based on Log-transformed data. SEM- standard error of the measure. r- Pearson correlation coefficient. ICC- Intraclass correlation coefficient.

Table 4: Grit Sub-scale Consistency Interest Reliability Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Upper Limit</th>
<th>Lower Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>∆ Means (trials)</td>
<td>-0.8±2.9</td>
<td>2.0</td>
</tr>
<tr>
<td>r</td>
<td>0.74</td>
<td>0.88</td>
</tr>
<tr>
<td>ICC</td>
<td>0.76</td>
<td>0.87</td>
</tr>
<tr>
<td>Typical Error (CV%)*</td>
<td>11.6</td>
<td>16.5</td>
</tr>
<tr>
<td>SEM</td>
<td>2.0</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Note: 90% Confidence limits for selected reliability statistics. *Typical error expressed as a CV% based on Log-transformed data. SEM- standard error of the measure. r- Pearson correlation coefficient. ICC- Intraclass correlation coefficient.
Figure 1: Crossfit participants in action

Figure 2: Scatter plot of trial 1 and 2 12-Item Grit scale scores
Figure 3: Bland-Altman plot of trial 1 and 2 12-Item Grit scale scores. Limits of agreement are the mean±2 SD in the differences in trial scores.