**Introduction**

The Beta–4 Manual Supplement provides additional information referred to in the Beta–4 Manual. This supplement is intended to augment the manual, allowing it to remain simple, uncluttered by detail, and portable.

**Test–Retest Reliability by Age Band**

Table 1 presents the mean Beta–4 IQ scores and SDs for the first and second testing, the test–retest stability coefficient corrected for the normative sample’s variability (Allen & Yen, 2002; Magnusson, 1967), the standard difference (i.e., effect size) between the first and second test sessions, and the standard error of measurement (SEM) for the Beta–4 IQ for each of two age bands (16–44 and 45+). The values for the overall sample appear in the Beta–4 Manual, Table 4.2.

<table>
<thead>
<tr>
<th>Ages 16–44</th>
<th>First testing</th>
<th>Second testing</th>
<th>$r_{xx}$</th>
<th>Corrected $r$</th>
<th>Standard difference</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta–4 IQ</td>
<td>97.6</td>
<td>13.8</td>
<td>104.5</td>
<td>15.6</td>
<td>.89</td>
<td>.91</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ages 45–99+</th>
<th>First testing</th>
<th>Second testing</th>
<th>$r_{xx}$</th>
<th>Corrected $r$</th>
<th>Standard difference</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta–4 IQ</td>
<td>103.2</td>
<td>15.0</td>
<td>108.5</td>
<td>16.4</td>
<td>.91</td>
<td>.91</td>
</tr>
</tbody>
</table>

**SEM and Confidence Interval Formulas**

The formulas for calculating the SEM and confidence intervals use various methods, including one used to develop the confidence intervals in the Beta–4 Manual, Table A.2. This section describes these methods.

The SEM is calculated with the formula:

$SEM = SD \sqrt{1 - r_{xx}}$

where $SD$ is the standard deviation of the scale, and $r_{xx}$ is the reliability coefficient of the scale.

The SEM is used to calculate confidence intervals, or bands of scores around observed scores, in which true scores are likely to fall. Confidence intervals express test score precision and serve as reminders that measurement error is inherent in all test scores and that observed test scores are only estimates of true ability. Confidence intervals can be used to report an examinee’s score as an interval that is likely to contain the examinee’s true score.

There are multiple methods that can be used to calculate confidence intervals, and they all produce slightly different confidence interval values. All confidence interval calculation formulas take the SEM into account. Some local education agencies and other practice settings prefer to calculate confidence intervals in the most parsimonious manner, based only on the observed score and the SEM. Confidence intervals based on the observed score and the SEM are calculated by the following formula:

$p\%$ Confidence Interval = Observed Score $\pm z_p(SEM)$

where $p$ is the confidence level (e.g., 90% or 95%), and $z_p$ is the z value associated with the confidence level, located in normal probability tables.
Depending on the preference of the practitioner or practice settings, the SEM for either the overall sample or for the examinee’s age group can be used to calculate confidence intervals centered on the observed scores. The examples provided in this section utilize SEMs based on the overall sample. If an examinee obtained a Beta–4 IQ score of 108, the practitioner can be 95% confident that the examinee’s true Beta–4 IQ score falls in the range of 99–117 (because the 95% confidence interval is $108 \pm 1.96 \text{SEM}$, where the SEM is 4.74), and 90% confident that the examinee’s true Beta–4 IQ score is in the range of 100–116 ($108 \pm 1.65 \text{SEM}$).

The confidence intervals for the Beta–4 IQ score ranges reported in the Beta–4 Manual, Table A.2, were calculated around the estimated true scores using the standard error of estimation (SEE), according to the method proposed by Dudek (1979) and Glutting, McDermott, and Stanley (1987). The estimated true score is obtained by the formula:

$$\text{Estimated True Score} = 100 + r_{xx}(X - 100)$$

where $X$ is the observed Beta–4 IQ score and $r_{xx}$ is the reliability coefficient of the Beta–4 IQ score. The SEE is derived by the formula proposed by Stanley (1971):

$$\text{SEE} = SD \sqrt{1 - r_{xx}}$$

where SD is the theoretical standard deviation of the IQ score, and $r_{xx}$ is the reliability coefficient of the IQ score.

This method centers the confidence interval on the estimated true score rather than on the observed score, and in turn, results in an asymmetrical confidence interval around the observed score. This asymmetry occurs because the estimated true score typically is closer to the mean of the scale than is the observed score. A confidence interval based on the estimated true score and the SEE is a correction for true-score regression toward the mean. For example, if an examinee obtains a Beta–4 IQ score of 113, the examinee’s estimated true score will be 111.7, the 95% confidence interval of the examinee’s true score will be 103–120 (because the 95% confidence interval is $111.7 \pm 1.96 \text{SEE}$, where the SEE is 4.27), and the 90% confidence interval of the examinee’s true score will be 105–119 ($111.7 \pm 1.65 \text{SEE}$).

