October 2023

Bayley-4 GSV Technical Supplement

Mark H. Daniel, PhD

Abstract

This report provides a technical description of the Rasch analyses used for each of Bayley-4 subtests to develop growth scale values (GSVs). Using a partial-credit model and joint maximum likelihood estimation, Rasch analysis was done from the results of the standardization sample of 1,700 children between the ages of 16 days and 42 months. Both person ability and item-threshold difficulty were identified in the sample. The Rasch logit values that represent ability and difficulty were linearly transformed into GSVs. The change in item success rate as GSV changes is identified, and diagnostic information about the Rasch calibration, such as item fit, local independence, and dimensionality, is discussed.

Cite this document as:

Daniel, M. H. (2023). *Bayley-4 GSV technical supplement* [GSV Technical Report]. NCS Pearson.

Pearson

Product Number 66782 v.B

Contents

Test Content and Item Type

The *Bayley Scales of Infant and Toddler Development,* (4th ed.; Bayley-4; Bayley & Aylward, 2019) measures cognitive, language, motor, social-emotional, and adaptive behavior characteristics of children ages 16 days to 42 months. The first three of these areas are assessed by five individually administered subtests:

- Cognitive (81 items)
- Language
  - Receptive Communication (42 items)
  - Expressive Communication (37 items)
- Motor
  - Fine Motor (46 items)
  - Gross Motor (58 items)

The items on these subtests are scored on a three-point scale where 0 = failure, 1 = partial success, and 2 = complete success, as judged by the examiner. A Rasch scaling of each of these subtests was used to develop growth scale values (GSVs) for each subtest. This report provides a technical description of those Rasch analyses and their results.

Social-emotional skills and adaptive behavior are assessed through a caregiver questionnaire, which is not discussed in this report.

Rasch Analysis and Applications

Method and Sample

Rasch analysis was done with the Winsteps software program (version 5.1.4; Linacre, 2021a), using the partial-credit model (Wright & Masters, 1982) and joint maximum likelihood estimation. The data for the Rasch analysis came from the standardization sample, updated in 2023, consisting of 1,700 children ages 16 days to 42 months. The updated sample was representative of the U.S. population by sex, race/ethnicity, and parent education. It consisted of children without known clinical conditions, except for 1.2% with Down syndrome who were included to increase variance at the low end of the score distribution. Details may be found in the Bayley-4 *Technical Manual* (Bayley & Aylward, 2023).

In general, each child started at an age-based starting point expected to be very easy. If they did not set a "basal" by scoring 2 on each of the first three items, they reverted to the next lower starting point, and repeated this process until setting a basal. For all children, testing stopped when the child scored 0 on five consecutive items. Unadministered items were assigned a score of 2 if they preceded the basal or a score of 0 if they followed the discontinue point. Because there were no missing scores in any of the administered items, the Rasch analysis did not include missing data.

Item Difficulty and Person Ability

In the Rasch model, the probability that a person will succeed on a dichotomously scored item depends on the difference between the person's ability and the item's difficulty. Both ability and difficulty are measured on the same scale whose units are called *logits.* Higher logit values correspond to higher abilities and more difficult items. As ability increases relative to difficulty, the probability of success increases. When person ability and item difficulty are equal, the person has an equal (.50) probability of failing or passing the item. When the item difficulty is within two logits of a person's ability, the person's probability of success is between .12 and .88. Items within this range are moderately difficult for the person and provide more information about their ability than items that are very easy or very difficult.

If items are polytomous (i.e., have more than two score categories) rather than dichotomous, the same principle applies, but at the level of transitions between score categories. An item *threshold* refers to the level of ability at which a person has an equal probability of earning the lower or higher of two adjacent score categories. Each Bayley-4 item has three score categories and, therefore, two thresholds. Thresholds can be operationalized in several ways, but for the current purpose, it is useful to employ the Rasch-Thurstonian method in which the first threshold difficulty is the ability level where the person is equally likely to score 0 or greater than 0, and the second is the ability at which there is an equal likelihood of scoring 2 or less than 2. Each item also has an overall difficulty value which is the average of the two threshold difficulties.

Figure 1 presents a Wright map for each subtest, showing the frequency distribution of item-threshold difficulties along with the distribution of abilities of the children in the calibration sample. By comparing the upper and lower distributions, one can see how many moderately difficult item thresholds a child with a given level of ability is likely to encounter on that subtest. Each Bayley-4 subtest contains item-threshold difficulties spanning virtually the entire range of abilities of the children in the standardization sample, meaning that virtually all children in this age range will encounter at least a few moderately difficult item thresholds.

Figure 1. Wright Map of Person Abilities and Item-threshold Difficulties in Logits, by Subtest

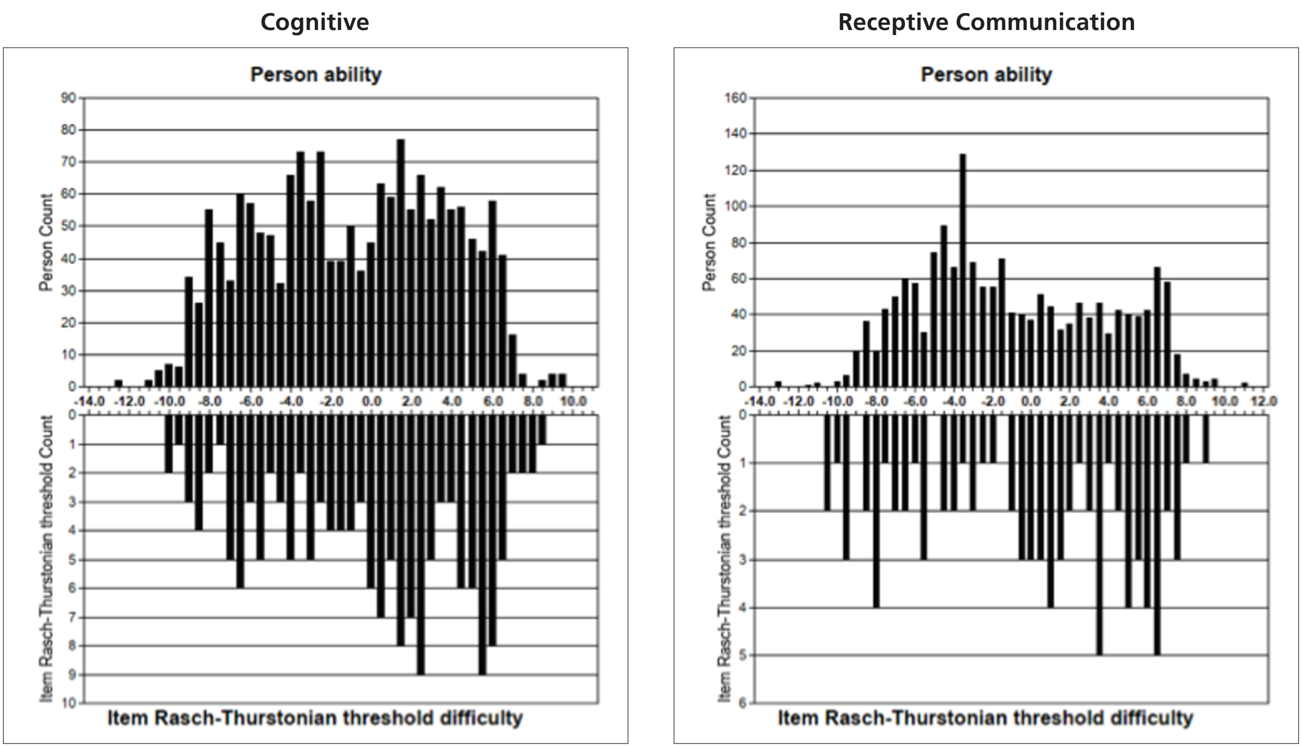
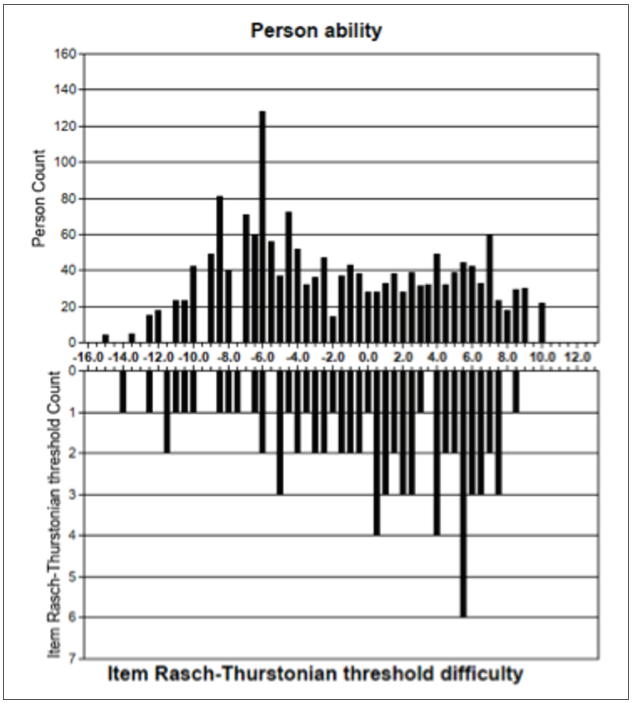
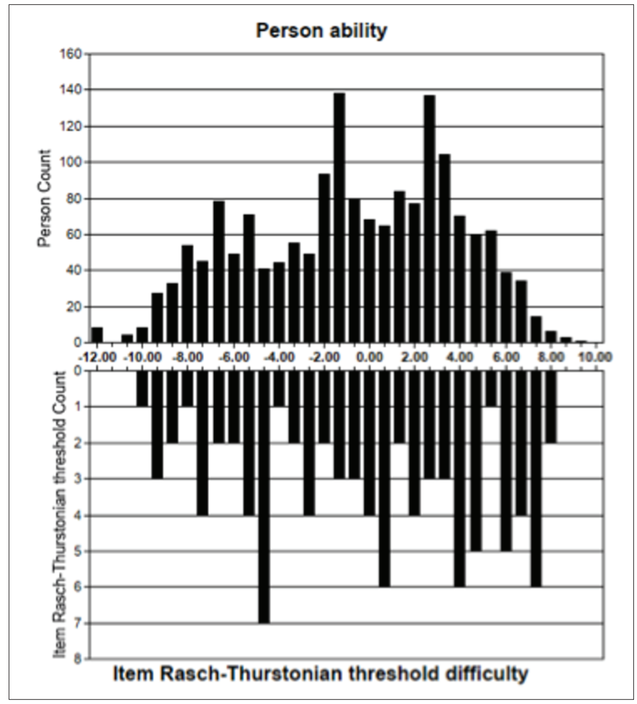


