**Exercise 4.1, page 86**

**Interpreting Test Scores: Marisol**

1. Interpret the standard scores on the Reading Total, Written Language Total, and Math Total.

* *Reading Total: Marisol’s standard score on the Reading Total was 88. This score falls slightly less than one standard deviation below the mean of the norm group.*
* *Written Language Total: The standard score on the Written Language Total was 91, which is slightly less than half of a standard deviation below the mean of the norm group.*
* *Math Total: The standard score was 126, which is close to 2 standard deviations above the mean.*

1. Interpret the percentile ranks on the Reading Total, Written Language Total, and Math Total.

* *Reading Total: The percentile rank of the Reading Total was 21, which indicates that Marisol scored the same or better than 21 percent of the norm group.*
* *Written Language Total: On the Written Language Total, Marisol scored in the 27th percentile, meaning her score equals or exceeds that of 27 percent of the norm group.*
* *Math Total: Marisol scored in the 97th percentile on the Math Total, which means that she scored the same or better than 97 percent of the norm group.*

1. Interpret the stanine scores on the Reading Total, Written Language Total, and Math Total.

* *None reported.*

1. Interpret the grade equivalent scores on the Vocabulary and Math Computation subtests.

* *Vocabulary: Marisol’s grade equivalent score on the Vocabulary subtest was less than the 3.0. This means that her performance falls below the average for third grade students (at the beginning of the school year).*
* *Math Computation: The grade equivalent score on the Math Computation subtest is 9.1. This indicates that Marisol correctly answered the same number of items that the average ninth grader (in the 1st month of the school year [September]) would answer on the same test.*

1. Interpret the age equivalent score on the Reading Comprehension and the Math Application subtests.

* *Reading Comprehension: On the Reading Comprehension subtest, Marisol’s age equivalent score was 8-0. This indicates that her score was similar to the average performance of children who are 8 years, 0 months old.*
* *Math Application: Marisol’s age equivalent score on the Math Application subtest is 17-11. This means that her score was similar to the average performance of children who are 17 years, 11 months old on the same test.*

1. Explain the meaning of the confidence intervals for Reading Total, Written Language Total, and Math Total.

* *Reading Total: This confidence interval indicates that Marisol’s true test scores falls somewhere between 83 and 94.*
* *Written Language Total: This confidence interval of 85-07 indicates the range of standard scores that is likely to include Marisol’s true score.*
* *Math Total: The confidence interval on the Math Total is 116-132, which indicates the range of standard scores that is likely to include Marisol’s true score.*

1. Based on these scores, how would you describe Marisol’s strengths and weaknesses regarding achievement?

* *In general, Marisol appears to show relative strengths in achievement on the Math Total composite and the Math Computation and Math Application subtests. Relative weaknesses in achievement appear to fall in the Reading Total and Written Language Total composites and their related subtests.*

**Exercise 7.1, pages 142-143**

**Conducting a Test Review: Beck Depression Inventory-II**

1. Describe and evaluate the norm group. Do you think it is representative? Do you think the norm group is current? Do you believe the size of the norm group was large enough? Are the samples related to the population you intend to use the test with? Explain.

* *The BDI norm group appears representative for individuals with depressive symptoms since it is make up individuals from rural and suburban areas who sought outpatient therapy. However, no attempt was made to account for such variables as ethnicity; for example, 91% of the sample consisted of individuals identified as “White,” which clearly does not represent a diverse population.*
* *The BDI-II was published in 1996, which indicates that the norm group may be somewhat out of date.*
* *Because the BDI-II has a narrow focus and was developed to measure the severity of depression in individuals, the relatively small sample size of 500 is appropriate.*
* *As a mental health counselor who is working in an outpatient clinic with adult clients, and who is looking for an instrument that specifically assess depressive symptoms among my clients, the BDI-II standardization sample is related to the population I intend to use it with. Because it was developed for persons 13 years of age and older, it is also representative of the age group of my clients.*

1. Describe and evaluate each method used to estimate reliability. Does the reliability evidence support a decision to use the instrument? Explain.