### Confirmatory Factor Analysis Interpretation

One of the most important outcomes of factor analysis is an understanding of the number and nature of the factors necessary to explain how the tests interrelate. For example, does the evidence suggest that there are one, two, or three distinct ability dimensions influencing performance on all the tests? Factor analysis is a tool for identifying the fewest factors that account for the data. Is there evidence that particular cognitive abilities (e.g., nonverbal ability) are among those dimensions? Examination of the content of the tests that are strongly related to a factor supports inferences about the nature of the ability represented by that factor. In this way, factor analysis links the measures in a test battery to existing measurement models.

The factor model proposed for a scale can be evaluated with a type of structural equation modeling (Jöreskog & Sörbom, 1993) known as confirmatory factor analysis. This technique is designed to evaluate a factor structure specified by researchers on the basis of theory and empirical research. The specific relations between observed variables (test scores) and latent variables (constructs such as nonverbal reasoning), and among the latent variables are specified in the model. In this sense, confirmatory factor analysis is different from exploratory factor analysis, because the tests are sorted into groups representing different factors in advance rather than generating the groups by applying a statistical algorithm to the data. The prespecified model is tested to determine if it provides a reasonably good, yet parsimonious, explanation of the actual correlations among the tests. The outcome of the analysis may suggest ways that the model could be improved to do a better job of explaining the data; however, such changes would be made only after careful consideration of whether they align with theory.

For several reasons, the technique of confirmatory factor analysis is preferred to exploratory factor analysis when an explicit theory of the factor structure is present or when there are competing models in the research literature (Schmitt, 2011; Stevens, 1996). The confirmatory method provides a numerical evaluation of the quality of a specified model, and it enables the researcher to compare the effectiveness of different models. Also, when the hypothesized ability dimensions are significantly correlated with one another, the confirmatory approach is more robust than exploratory factor analysis, whose results can be quite sensitive to small differences in the pattern of correlations.
FIT STATISTICS

The effectiveness of a factor model is evaluated using goodness-of-fit statistics. Most such indicators are based on the chi-square statistic. The likelihood-ratio chi-square is a statistic used to test the null hypothesis that the matrix of subtest variances and covariances implied by the factor loadings and factor intercorrelations is the same as the observed variance-covariance matrix (Byrne, 2001; Schumacker & Lomax, 2004; Thompson, 2000). Therefore, if the model adequately fits the data, the null hypothesis is not rejected.

However, when the sample is large, chi-square tends to detect differences between the actual and implied variance-covariance matrices and to reject the null hypothesis even when the differences are small. For that reason, evaluation of model fit relies on other statistics that are not sample-size dependent (Byrne, 2001; Schumacker & Lomax, 2004; Thompson, 2000). Chi-square remains necessary, however, because of its utility in testing the statistical significance of a difference in fit between two models.

The Beta–4 analyses report additional model-fit statistics that are commonly used to determine the degree to which the observed and implied variance-covariance matrices are equivalent. The comparative fit index (CFI; Bentler, 1990) evaluates how well the specified model reproduces the observed variance-covariance matrix as compared to a model that presumes that all of the tests are uncorrelated. Hu and Bentler (1999) recommend a CFI value of .95 or higher as a criterion for good fit. The Tucker-Lewis non-normed fit index (TLI; Tucker & Lewis, 1973) measures the improved fit in a manner similar to the CFI, with an adjustment for the degrees of freedom in the model. Steiger’s (1990) root mean squared error of approximation (RMSEA) is used to explore model fit relative to a population covariance matrix. The RMSEA is an estimate of how well the estimated variance-covariance matrix matches the population matrix, adjusting for the degrees of freedom in the fitted model. Browne and Cudeck (1993) suggest that a RMSEA value of .05 or less indicates a close model fit and that values of up to .08 represent adequate model fit with reasonable errors of approximation in the population. The Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978) are also included. When two models are compared on these statistics, smaller values suggest a better fit.
References


Appendix A. Special Group Criteria

Individuals were eligible to participate if they met all of the following criteria:

Primary language is English (with the exception of the Spanish-speaker and the English as a second language groups);

Able to communicate at a level commensurate with diagnosis, and not completely uncommunicative;

Normal hearing and vision (with aid);

Normal fine and gross motor ability (with the exception of mild motor impairment occurring in the intellectual disability group);

No physical conditions, illnesses, or impairments that could impact cognitive functioning or test performance (with the exception of conditions or impairments associated with a specific special group);

No diagnosis of a neurological condition (e.g., seizure disorder, epilepsy, encephalitis, brain surgery, brain tumor) other than the condition of interest or as allowed for a given special group;

No period of unconsciousness not related to surgery or greater than 20 minutes related to a medical condition;