Figure 1. Wright Map of Person Abilities and Item-threshold Difficulties in Logits, by Subtest (*continued*)

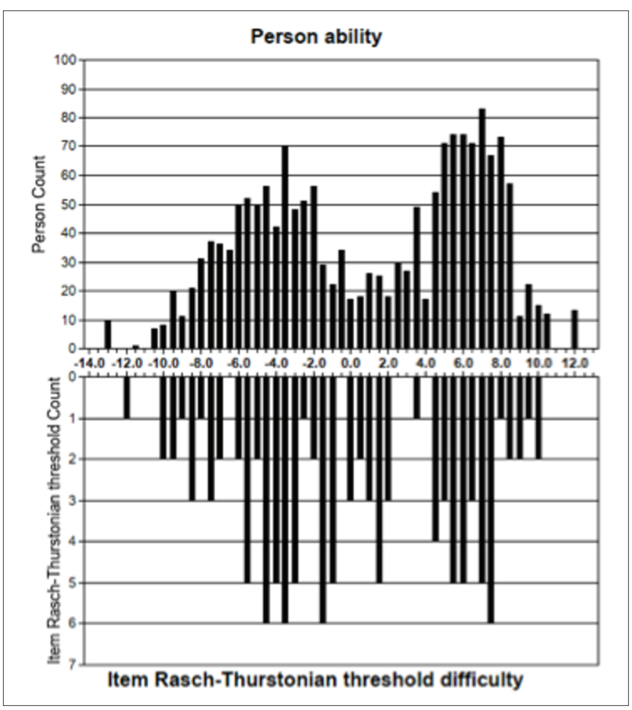
Expressive Communication



Fine Motor



Gross Motor



Transformation of Rasch Ability Scores to GSVs

GSVs are a linear transformation of the logit values of Rasch ability scores: $GSV_i = a \times ability_i + b$. Linear transformation preserves the shape of the ability-score distribution. The coefficient *a* and constant *b* for each Bayley-4 subtest, shown in Table 1, were chosen so that GSVs would have a standard deviation of 25 and mean of 500 in the full standardization sample.

Table 1. Parameters for the Linear Transformation of Ability Scores from Logits to GSVs

Subtest	Coefficient (*a*)	Constant (*b*)
Cognitive	5.401	505.1
Receptive Communication	5.300	505.5
Expressive Communication	4.190	507.2
Fine Motor	5.624	504.1
Gross Motor	4.267	496.0

The Appendix presents the GSV corresponding to each raw score on each subtest. It should be noted that these GSVs are slightly different from those originally provided for Bayley-4, prior to the 2023 update of the norm sample.

The charts in Figure 2 show the shapes of the relationships of raw scores to GSVs on each subtest. Each chart also includes the frequency distribution of item threshold difficulties in GSV units to illustrate how the density of item difficulties affects the slope of the line: the more item thresholds there are in a region of difficulty, the faster the raw score increases as ability increases, because there are more item thresholds on which the person can demonstrate their ability increase.

Figure 2. Raw Score Versus GSV and Frequency Distribution of Item Threshold Difficulties, by Subtest

