**Summary**

This paper reports on DIF analysis strategies applied in creating an improved Norwegian hardness (resilience) measure. Hardiness is a personality style marked by high levels of commitment, control and challenge (Kobasa, 1979), and has been linked to resilient, healthy responses to stress (Maddi & Kobasa, 1984). A short 15-item hardness measure (“Dispositional Resilience Scale” or DRS-15) was administered to groups of Norwegian and American military cadets in separate research studies. Simple Delta Plot methods, and more elaborate DIF techniques including Mantel-Haenszel and logistic regression analyses were applied to American and Norwegian data sets, revealing uniform and nonuniform differential item functioning of hardness items. While most of the DRS-15 items appear to be operating similarly in the Norwegian and American versions, five items show some evidence of DIF. These results are being used to inform a new revision of the Norwegian DRS-15 to measure personality hardiness, and have also led to improvements in the original English-language version.

**Introduction**

A brief 15-item hardness (“dispositional resilience”) scale developed by Bartone (1995) has been used extensively, with good results (e.g., Bartone, Ursano, Wright & Ingraham, 1989; Britt, Adler & Bartone, 2000). This scale was translated into Norwegian in 1998, and has since been used in multiple studies in Norway (Johnsen, Eid & Bartone, 2004; Bartone, Johnsen, Eid, Laberg & Brun, 2002). Results are promising, but there is evidence that the measure is in need of psychometric improvement in order to capture the concept of hardiness more effectively in the Norwegian language and culture (Johnsen, Eid, & Bartone, 2004). For example, the overall scale reliability (Cronbach’s alpha) in a sample of Norwegian Navy cadets was low (.61; Bartone et al., 2002), and reliability coefficients for the three facets were also somewhat low. A useful step in the test adaptation process is to conduct a DIF – Differential Item Functioning - analysis in order to identify items that may not be operating the same way in the two different languages and cultures (Allalouf, Hambleton & Sireci, 1999; Holland & Wainer, 1993). A number of techniques are available for this, including methods based on Item Response Theory (Thissen, Steinberg & Wainer, 1993). The DIF – Delta Plot method first elaborated by Angoff (1972) is still a very useful technique for identifying potential problem items (Angoff, 1993).

**Method**

In the present study, two similar samples of U.S. and Norwegian military cadets were used for a DIF analysis of hardness items. The U.S. cadets (N = 430) were in their senior year at the U.S. Military Academy – West Point when they completed the English version of the short hardness scale - DRS-15. The Norwegian cadets (N=297) all have some military experience before entering the Naval Academy, and are close in age (Mean=23.1 years, s.d.=2.6) to the West Point cadets (Mean=22.1, s.d.=1.1). Both samples were primarily men, with women a comparable minority in the U.S. (13%) and Norwegian (10%) samples. Mean overall hardness levels were slightly higher in the American sample (Mean=30.76, s.d.=4.65) compared to the Norwegian (Mean=30.03, s.d.=4.42).

**Results – over →**

**Discussion**

Results of the DIF analysis were followed up with a detailed conceptual analysis of Norwegian and English items by bilingual experts, and also administration of both language forms to available groups in a bilingual research design. This has led to an improved Norwegian version, now under evaluation with Norwegian forces deployed to Lebanon and Afghanistan. These results and additional validity data will be presented on the improved Norwegian Dispositional Resilience Scale.

Resiliency under stress is a consequential matter in many occupations. The costs of non-adaptive responses to stress can be devastating not only for individuals, but also for organizations. The personality style of hardiness has been linked to resilient responding, and helps to explain how it is that some people remain healthy, while others become symptomatic when exposed to highly stressful conditions. More research is needed to clarify how hardiness and other possible resiliency factors operate. The present study contributes to this line of investigation by providing improved measures for instrumenting hardiness. Ultimately, the results of this work will allow practitioners to design more effective training programs for building-up the kinds of healthy cognitions and behaviors that resilient persons display in response to stress.

**References**


*Presented at the Association for Psychological Science Annual Convention, Washington, DC, May 25, 2007. This project was supported in part by the U.S. – Norwegian Fulbright Foundation. For further information contact: bartonep@ndu.edu or paul.bartone@psyssp.ubn.no
Results of the DIF delta-plot analysis (Holland & Wainer, 1993) suggest five hardiness items (out of 15) that show evidence of DIF. These are items 12 and 7, which the Norwegian sample responds more positively to, and items 1, 15 and 2, which the U.S. sample responds more positively to. Further DIF analysis were conducted using the "EZDIF" program developed by Niels Waller, which yields Mantel-Haenszel as well as logistic regression (odds-ratio) statistics, effect sizes and statistical significance (Waller, 2004). Here, only four hardiness items showed significant DIF. While caution is appropriate (considering sample specificity) in interpreting these results, several items were identified that should be closely scrutinized for possible modification or elimination in an improved Norwegian hardiness scale. DIF techniques merit wider application in the psychological test adaptation process.

DIF analysis by the delta-plot technique (also known as “transformed item difficulty” or TID) followed these steps. Item responses were collapsed into binary categories, with codes 0 or 1 = 1, and 2 or 3 = 2. Next, the proportion of 2 (positive) responses was calculated for each item, within each sample. These scores were standardized within samples and converted to the standard metric used by ETS (Educational Testing Service), with Mean=13 and s.d.=4, as recommended for delta plots (Dorans & Holland, 1993). Finally, delta scores for each group were plotted, with the American scores on the x – horizontal axis, and Norwegian scores on the y – vertical axis. Each data point represents the delta scores for both groups on a particular hardiness scale item. A line is drawn diagonally approximating the regression line that minimizes the distances between points. This line shows the area of perfect agreement. If both samples answered items exactly the same way, all points would fall directly on the line.