No diagnosis of intellectual disability (with the exception of the intellectual disability group) or of a pervasive developmental disorder;

No diagnosis of a psychiatric disorder (e.g., psychotic disorders, mood disorders) other than that defined by the special group criteria;

Not currently admitted to a hospital, inpatient treatment, or psychiatric facility (with the exception of specified treatment settings that are pertinent to a given condition of interest [e.g., placements for correctional group or intellectual disability group]);

Not currently taking medication that might impact test performance, except as appropriate to treat condition of interest or associated conditions;

Meets the criteria for one special group only (with the exception of dual diagnoses of ADHD secondary to intellectual disability); and

Has not completed the Beta III or any other measure of cognitive ability in the 6 months prior to the testing date.

SPECIFIC INCLUSION CRITERIA FOR SPECIAL GROUPS

Correctional

Participation criteria included:

currently incarcerated in a correctional facility.

Intellectual Disability

Participation criteria included:

meets DSM-5™ criteria for a current diagnosis of intellectual disability, mild or moderate severity;

OR

full scale score 2–4 SDs below the mean on a standardized, individually-administered measure of cognitive ability (e.g., IQ = 40–70).
Attention-Deficit/Hyperactivity Disorder

Participation criteria included:

- meets DSM-5 criteria for a current diagnosis of ADHD;
- has not taken ADHD medication for at least 24 hours prior to testing.

Spanish Speakers

Participation criteria included:

- endorsed Spanish as his or her primary language.

English as a Second Language

Participation criteria included:

- speaks and understands English well;
- endorsed another language as his or her primary language.

Specific Learning Disorder or Disability

Participation criteria included:

- history of specific learning disorder, by self-report;
- OR
- history of having received school services for specific learning disability, by self-report.
Appendix B. Beta–4 Examiners and Sites

BETA–4 EXAMINERS

Dawn Hafling Magers  Shelly Stripling  Heather Giao
James P. Simonds  Maleah Bufford  Patricia Guarini-Reyes
Christopher Sumner  Terri Garmon  Amy Holloway
Sarah Kennedy  Wendy Mickley  Jane Russell
Barbara Vines  Patricia Hollinger  Khristy Garrett
Sylvia Sanchez  James Wahrer  Brenda (BJ) Crossley
Patrick Dallas  Ronni Laing  Cheryl O’Heir
Bonnie Campbell  Cathy Abels  Lisa Leon
Tracy Larson  Lori A. Bailey  Beth Miller
Lisa Brockhuizen  Rena Wechter  Jacqueline Johnson
Linda Guttman  Ashley Holt  Monica Galvez
Aubrey M. Johnson  Elaina LeGros  Ana Magallanes
JoAnn Salerno  Kalli Kronmiller  Christopher Zuccaro
Jamira Grana  Marie Anne Hoffmann  Jacqueline Harrison
Kevin Crippen  Germaine Graham  Bronwyn Guiles
Yvette Lightbourn  Bobbi Smith  Diana Iborg
Yael Horowitz  Crystal Joson  Kelly Paolisso
Gail Tonkovich  Patricia Sullivan  Stefanie Nance
Loraine Alderman  Valerie Caruso  Joanna Meyer
Denise Fancher Zaldana  Barbara Rhinehart  Mariangela Cavagna
Jennifer Japhet  Kathryn Salter  Melissa Thompson
Amy Warren  Nancy Klein  Madeline Wesh
Ann Roberts  Brittany Ubben  Darlene Gustavson
Candace McKinley  Maria Schreiber  Kylie Walker
Paige Beal  Alejandro Mancha  Melissa Ussery
Sheelah Zink  Mikala Wilder  Christine Kane
Helen Stayna  Misty Lay  Rachel Burkhalter
Julie Ferguson  Danita Lackey  Reginald Mitchell
Lori Liebing  Krysta Adams  Sharna Wood
Heidi Benson Rodriguez  Alanna Lipinski  Barbara Shumard
# BETA-4 SITES

<table>
<thead>
<tr>
<th>Organization</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFG &amp; Associates, LLC</td>
<td>Round Rock, TX</td>
</tr>
<tr>
<td>Tracy Larson Private Practice</td>
<td>Newark, NY</td>
</tr>
<tr>
<td>Dr. Loraine Alderman Private Practice</td>
<td>East Meadow, NY</td>
</tr>
<tr>
<td>Sandilands, Inc.</td>
<td>Detroit, MI</td>
</tr>
<tr>
<td>Therapy Group of Waco</td>
<td>Waco, TX</td>
</tr>
<tr>
<td>CD Fulkes Middle School</td>
<td>Round Rock, TX</td>
</tr>
<tr>
<td>Victor Child Care</td>
<td>Victor, NY</td>
</tr>
</tbody>
</table>