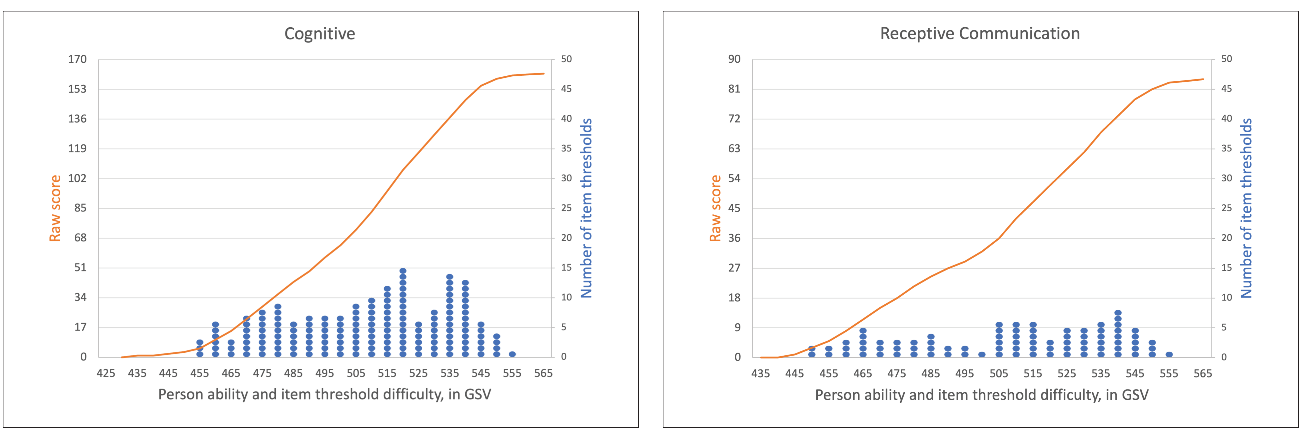
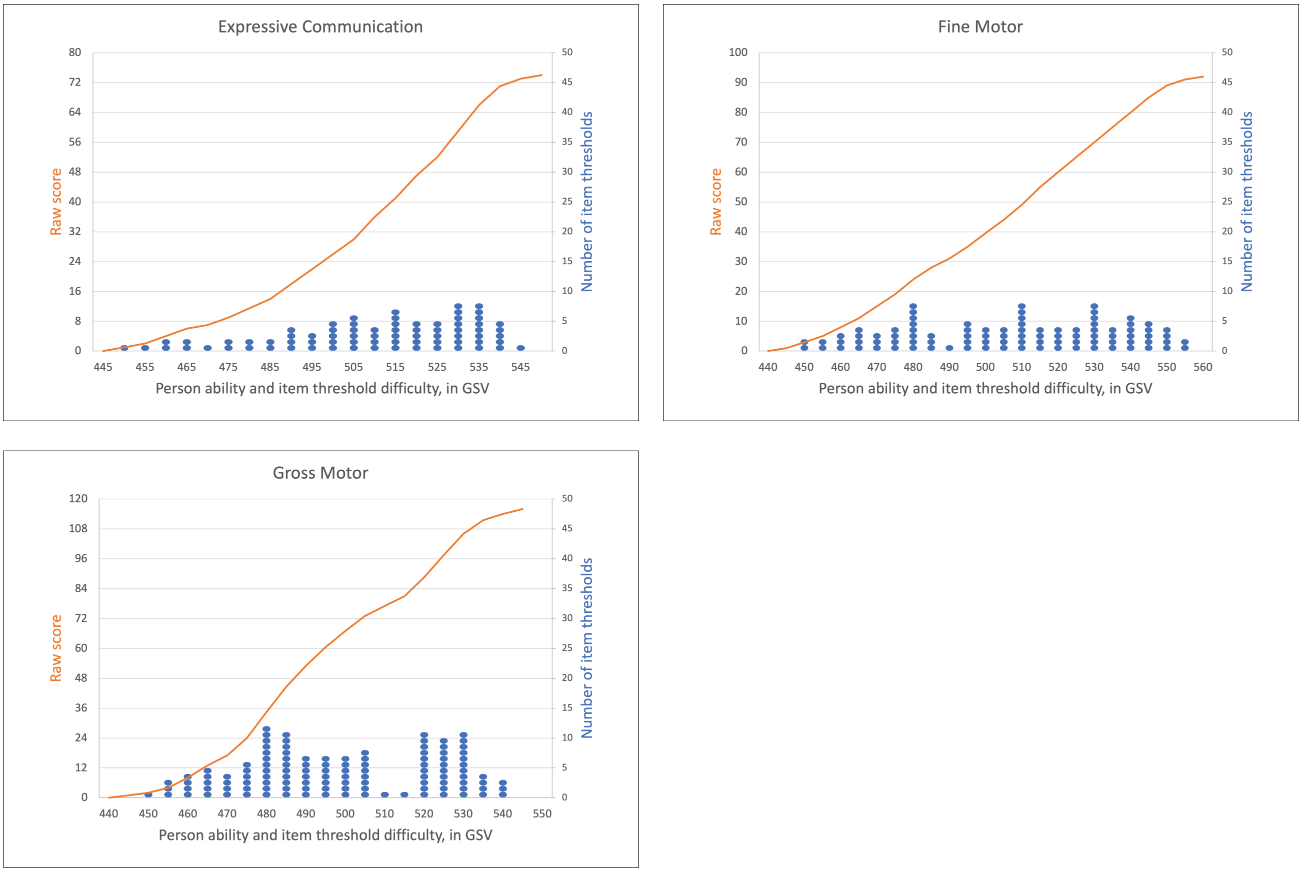


Figure 2. Raw Score Versus GSV and Frequency Distribution of Item Threshold Difficulties, by Subtest (*continued*)



Change in Item Success Probability as GSV Changes

Knowing how a change in GSV affects children's probability of mastering a particular test item can be useful for attaching meaning to GSV changes, such as for estimating a minimal clinically important difference (MCID). The relationship between GSV change and change in success probability differs across Bayley-4 subtests because they use different coefficients to transform Rasch ability scores to GSVs. Table 2 shows, for each subtest, the probability of success following various amounts of GSV change, starting from various initial probability values. For example, if a child initially had a .30 probability of scoring 2 on a particular Cognitive item, and their GSV increased by 10 points, their new probability of scoring 2 would be .73.

Table 2. Probability of Success After Change in GSV, by Subtest and Initial Probability

Cognitive

	Change in GSV						
Initial *p*	–15	–10	–5	0	5	10	15
0.95	0.54	0.75	0.88	0.95	0.98	0.99	1.00
0.90	0.36	0.59	0.78	0.90	0.96	0.98	0.99
0.80	0.20	0.39	0.61	0.80	0.91	0.96	0.98
0.70	0.13	0.27	0.48	0.70	0.85	0.94	0.97
0.60	0.09	0.19	0.37	0.60	0.79	0.91	0.96
0.50	0.06	0.14	0.28	0.50	0.72	0.86	0.94
0.40	0.04	0.09	0.21	0.40	0.63	0.81	0.91
0.30	0.03	0.06	0.15	0.30	0.52	0.73	0.87
0.20	0.02	0.04	0.09	0.20	0.39	0.61	0.80
0.10	0.01	0.02	0.04	0.10	0.22	0.41	0.64
0.05	0.00	0.01	0.02	0.05	0.12	0.25	0.46

Receptive Communication

	Change in GSV						
Initial *p*	–15	–10	–5	0	5	10	15
0.95	0.53	0.74	0.88	0.95	0.98	0.99	1.00
0.90	0.35	0.58	0.78	0.90	0.96	0.98	0.99
0.80	0.19	0.38	0.61	0.80	0.91	0.96	0.99
0.70	0.12	0.26	0.48	0.70	0.86	0.94	0.98
0.60	0.08	0.19	0.37	0.60	0.79	0.91	0.96
0.50	0.06	0.13	0.28	0.50	0.72	0.87	0.94
0.40	0.04	0.09	0.21	0.40	0.63	0.81	0.92
0.30	0.02	0.06	0.14	0.30	0.52	0.74	0.88
0.20	0.01	0.04	0.09	0.20	0.39	0.62	0.81
0.10	0.01	0.02	0.04	0.10	0.22	0.42	0.65
0.05	0.00	0.01	0.02	0.05	0.12	0.26	0.47

Expressive Communication

	Change in GSV						
Initial *p*	–15	–10	–5	0	5	10	15
0.95	0.35	0.64	0.85	0.95	0.98	1.00	1.00
0.90	0.20	0.45	0.73	0.90	0.97	0.99	1.00
0.80	0.10	0.27	0.55	0.80	0.93	0.98	0.99
0.70	0.06	0.18	0.41	0.70	0.89	0.96	0.99
0.60	0.04	0.12	0.31	0.60	0.83	0.94	0.98
0.50	0.03	0.08	0.23	0.50	0.77	0.92	0.97
0.40	0.02	0.06	0.17	0.40	0.69	0.88	0.96
0.30	0.01	0.04	0.11	0.30	0.59	0.82	0.94
0.20	0.01	0.02	0.07	0.20	0.45	0.73	0.90
0.10	0.00	0.01	0.03	0.10	0.27	0.55	0.80
0.05	0.00	0.00	0.02	0.05	0.15	0.36	0.65

Fine Motor

	Change in GSV						
Initial *p*	–15	–10	–5	0	5	10	15
0.95	0.57	0.76	0.89	0.95	0.98	0.99	1.00
0.90	0.38	0.60	0.79	0.90	0.96	0.98	0.99
0.80	0.22	0.40	0.62	0.80	0.91	0.96	0.98
0.70	0.14	0.28	0.49	0.70	0.85	0.93	0.97
0.60	0.09	0.20	0.38	0.60	0.78	0.90	0.96
0.50	0.06	0.14	0.29	0.50	0.71	0.86	0.94
0.40	0.04	0.10	0.22	0.40	0.62	0.80	0.91
0.30	0.03	0.07	0.15	0.30	0.51	0.72	0.86
0.20	0.02	0.04	0.09	0.20	0.38	0.60	0.78
0.10	0.01	0.02	0.04	0.10	0.21	0.40	0.62
0.05	0.00	0.01	0.02	0.05	0.11	0.24	0.43