Results of the DIF delta-plot analysis (Holland & Wainer, 1993) suggest five hardiness items (out of 15) that show evidence of DIF. These are items 12 and 7, which the Norwegian sample responds more positively to, and items 1, 15 and 2, which the U.S. sample responds more positively to. Further DIF analysis were conducted using the "EZDIF" program developed by Niels Waller, which yields Mantel-Haenszel as well as logistic regression (odds-ratio) statistics, effect sizes and statistical significance (Waller, 2004). Here, only four hardiness items showed significant DIF. While caution is appropriate (considering sample specificity) in interpreting these results, several items were identified that should be closely scrutinized for possible modification or elimination in an improved Norwegian hardiness scale. DIF techniques merit wider application in the psychological test adaptation process.

DIF analysis by the delta-plot technique (also known as “transformed item difficulty” or TID) followed these steps. Item responses were collapsed into binary categories, with codes 0 or 1 = 1, and 2 or 3 = 2. Next, the proportion of 2 (positive) responses was calculated for each item, within each sample. These scores were standardized within samples and converted to the standard metric used by ETS (Educational Testing Service), with Mean=13 and s.d.=4, as recommended for delta plots (Dorans & Holland, 1993). Finally, delta scores for each group were plotted, with the American scores on the x – horizontal axis, and Norwegian scores on the y – vertical axis. Each data point represents the delta scores for both groups on a particular hardiness scale item. A line is drawn diagonally approximating the regression line that minimizes the distances between points. This line shows the area of perfect agreement. If both samples answered items exactly the same way, all points would fall directly on the line.

**Figure 2: Sample output from Mantel-Haenszel and Logistic Regression Analysis of DIFFERENTIAL ITEM FUNCTIONING**

**EZDIF analysis of Norwegian vs West Point cadet hardiness data**

Reference Group: F:/dif-anal/us-436-bi.dat
Focal Group: F:/dif-anal/no-297-bi.dat
Number of Cases in Reference Group: 436
Number of Cases in Focal Group: 297

**Conditioning Levels**

0 4 8 12
3 7 11 15

**Note:**

Alpha > 1.00 favors Reference Group; Alpha < 1.00 favors Focal Group
D-DIF < 0.00 favors Reference Group, D-DIF > 0.00 favors Focal Group

**Results for Pass Number: 1**

**WARNING: Insufficient Data Found in Level: 0 - 3**

**ETS (effect size) Code:**
A = negligible; B = Moderate; C = Large

<table>
<thead>
<tr>
<th>Code</th>
<th>ITEM</th>
<th>Alpha</th>
<th>X^2</th>
<th>P-Value</th>
<th>MH D-DIF</th>
<th>(MH D-DIF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>1***</td>
<td>2.045</td>
<td>14.330</td>
<td>0.000</td>
<td>-1.681</td>
<td>0.440</td>
</tr>
<tr>
<td>CR</td>
<td>2***</td>
<td>2.850</td>
<td>19.864</td>
<td>0.000</td>
<td>-2.461</td>
<td>0.556</td>
</tr>
<tr>
<td>A</td>
<td>3</td>
<td>1.188</td>
<td>0.781</td>
<td>0.377</td>
<td>-0.405</td>
<td>0.417</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>2.214</td>
<td>3.265</td>
<td>0.071</td>
<td>-1.686</td>
<td>0.952</td>
</tr>
<tr>
<td>A</td>
<td>5</td>
<td>1.311</td>
<td>1.886</td>
<td>0.170</td>
<td>-0.637</td>
<td>0.435</td>
</tr>
<tr>
<td>A</td>
<td>6</td>
<td>0.708</td>
<td>1.384</td>
<td>0.239</td>
<td>0.810</td>
<td>0.623</td>
</tr>
<tr>
<td>CF</td>
<td>7***</td>
<td>0.369</td>
<td>20.844</td>
<td>0.000</td>
<td>2.341</td>
<td>0.513</td>
</tr>
<tr>
<td>A</td>
<td>8</td>
<td>1.166</td>
<td>0.446</td>
<td>0.504</td>
<td>-0.361</td>
<td>0.469</td>
</tr>
<tr>
<td>A</td>
<td>9</td>
<td>1.529</td>
<td>0.976</td>
<td>0.366</td>
<td>-0.666</td>
<td>0.492</td>
</tr>
<tr>
<td>B</td>
<td>10*</td>
<td>0.298</td>
<td>0.093</td>
<td>0.224</td>
<td>2.924</td>
<td>1.242</td>
</tr>
<tr>
<td>A</td>
<td>11</td>
<td>0.828</td>
<td>0.784</td>
<td>0.376</td>
<td>0.444</td>
<td>0.453</td>
</tr>
<tr>
<td>CF</td>
<td>12***</td>
<td>0.434</td>
<td>20.456</td>
<td>0.000</td>
<td>1.964</td>
<td>0.431</td>
</tr>
<tr>
<td>A</td>
<td>13</td>
<td>0.788</td>
<td>1.170</td>
<td>0.279</td>
<td>0.559</td>
<td>0.473</td>
</tr>
<tr>
<td>A</td>
<td>14</td>
<td>1.199</td>
<td>0.674</td>
<td>0.412</td>
<td>-0.367</td>
<td>0.403</td>
</tr>
<tr>
<td>CR</td>
<td>15***</td>
<td>3.066</td>
<td>25.241</td>
<td>0.000</td>
<td>-2.633</td>
<td>0.528</td>
</tr>
</tbody>
</table>

**Number of Items Purged in Pass 1: 4**

**ETS (effect size) Code:**
A = negligible; B = Moderate; C = Large

**Figure 1: DIF – Delta Plot Analysis of Hardiness Scale American vs. Norwegian Military Cadets**