* *Analysis of internal consistency reliability for the BDI-II yielded a Cronbach alpha of .92 and .93. These are very high reliability coefficients, which indicate that the items within the inventory are homogeneous.*
* *For the test-retest method of estimating reliability, the correlation coefficient over a one week interval was .93 (for a subsample of 26). This coefficient is very high and indicates that scores are consistent from one test administration to the next (with a one week interval).*

1. Describe and evaluate each type of validity evidence. Does the validity evidence support a decision to use the instrument? Explain.

* *Content validity evidence for the BDI-II appears to be strong since the items were designed to be consistent with the diagnostic criteria for depression in the DSM-5.*
* *Convergent validity evidence includes correlations between the BDI-II and other scales measuring depressive symptoms with coefficients of .68, .71, and .89. These coefficients are all very high indicating a strong relationship between BDI-II scores and scores on other similar instruments.*
* *For discriminant validity, correlation coefficients between the BDI-II and two measures of anxiety were provided. These correlations (.47 and .60) indicated a strong relationship between the scores on the BDI-II and the scores on the anxiety measures. This is not strong evidence of discriminant validity.*

1. Describe the practical aspects of the instrument, focusing on issues related to time required for administration, ease of administration and ease of scoring.

* *The BDI-II has 21 items, is a self-report inventory, takes 5-10 minute to complete, and scores are easily calculated by summing the item totals.*

1. Summarize the strengths and weakness of the inventory.

Strengths:

* *Norm group consists of individuals seeking therapy in outpatient clinics and appears large enough.*
* *Strong internal consistency reliability and test-retest reliability*
* *Strong evidence of content validity and convergent validity*
* *The instrument appears very practical to use*

Weaknesses:

* *Norm group does not appear to represent individuals from diverse racial/ethnic backgrounds*
* *Poor discriminant validity evidence*

1. Based on your review of the BDI-II, would you adopt this instrument? Explain your answer.

* *Based on the strong evidence of validity and reliability alone, I would adopt this instrument. However, I would be cautious about its results based on the lack of a diverse norm group, particularly when administering it to non-White individuals.*

**Exercise 7.2, pages 144-145**

**Conducting a Test Review: Coopersmith Self-Esteem Inventory**

1. Describe and evaluate the norm group. Do you think it is representative? Do you think the norm group is current? Do you believe the size of the norm group was large enough? Are the samples related to the population you intend to use the test with? Explain.

* *The SEI norm group appears representative of children in grades 3 to 8. Although the test manual states that “a considerable number of Spanish surnamed and Black children were included in the sample,” it is unclear as to whether the norm group represents the U.S. population of children based on such variables as race, ethnicity, gender, etc. Furthermore, the manual recommends that users develop local norm groups, which imply that the national norm group used to develop the instrument may not be representative.*
* *The SEI was published in 1981. This indicates that the instrument is not current.*
* *The size of the norm group was 643. Because the instrument assesses a relatively narrow construct (i.e., self-esteem), this appears large enough*
* *As a school counselor who is working with culturally and linguistically diverse students, I cannot determine if the SEI norm group is related to the population of students I work with. The SEI is appropriate in terms of grade level; however, specific information regarding the characteristics of the norm group (e.g., race, ethnicity, language, etc.) is not available.*

1. Describe and evaluate each method used to estimate reliability. Does the reliability evidence support a decision to use the instrument? Explain.

* *For the test-retest method of estimating reliability, the correlation coefficient after a five week interval was .88 (for a subsample of 30 fifth graders), which is a high coefficient that indicates stability of test results from the first administration to the second administration. For a three year interval, the test-retest reliability coefficient was .70. Although this coefficient is lower, it is considered acceptable, showing some stability of scores over a three year period.*
* *Studies analyzing the internal consistency reliability for the SEI reported KR20 coefficients ranging from .87 to .92 (for students in grades 4 to 8). Thiese are high coefficients, indicating that the items within the inventory are homogeneous.*
* *Alternate forms reliability information was provided that compared the SEI to a Canadian version of the test. Coefficients ranged from .71 to 80, which are considered acceptable to high.*

1. Describe and evaluate each type of validity evidence. Does the validity evidence support a decision to use the instrument? Explain.