Gross Motor

	Change in GSV						
Initial *p*	–15	–10	–5	0	5	10	15
0.95	0.36	0.65	0.85	0.95	0.98	0.99	1.00
0.90	0.21	0.46	0.74	0.90	0.97	0.99	1.00
0.80	0.11	0.28	0.55	0.80	0.93	0.98	0.99
0.70	0.06	0.18	0.42	0.70	0.88	0.96	0.99
0.60	0.04	0.13	0.32	0.60	0.83	0.94	0.98
0.50	0.03	0.09	0.24	0.50	0.76	0.91	0.97
0.40	0.02	0.06	0.17	0.40	0.68	0.87	0.96
0.30	0.01	0.04	0.12	0.30	0.58	0.82	0.94
0.20	0.01	0.02	0.07	0.20	0.45	0.72	0.89
0.10	0.00	0.01	0.03	0.10	0.26	0.54	0.79
0.05	0.00	0.01	0.02	0.05	0.15	0.35	0.64

Conditional Standard Error of Measurement of GSV

The standard error of measurement (*SEM*) provides an estimate of the amount of error in a child's observed test score. The Bayley-4 Technical Manual describes how to evaluate GSV differences and determine whether the absolute value of the GSV difference is statistically significant based on traditional *SEM.*

The conditional *SEM* has the same meaning as the traditional *SEM*, but it is specific to each GSV value. Conditional *SEM* is a function of the number of moderately difficult items (or item thresholds) a child at that ability level will encounter on the subtest. The score from each such encounter provides information about the child's ability, and the more such encounters, the better the estimate of ability and the smaller the conditional *SEM*. This may be seen in Figure 3, showing both the conditional *SEM* and the number of item thresholds at each level of ability/difficulty. Conditional *SEM* rises and falls in accordance with the number of item thresholds at that GSV level. On every subtest the conditional *SEM is* large for very low or very high GSVs, where the number of item thresholds is small. For children with such extreme levels of ability, the test is too difficult or too easy, respectively, to provide highly precise measurement. The Appendix reports the conditional *SEM* of each GSV for each subtest.

Figure 3. Conditional *SEM* Versus GSV, and Frequency Distribution of Item Difficulties, by Subtest

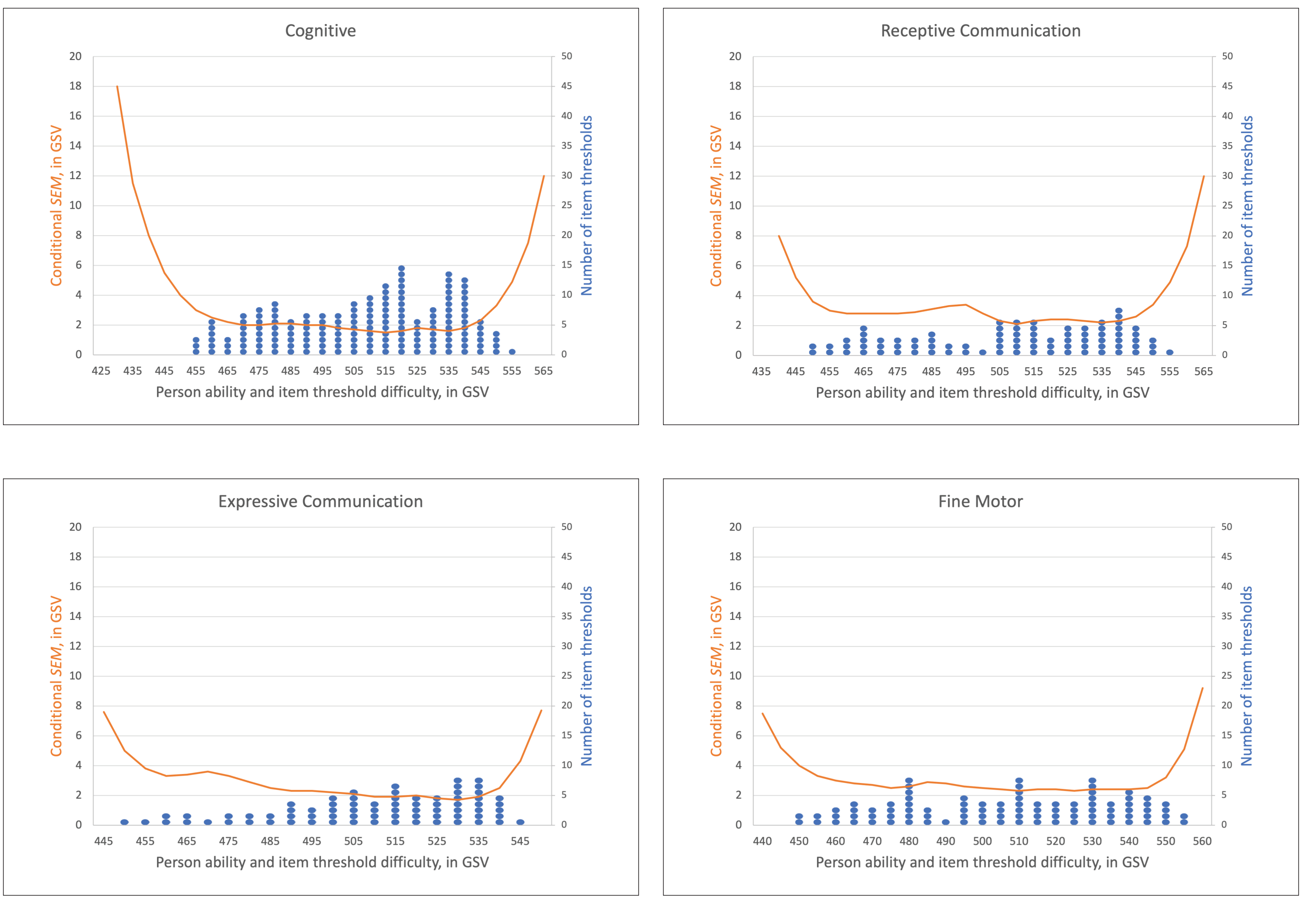
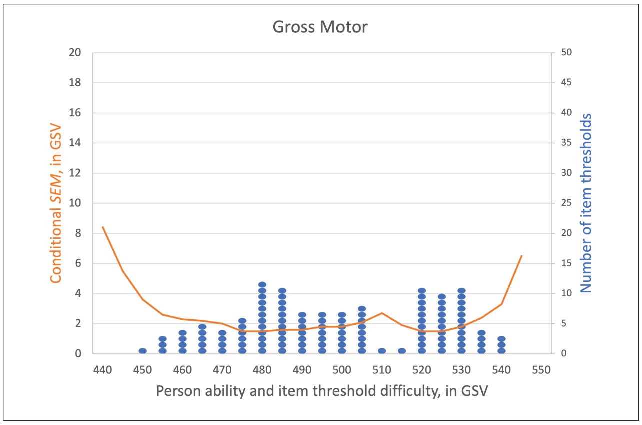


Figure 3. Conditional *SEM* Versus GSV, and Frequency Distribution of Item Difficulties, by Subtest (*continued*)



A common application of conditional *SEM* is to determine whether the difference between two GSVs on the same test (such as a child's GSVs at two points in time) is statistically significant. This is done by dividing the difference by the standard error of the difference:

$$\frac{GSV_1 - GSV_2}{\sqrt{cSEM_1{}^2 + cSEM_2{}^2}}$$

Values of 1.65 and 1.96 indicate significant differences at $p<.10$ and $p<.05$, respectively.

Diagnostic Information About Rasch Calibration

The Rasch model makes several assumptions about how test data behave: score categories should be well-ordered, item characteristic curves should have similar slope, the test should measure a single dimension of ability, and items should be *locally independent* (i.e., pairs of items should correlate only because they both measure the single ability dimension). This section presents evidence regarding each of these assumptions.

Score category ordering

Each Bayley-4 item has three score categories: 0, 1, and 2. An item's categories are said to be *well-ordered* if people who obtain a score of 1 have greater average ability than those who score 0, and people scoring 2 have greater ability than those scoring 1. Figure 4 shows the observed ordering of these averages on each item, using data from the calibration sample. In each chart, the items are in descending order of difficulty. All items on all subtests have well-ordered categories, which is to be expected given that the categories represent failure, partial success, and complete success.

Figure 4. Average Observed Person Ability by Item Score: Cognitive

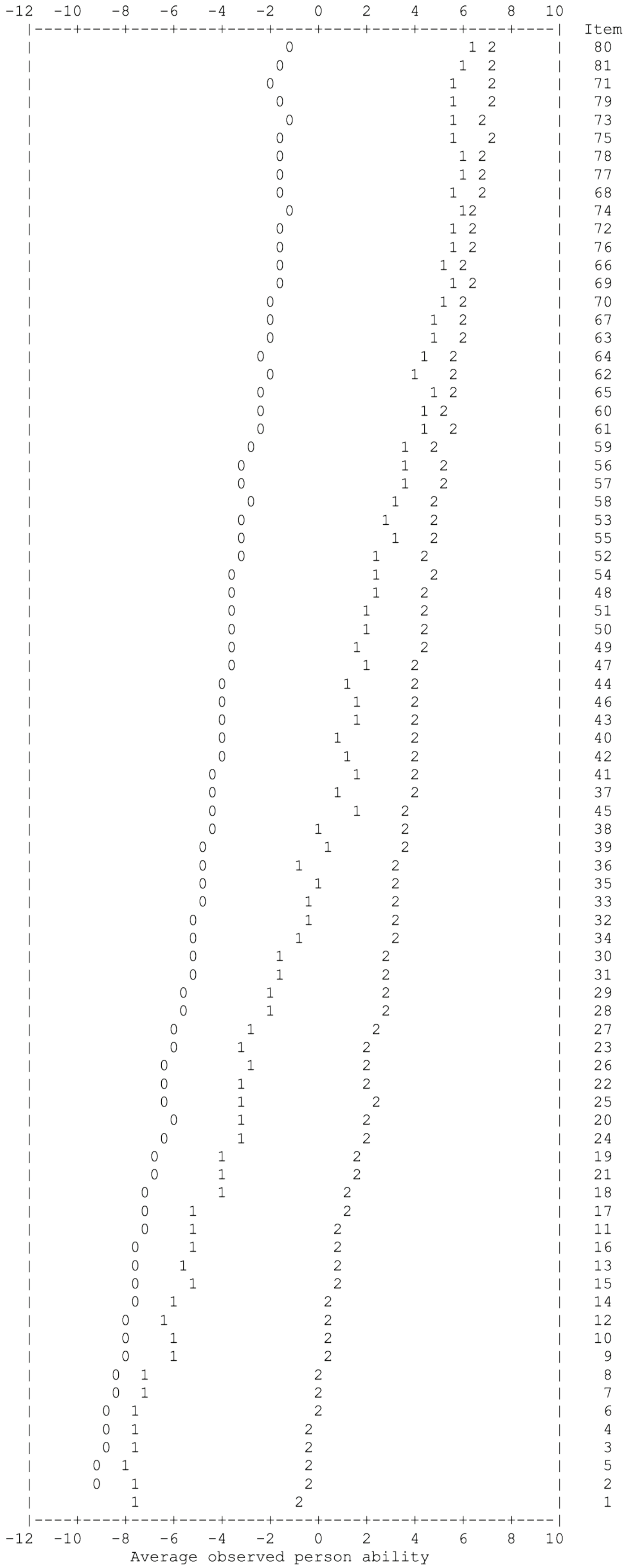


Figure 4 (*continued*). Average Observed Person Ability by Item Score: Receptive Communication

```
-12  -10   -8   -6   -4   -2    0    2    4    6    8   10
  |----+----+----+----+----+----+----+----+----+----+----|  Item
  |                         0                   1  2     |   40
  |                         0                    1 2     |   42
  |                        0                    1  2     |   36
  |                        0                    1 2      |   41
  |                        0                    1  2     |   39
  |                        0                    12       |   38
  |                       0                    1  2      |   35
  |                       0                   1  2       |   37
  |                       0                   1 2        |   34
  |                      0                   1   2       |   31
  |                      0                   1  2        |   32
  |                      0                   1  2        |   33
  |                       0                 1   2        |   30
  |                     0                  1    2        |   27
  |                     0                1    2          |   28
  |                     0               1     2          |   29
  |                     0               1     2          |   26
  |                    0            1        2           |   25
  |                   0             1       2            |   21
  |                   0             1       2            |   23
  |                   0           1        2             |   24
  |                   0            1       2             |   22
  |                   0          1         2             |   19
  |                  0        1           2              |   20
  |                  0        1           2              |   18
  |                 0          1          2              |   16
  |                 0         1          2               |   17
  |               0       1            2                 |   14
  |               0     1             2                  |   13
  |              0      1            2                   |   15
  |             0     1            2                     |   11
  |            0     1             2                     |   12
  |          0    1              2                       |    9
  |           0  1               2                       |   10
  |         0  1                2                        |    7
  |         0  1                2                        |    8
  |        0  1                2                         |    6
  |      0   1                 2                         |    4
  |       0   1                2                         |    2
  |      0  1                  2                         |    3
  |    0      1                2                         |    5
  |    0    1                 2                          |    1
  |----+----+----+----+----+----+----+----+----+----+----|
-12  -10   -8   -6   -4   -2    0    2    4    6    8   10
            Average observed person ability
```

Figure 4 (*continued*). Average Observed Person Ability by Item Score: Expressive Communication

```
-16 -14 -12 -10  -8  -6  -4  -2   2   4   6   8  10  12
  |---+---+---+---+---+---+---+---+---+---+---+---+---|  Item
  |                       0                   1  2    |   36
  |                       0                   1  2    |   33
  |                       0                   1 2     |   35
  |                       0                   1 2     |   32
  |                      0                  1  2      |   37
  |                      0                  1  2      |   34
  |                      0                  1  2      |   29
  |                      0                  1 2       |   31
  |                     0                1    2       |   28
  |                     0                1    2       |   26
  |                     0               1     2       |   27
  |                    0                 1    2       |   30
  |                    0               1     2        |   25
  |                   0                1     2        |   24
  |                  0               1      2         |   23
  |                  0              1       2         |   22
  |                  0             1       2          |   20
  |                  0              1      2          |   21
  |                  0            1        2          |   18
  |                 0             1       2           |   19
  |                 0           1         2           |   16
  |                0             1        2           |   17
  |                0           1          2           |   15
  |               0          1           2            |   13
  |               0         1           2             |   11
  |               0          1          2             |   14
  |               0        1           2              |   12
  |              0     1             2                |    9
  |             0      1             2                |    8
  |             0       1            2                |   10
  |            0     1              2                 |    7
  |           0    1              2                   |    5
  |          0     1             2                    |    6
  |       0    1               2                      |    4
  |     0    1                2                       |    3
  |   0    1                  2                       |    2
  |    0    1                2                        |    1
  |---+---+---+---+---+---+---+---+---+---+---+---+---|
-16 -14 -12 -10  -8  -6  -4  -2   2   4   6   8  10  12
              Average observed person ability
```