* *Content validity evidence for the SEI appears to be strong since the items were developed based on items adapted from the Rogers and Dymond (1954) scale. Furthermore, items were evaluated by “experts” (i.e., five psychologists) who reviewed the items to determine if they represent the content domain.*
* *Concurrent validity evidence was reported by citing correlations between SEI scores and an achievement test and intelligence test. Coefficients were .33 and .30, which indicated that SEI scores are moderately correlated with achievement test and intelligence test scores.*
* *Predictive validity evidence was provided by showing correlations between the SEI General Self subscale and the SEI Lie scale with measure of reading achievement. Correlations were .35 and .39 respectively, indicating that the reading achievement was moderately correlated with the two SEI scales.*
* *Convergent validity evidence included the correlation between the SEI and the California Psychological Inventory Self-Acceptance scale at .45. This is a high correlation, which indicates a strong relationship between the two instruments.*
* *Subscale intercorrelations were provided that showed high correlation coefficients between several of the subscales on SEI. Moderate to very high correlations were found among four of the subscales (general self, social self-peers, home-parents, school-academic), which provides evidence of homogeneity among subscales. The only subscale that showed very low/unacceptable correlations with other subscales was the Lie scale.*

1. Describe the practical aspects of the instrument, focusing on issues related to time required for administration, ease of administration and ease of scoring.

* *The SEI School Form has 50 items, takes 10 minutes to administer, and can be hand scored in a few minutes.*

1. Summarize the strengths and weakness of the inventory.

Strengths:

* *Norm group appears large enough.*
* *Test-retest, internal consistency, and alternate form reliability evidence is moderate to strong.*
* *Acceptable to strong evidence of several types of validity provided.*
* *The instrument appears very practical to use*

Weaknesses:

* *It is unclear as to whether the SEI norm group represents the U.S. population of children based on such variables as race, ethnicity, gender, etc.*
* *The SEI was published in 1981. This indicates that the instrument is not current.*

1. Based on your review of the SEI, would you adopt this instrument? Explain your answer.

* *Because I’m working with culturally and linguistically diverse students, I would hesitate to adopt this instrument (due to the lack of information about the characteristics of the norm group). Based on the strong evidence of validity and reliability alone, I would adopt this instrument.*

**Exercise 8.1, pages 179-180**

**Understanding Intelligence Assessment: Jackie**

1. Identify the descriptive classifications of the FSIQ, Index scores, and subtest scaled scores.

|  |  |  |  |
| --- | --- | --- | --- |
| **FSIQ and Index Scores** | **Composite Score** | **Percentile** | **Descriptive Classification** |
| Verbal Comprehension (VCI) | 104 | 61 | *Average* |
| Perceptual Reasoning Index (PRI) | 102 | 55 | *Average* |
| Working Memory Index (WMI) | 86 | 18 | *Low Average* |
| Processing Speed Index (PSI) | 91 | 27 | *Average* |
| Full Scale IQ (FSIQ) | 97 | 42 | *Average* |
|  |  |  |  |
| **Subtest Scores** | **Scaled Score** | **Percentile** | **Descriptive Classification** |
| Verbal Comprehension Subtests |  |  |  |
| *Similarities* | 11 | 63 | *Average* |
| *Vocabulary* | 11 | 63 | *Average* |
| *Comprehension* | 11 | 63 | *Average* |
| *Information* | 10 | 50 | *Average* |
| *Word Reasoning* | 10 | 50 | *Average* |
| Perceptual Reasoning Subtests |  |  |  |
| *Block Design* | 9 | 37 | *Average* |
| *Picture Concepts* | 10 | 50 | *Average* |
| *Matrix Reasoning* | 13 | 84 | *Relative Strength* |
| *Picture Completion* | 10 | 50 | *Average* |
| Working Memory Subtests |  |  |  |
| *Digit Span* | 8 | 25 | *Average* |
| *Letter–Number Sequencing* | 7 | 16 | *Relative Weakness* |
| *Arithmetic* | 10 | 50 | *Average* |
| Processing Speed Subtests |  |  |  |
| *Coding* | 8 | 25 | *Average* |
| *Symbol Search* | 9 | 37 | *Average* |
| *Cancellation* | 8 | 25 | *Average* |