Figure 4 (*continued*). Average Observed Person Ability by Item Score: Fine Motor

```
-12  -10   -8   -6   -4   -2    0    2    4    6    8   10
  |----+----+----+----+----+----+----+----+----+----+----|   Item
  |                           0                  1  2    |    46
  |                           0                 1   2    |    42
  |                           0               1  2       |    45
  |                          0                  1 2      |    43
  |                          0                 1  2      |    41
  |                          0                 1 2       |    44
  |                          0                1 2        |    40
  |                          0                1  2       |    39
  |                         0               1  2         |    37
  |                         0                1 2         |    38
  |                        0              1   2          |    35
  |                        0               1  2          |    34
  |                        0              1   2          |    36
  |                        0               1 2           |    33
  |                       0              1   2           |    31
  |                        0           1   2             |    29
  |                      0             1   2             |    32
  |                      0          1     2              |    26
  |                     0             1    2             |    30
  |                    0            1     2              |    25
  |                   0              1    2              |    28
  |                    0           1     2               |    27
  |                   0            1     2               |    24
  |                   0          1       2               |    23
  |                  0         1        2                |    22
  |                 0          1        2                |    20
  |                 0          1       2                 |    19
  |                0         1        2                  |    21
  |               0          1         2                 |    18
  |               0         1         2                  |    17
  |               0        1          2                  |    16
  |             0          1         2                   |    15
  |            0       1            2                    |    14
  |           0     1              2                     |    12
  |           0         1          2                     |    13
  |           0    1               2                     |    10
  |          0      1              2                     |    11
  |           0   1                2                     |     9
  |         0    1                 2                     |     8
  |         0   1                 2                      |     7
  |        0    1                2                       |     6
  |        0   1                 2                       |     5
  |       0   1                 2                        |     3
  |      0   1                  2                        |     2
  |      0  1                  2                         |     4
  |     0  1                   2                         |     1
  |----+----+----+----+----+----+----+----+----+----+----|
-12  -10   -8   -6   -4   -2    0    2    4    6    8   10
            Average observed person ability
```

Figure 4 (*continued*). Average Observed Person Ability by Item Score: Gross Motor

```
-14 -12 -10  -8  -6  -4  -2   2   4   6   8  10  12  14
  |---+---+---+---+---+---+---+---+---+---+---+---+---+---|  Item
  |                          0               1  2     |   57
  |                          0               1  2     |   55
  |                         0               1  2      |   56
  |                        0                1 2       |   58
  |                        0              1  2        |   54
  |                        0               1 2        |   49
  |                       0               1  2        |   51
  |                        0              1 2         |   52
  |                       0               1  2        |   53
  |                       0              1  2         |   50
  |                      0               1  2         |   47
  |                     0                1  2         |   48
  |                     0               1   2         |   44
  |                     0               1  2          |   46
  |                     0               1  2          |   43
  |                    0                1  2          |   41
  |                   0                1   2          |   39
  |                    0             1    2           |   40
  |                   0                1   2          |   45
  |                  0                1   2           |   42
  |                0             1       2            |   38
  |                0            1        2            |   35
  |                0          1         2             |   36
  |                0             1      2             |   37
  |                0         1          2             |   34
  |                0          1         2             |   33
  |               0          1          2             |   31
  |               0          1          2             |   32
  |               0       1            2              |   30
  |               0      1             2              |   28
  |              0      1              2              |   25
  |              0       1             2              |   29
  |              0      1             2               |   27
  |              0      1             2               |   26
  |             0      1              2               |   20
  |             0     1              2                |   24
  |             0    1               2                |   19
  |             0    1              2                 |   21
  |            0     1              2                 |   18
  |            0      1             2                 |   22
  |            0     1              2                 |   23
  |            0   1                2                 |   14
  |           0     1              2                  |   16
  |           0    1               2                  |   17
  |           0   1               2                   |   13
  |           0   1                2                  |   10
  |          0    1                2                  |   11
  |           0   1               2                   |   12
  |          0   1                2                   |    9
  |          0    1               2                   |   15
  |        0   1                 2                    |    8
  |        0   1                2                     |    7
  |      0   1                  2                     |    4
  |      0    1                2                      |    6
  |      0   1                  2                     |    5
  |     0   1                  2                      |    3
  |    0  1                   2                       |    2
  | 0       1                 2                       |    1
  |---+---+---+---+---+---+---+---+---+---+---+---+---+---|
-14 -12 -10  -8  -6  -4  -2   2   4   6   8  10  12  14
            Average observed person ability
```

Item Fit

An *item characteristic curve* (ICC) describes how the expected score on an item increases as ability increases. (For simplicity, this discussion refers to dichotomous items, but the same principles apply to polytomous items.) The Rasch model assumes that every item's ICC is a logistic curve and that all ICCs on a test have the same slope. This assumption can be tested by computing the actual success rate of persons at each ability level and comparing the resulting empirical trend with the expected ICC. To the extent that these differ, the item is said to *misfit*. If the slope of the actual data is steeper than expected (*overfit*), the item does a better job of differentiating between people at higher and lower ability levels—that is, it correlates higher with the overall ability dimension. Conversely, a flatter ICC (*underfit*) means that the item is below-average in its ability to identify different levels of ability. Overfit does not compromise the reliability or validity of scores but can cause underestimation of conditional *SEM*s and affect the accuracy of estimates of success probability (Bond et al., 2021). Underfit, on the other hand, may affect the quality of measurement. It should be noted that underfit also affects raw scores and other scores derived from raw scores, such as standard scores and age equivalents.

Misfit can be measured by several statistics, one of which is *infit mean square*. On any test, the average value of infit mean square is expected to be 1.00; larger values indicate ICCs that are flatter than expected, and smaller values indicate ICCs that are relatively steep. There is no generally accepted standard for the desirable range of infit mean square, but 0.50 to 1.50 (Linacre, 2021b) and 0.75 to 1.30 (Bond et al., 2021) are typical recommendations.

Table 3 reports, for each subtest, the mean and standard deviation of infit mean square and the number of items at different levels of infit mean square. Overall, 97% of items had infit mean square between 0.50 and 1.50, and 78% had infit mean square between 0.76 and 1.30. Most of the values that fell outside the latter range were small (i.e., the items overfit). The three Fine Motor items with infit mean square values greater than 1.50 were all measures of pencil grasp, a distinct subskill that nevertheless is part of the fine motor construct.

Table 3. Descriptive Statistics for Infit Mean Square, by Subtest

		Infit mean square		Number of items with infit mean square in range:				
Subtest	No. items	Mean	*SD*	<0.50	0.50–0.75	.76–1.30	1.31–1.50	>1.50
Cognitive	81	1.00	0.21	0	7	69	3	2
Receptive Communication	42	0.97	0.18	0	4	36	1	1
Expressive Communication	37	0.96	0.20	0	5	30	2	0
Fine Motor	46	0.99	0.29	0	9	30	4	3
Gross Motor	58	0.96	0.25	2	9	41	5	1

Local Independence and Item Intercorrelations

The local independence and dimensionality assumptions are both evaluated, in whole or in part, through the intercorrelations of item residuals (the differences between persons' expected and observed item scores). If a person's item performance is solely a function of their level on the underlying ability, then residuals on different items should be uncorrelated. A substantial correlation might reflect the fact that the two items measure a secondary ability in addition to the primary ability; depending on the strength and nature of this secondary ability, the test might be considered multidimensional. Alternatively, there could be a relationship between the content, administration, or scoring of the items. For example, performance on one item might constrain the score on another item or determine whether it is administered; the content of one item might give information useful to solving the other; or both items might require interpreting the same stimulus such as a chart or a reading passage. The term *local dependence* will be used to refer to these latter types of relationships, regardless of whether the item residuals are highly correlated. The effects of local dependence on test usage are generally benign (Bond et al., 2021; Linacre, 2021b). Locally dependent items will tend to overfit, which may cause underestimation of the conditional *SEM* at some score values and may affect the accuracy of estimates of success probability on individual items.

It is desirable to evaluate local dependence first so that any high residual correlations it causes are not misinterpreted as evidence of multidimensionality. For Bayley-4, this was done by examining the content, administration, and scoring procedures of all items, with special attention to pairs of items with highly correlated item residuals. The following locally dependent item sets were found. Except as noted, they consisted of different levels of performance on the same task.

- Cognitive: Five pairs and two triads. In two sets, performance on the first item determined whether the second was administered, and in one, the tasks were partly the same. Correlations between raw residuals on dependent items ranged from .15 to .58 (median = .39).
- Receptive Communication: Three pairs. Residual correlations: .35 to .50 (median = .41).
- Expressive Communication: Three pairs and one triad. Residual correlations: .08 to .58 (median = .25).
- Fine Motor: Five pairs, one triad, and one set of four. Three sets consisted of highly similar tasks. Residual correlations: .00 to .66 (median = .15).
- Gross Motor: Ten pairs. Residual correlations: .09 to .71 (median = .25).

The effect of these locally dependent items on the subtests' measurement properties was evaluated by consolidating each pair or set of locally dependent items into a single item (by summing their scores), performing a Rasch calibration of the reduced item set, and comparing the resulting ability scores and conditional *SEM*s with the original values for the same raw scores. Table 4 reports these correlations in the Bayley-4 norm sample. The relationships were nearly perfect. The standard deviations of ability scores were 4% to 6% larger when locally dependent items were separate, reflecting the artificially high correlations between them. Average conditional *SEM*s were 2% to 3% larger, meaning that the presence of locally dependent items did not cause underestimation of conditional *SEM*s. Overall, the data indicate that the presence of locally dependent items had a negligible effect on the measurement properties of GSVs.

Table 4. Effects of Combining Locally Dependent Items on Ability Scores and Conditional *SEM*s

	Correlation		*SD* of ability scores		Mean conditional *SEM*	
Subtest	Ability scores	Conditional *SEM*	Separate	Summed	Separate	Summed
Cognitive	0.9999	0.996	4.63	4.45	0.35	0.34
Receptive Communication	0.9998	0.994	4.72	4.53	0.52	0.51
Expressive Communication	0.9997	0.998	5.97	5.67	0.62	0.61
Fine Motor	0.9999	0.993	4.45	4.23	0.46	0.45
Gross Motor	0.9999	0.999	5.86	5.52	0.45	0.43

In the reduced item sets containing sums of locally dependent items, only two pairs of items had residual correlations exceeding .40 in absolute value: two Cognitive items that involve finding a hidden object (.41), and two Gross Motor items related to climbing up or down stairs (.51). Each of these pairs included tasks that were highly similar but less so than the items identified as locally dependent.

Dimensionality

Dimensionality can be evaluated by using a principal components analysis of the item residuals to see if there are any components large enough to constitute secondary ability dimensions. Linacre (2021b) recommends that components with eigenvalues greater than 2.0 are worthy of investigation because they have the "strength" of two or more items. Three features of such components should be examined. One is the percentage of variance accounted for by the component. A second is its impact on the overall score, assessed by scoring each person on three subsets of items: items with high positive loadings on the component, those with high negative loadings, and those in between. If these three subset scores intercorrelate highly, the component has little effect on subtest scores. The third feature to examine is the content of the items with large loadings on the component, to infer what construct the component represents and whether it is outside the conceptual domain of the test. As Smith (2004) notes, "multidimensionality only becomes a problem when data represent two or more dimensions so disparate or distinct that it is no longer clear what dimension the Rasch model is defining (lacks construct validity) or when the different subsets of items would lead to different norm (NR) or criterion-referenced (CR) decisions."

A principal components analysis was performed on the raw item residuals in the reduced item set for each subtest. A total of three components with eigenvalues greater than 2 were found: one on Cognitive (eigenvalue 2.2), one on Fine Motor (2.0), and one on Gross Motor (2.1). Each accounted for 0.5% or less of score variance on the subtest, and all disattenuated correlations between item subsets were 1.00. The Fine Motor component reflected visual versus manual activity; the Gross Motor component reflected walking up/down stairs versus walking or kicking a ball; and the Cognitive component was not interpretable. These findings indicate that each subtest met the unidimensionality assumption once the effects of procedurally induced local dependence were controlled.

References

Bayley, N., & Aylward, G. P. (2019). *Bayley Scales of Infant and Toddler Development* (4th ed.): *Administration manual*. NCS Pearson.

Bayley, N., & Aylward, G. P. (2023). *Bayley Scales of Infant and Toddler Development* (4th ed.): *Technical manual*. NCS Pearson.

Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge.

Linacre, J. M. (2021a). *Winsteps Rasch measurement computer program*. (Version 5.1.4) [Computer software]. Winsteps.com. Available from https://www.winsteps.com.

Linacre, J. M. (2021b). *Winsteps Rasch measurement computer program user's guide*. (Version 5.1.4). Winsteps.com. Available from https://www.winsteps.com.

Smith, E. V. (2004). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement: Theory, models and applications* (pp. 575–600). JAM Press.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. MESA Press.

Appendix: Bayley-4 GSVs and Conditional GSV *SEM*s Corresponding to Raw Scores

Raw score	Cognitive GSV	Cognitive c*SEM*	Receptive GSV	Receptive c*SEM*	Expressive GSV	Expressive c*SEM*	Fine Motor GSV	Fine Motor c*SEM*	Gross Motor GSV	Gross Motor c*SEM*	Raw score
0	427		438	9.8	444	8.1	437	10.2	441	8.1	0
1	438	10.0	444	5.5	450	4.9	444	5.7	447	4.8	1
2	445	5.6	448	4.1	455	3.9	448	4.3	451	3.5	2
3	449	4.0	451	3.6	458	3.5	451	3.8	453	2.9	3
4	452	3.4	453	3.2	460	3.3	453	3.5	455	2.6	4
5	454	3.0	455	3.1	463	3.3	455	3.3	456	2.5	5
6	455	2.8	457	3.0	466	3.4	457	3.1	458	2.4	6
7	457	2.6	458	2.9	469	3.6	459	3.0	459	2.3	7
8	458	2.5	460	2.8	472	3.6	461	3.0	460	2.2	8
9	459	2.4	461	2.8	475	3.3	462	2.9	461	2.2	9
10	460	2.4	463	2.8	477	3.1	464	2.9	462	2.1	10
11	461	2.4	464	2.8	479	2.9	465	2.8	463	2.1	11
12	462	2.3	466	2.8	481	2.8	466	2.8	464	2.1	12
13	463	2.3	467	2.8	483	2.7	468	2.8	466	2.1	13
14	464	2.2	469	2.8	485	2.6	469	2.7	467	2.1	14
15	465	2.2	470	2.8	486	2.5	470	2.6	468	2.1	15
16	466	2.1	472	2.8	488	2.4	472	2.6	469	2.0	16
17	467	2.1	473	2.8	489	2.3	473	2.5	470	2.0	17
18	467	2.1	475	2.8	490	2.3	474	2.5	470	1.9	18
19	468	2.0	476	2.8	491	2.3	475	2.4	471	1.8	19
20	469	2.0	478	2.9	493	2.3	476	2.4	472	1.7	20
21	470	1.9	479	2.9	494	2.3	477	2.4	473	1.7	21
22	470	1.9	481	3.0	495	2.3	478	2.4	473	1.6	22
23	471	1.9	483	3.0	497	2.2	479	2.5	474	1.6	23
24	472	1.9	484	3.1	498	2.2	480	2.5	475	1.5	24
25	472	1.9	486	3.1	499	2.2	481	2.6	475	1.5	25
26	473	1.9	488	3.2	500	2.2	482	2.7	476	1.5	26
27	474	1.9	490	3.3	501	2.2	484	2.8	476	1.5	27
28	474	1.9	492	3.3	502	2.2	485	2.9	477	1.5	28
29	475	1.9	494	3.4	503	2.2	487	2.9	477	1.5	29
30	476	1.9	497	3.3	505	2.1	488	2.9	478	1.5	30
31	476	1.9	498	3.0	506	2.1	490	2.8	478	1.5	31
32	477	1.9	500	2.8	507	2.1	491	2.8	479	1.5	32
33	478	2.0	501	2.6	508	2.0	492	2.7	479	1.4	33
34	478	2.0	503	2.5	509	2.0	493	2.6	480	1.4	34
35	479	2.0	504	2.4	509	2.0	495	2.6	480	1.5	35
36	480	2.0	505	2.3	510	1.9	496	2.6	481	1.5	36
37	481	2.0	506	2.2	511	1.9	497	2.6	481	1.5	37
38	481	2.1	507	2.2	512	1.9	498	2.6	481	1.5	38
39	482	2.0	508	2.2	513	1.9	499	2.5	482	1.5	39
40	483	2.0	508	2.2	514	1.9	500	2.5	483	1.5	40
41	484	2.0	509	2.1	515	1.9	502	2.5	483	1.5	41
42	484	2.0	510	2.1	516	1.9	503	2.4	484	1.5	42
43	485	2.0	511	2.2	516	1.9	504	2.4	484	1.5	43
44	486	2.0	512	2.2	517	2.0	505	2.4	485	1.6	44
45	487	1.9	513	2.2	518	2.0	506	2.3	485	1.6	45

Appendix: Bayley-4 GSVs and Conditional GSV *SEM*s Corresponding to Raw Scores (*continued*)

Raw score	Cognitive		Receptive		Expressive		Fine Motor		Gross Motor		Raw score
	GSV	c*SEM*	GSV	c*SEM*	GSV	c*SEM*	GSV	c*SEM*	GSV	c*SEM*	
46	487	1.9	514	2.2	519	2.0	507	2.3	486	1.6	46
47	488	1.9	515	2.3	520	2.0	508	2.2	486	1.6	47
48	489	1.9	516	2.3	521	2.0	508	2.2	487	1.6	48
49	490	1.9	517	2.3	522	2.0	509	2.3	488	1.6	49
50	490	1.9	518	2.3	523	1.9	510	2.3	488	1.6	50
51	491	1.9	519	2.4	524	1.9	511	2.3	489	1.6	51
52	492	1.9	520	2.4	525	1.8	512	2.4	489	1.6	52
53	492	1.9	521	2.4	526	1.8	513	2.4	490	1.6	53
54	493	1.9	522	2.3	526	1.8	514	2.4	491	1.6	54
55	494	1.9	523	2.3	527	1.8	515	2.4	491	1.6	55
56	494	1.9	524	2.3	528	1.7	516	2.4	492	1.6	56
57	495	1.9	525	2.3	529	1.7	517	2.4	493	1.7	57
58	496	1.9	526	2.3	529	1.7	518	2.4	493	1.7	58
59	497	1.9	527	2.3	530	1.7	519	2.4	494	1.7	59
60	497	1.9	528	2.3	531	1.7	520	2.4	495	1.8	60
61	498	1.9	529	2.2	531	1.7	521	2.4	495	1.8	61
62	498	1.8	530	2.2	532	1.7	522	2.3	496	1.8	62
63	499	1.8	531	2.2	533	1.7	523	2.3	497	1.8	63
64	500	1.8	532	2.2	533	1.8	524	2.2	498	1.8	64
65	500	1.8	533	2.2	534	1.8	525	2.2	499	1.8	65
66	501	1.8	533	2.1	535	1.8	526	2.2	499	1.8	66
67	502	1.8	534	2.1	536	1.9	527	2.2	500	1.8	67
68	502	1.8	535	2.1	537	2.0	528	2.2	501	1.8	68
69	503	1.7	536	2.1	538	2.1	529	2.3	502	1.8	69
70	503	1.7	537	2.2	539	2.3	530	2.3	502	1.8	70
71	504	1.7	538	2.2	540	2.5	531	2.4	503	1.8	71
72	504	1.7	539	2.2	542	3.0	532	2.4	504	1.9	72
73	505	1.7	540	2.2	545	4.1	533	2.4	505	2.0	73
74	505	1.7	541	2.3	550	7.6	534	2.4	506	2.2	74
75	506	1.7	542	2.3			535	2.4	507	2.4	75
76	506	1.7	543	2.4			536	2.4	509	2.6	76
77	507	1.6	544	2.5			537	2.4	510	2.6	77
78	507	1.6	545	2.6			538	2.4	512	2.5	78
79	508	1.6	546	2.8			539	2.4	513	2.2	79
80	508	1.6	548	3.0			540	2.4	514	2.0	80
81	509	1.6	550	3.4			541	2.3	515	1.9	81
82	509	1.6	552	4.0			542	2.4	516	1.8	82
83	510	1.6	556	5.5			543	2.4	517	1.7	83
84	510	1.6	563	9.8			544	2.4	517	1.7	84
85	511	1.5					545	2.5	518	1.6	85
86	511	1.5					546	2.5	519	1.6	86
87	511	1.5					547	2.6	519	1.6	87
88	512	1.5					549	2.8	520	1.6	88
89	512	1.5					550	3.1	520	1.5	89
90	513	1.5					552	3.7	521	1.5	90

Appendix: Bayley-4 GSVs and Conditional GSV *SEM*s Corresponding to Raw Scores (*continued*)

Raw score	Cognitive		Receptive		Expressive		Fine Motor		Gross Motor		Raw score
	GSV	c*SEM*	GSV	c*SEM*	GSV	c*SEM*	GSV	c*SEM*	GSV	c*SEM*	
91	513	1.5					556	5.1	521	1.5	91
92	514	1.5					561	9.7	522	1.5	92
93	514	1.5							523	1.5	93
94	514	1.5							523	1.5	94
95	515	1.5							524	1.5	95
96	515	1.5							524	1.5	96
97	516	1.5							525	1.5	97
98	516	1.5							525	1.5	98
99	516	1.5							526	1.5	99
100	517	1.5							526	1.5	100
101	517	1.5							527	1.5	101
102	518	1.5							527	1.6	102
103	518	1.5							528	1.6	103
104	518	1.5							529	1.6	104
105	519	1.5							529	1.7	105
106	519	1.5							530	1.7	106
107	520	1.6							531	1.8	107
108	520	1.6							531	1.9	108
109	521	1.6							532	2.0	109
110	521	1.6							533	2.1	110
111	522	1.6							534	2.3	111
112	522	1.6							536	2.4	112
113	523	1.6							537	2.6	113
114	523	1.7							539	3.0	114
115	524	1.7							542	4.1	115
116	524	1.7							546	7.4	116
117	525	1.7									117
118	525	1.7									118
119	526	1.7									119
120	526	1.7									120
121	527	1.7									121
122	527	1.7									122
123	528	1.7									123
124	528	1.7									124
125	529	1.7									125
126	529	1.7									126
127	530	1.7									127
128	530	1.6									128
129	531	1.6									129
130	531	1.6									130
131	532	1.6									131
132	532	1.6									132
133	533	1.6									133
134	533	1.6									134
135	534	1.6									135

Appendix: Bayley-4 GSVs and Conditional GSV *SEM*s Corresponding to Raw Scores (*continued*)

Raw score	Cognitive		Receptive		Expressive		Fine Motor		Gross Motor		Raw score
	GSV	c*SEM*	GSV	c*SEM*	GSV	c*SEM*	GSV	c*SEM*	GSV	c*SEM*	
136	534	1.6									136
137	535	1.6									137
138	535	1.6									138
139	536	1.6									139
140	536	1.6									140
141	537	1.6									141
142	537	1.6									142
143	538	1.6									143
144	538	1.6									144
145	539	1.6									145
146	539	1.7									146
147	540	1.7									147
148	540	1.7									148
149	541	1.7									149
150	541	1.8									150
151	542	1.8									151
152	543	1.9									152
153	543	2.0									153
154	544	2.1									154
155	545	2.2									155
156	546	2.4									156
157	547	2.6									157
158	548	2.9									158
159	550	3.3									159
160	552	4.0									160
161	557	5.6									161
162	563	10.0									162