TEST REVIEW:
Slosson Intelligence Test - Revised (SIT-R)
Jennifer A. McKechnie, M.Ed. & Bradley T. Erford, Ph.D.
Loyola College in Maryland

I. General Information
   A. Title: Slosson Intelligence Test (SIT-R)
   B. Authors: Richard Slosson; revised by Charles L. Nicholson and Terry Hibpshman.
   C. Publisher: Slosson Educational Publications, Inc., P.O. Box 280, East Aurora, NY 14052.
   D. Forms; groups to which applicable: Ages 4 through adult.
   E. General type: Mental (verbal) ability of children and adults.
   F. Date of publication: 1990; 1998 calibrated norms revision.
   G. Practical features: The SIT-R must be administered individually and hand-scored. No computerized administration, scoring and interpretive software is available. Total Standard Score (M = 100, SD = 16), percentile rank, and mean age equivalent (MAE) interpretive statistics are available.
   H. Cost: $79.00 per test kit.
   I. Time required to administer: The time required to administer and score this test varies from about 15 to 20 minutes for the average person (Kunen, Overstreet, & Salles, 1996) or 20 to 30 minutes for the slow, the very gifted, or the person who is deficient in certain areas while normal or high in other areas (Nicholson & Hibpshman, 1998).

II. Purpose and Nature of the Instrument
   A. Stated purpose: “The SIT-R is designed to be an individual test for use in screening or estimating the cognitive ability of an individual, public school student, college student, mental patient or mentally handicapped person. Because the SIT-R is a screening instrument it alone should not be used in final placement decisions” (Nicholson & Hibpshman, 1990, p. 1). The original Slosson (SIT; Slosson, 1963) was frequently used as a quick screener for mental retardation (McCormick, Campbell, Pasnak, & Perry, 1990; McCuller, Salzberg, & Linnuris Kraft, 1987) and giftedness (Clark, McCallum, Edwards, & Hildman, 1987; Karnes, Whorton, & Currie, 1986). Others have raised concerns over its usefulness for intelligence screening (Harris & Reid, 1991; Salvia & Ysseldyke, 1995).
   B. Description of test, items and scores: Items on the SIT-R are not presented in a subtest format, because the scale is comprised of 187 questions presented in a unidimensional arrangement with age appropriate starting markers. Items were derived from six cognitive domains (Nicholson & Hibpshman, 1990): Vocabulary, General Information, Similarities and Differences, Comprehension, Quantitative, and Auditory Memory.
   C. Use in counseling: The SIT-R is used primarily for assessment of client verbal skills and ability.
III. **Practical Evaluation:**

A. **Usefulness of manual:** The manual for the SIT-R is user-friendly, albeit not comprehensive. Instructions are easy to follow and straightforward. The test is administered orally and has been scripted. Words to be repeated verbatim by the administrator are highlighted in blue.

B. **Adequacy of directions for administering the instrument:** Administration and scoring procedures are clearly explained in the manual. The SIT-R uses a small number of verbal item types and a dichotomous item scoring system making the testing process intuitive to use (Kemphaus, 1994). A concise summary of administration and scoring procedures is located on the inside of the manual’s front cover. Mean Age Equivalent (MAE) and Total Standard Score (TSS) are easily obtained from the norm table in the technical manual (Nicholson & Hibpshman, 1998).

C. **Qualifications of examiners:** According to Nicholson and Hibpshman (1990), to effectively administer, score and interpret the SIT-R the examiner should have Level B qualifications (graduate degree in psychology, education or a related field, and a course in testing and measurement).

D. **Scoring provisions:** A basal score is defined as the highest level at which the examinee obtains ten consecutive correct items. The ceiling item is the last correct item before the examinee misses ten consecutive items. The total raw score is a simple sum of scores and is the total number of correct items above the basal added to the basal item (which gives credit for all the items below the basal item).

IV. **Technical Considerations**

A. **Normative sample:** Children and adolescents aged 4 - 18 (51% female, 49% male) comprised the norm group. The authors suggested that the 18-year-old norms can be generalized to interpret the scores of adults. The 1998 technical manual involved a recalibration of the 1990 norms to reflect recent research and correct for apparent discrepancies between the WISC-III (Wechsler, 1991) and SIT-R. Protocols were collected on more than 2,400 individuals and a sample of 1,854 was selected to match, as close as possible, U.S. population demographics (Nicholson & Hibpshman, 1998). Geographically, the sample was comprised of participants from the Northeast (20%), North Central (22%), South (38%), and West (20%) United States. Racially, the sample was comprised of Black (14%), White (83%), and Other (3%) participants. The Other category included Native, Asian, Hispanic, Pacific Island Americans and other similar groups. The noted disproportion of these individuals was reportedly due to the testing protocol requiring the subject to speak English. The primary language for many of the individuals comprising minority populations was not English (Nicholson & Hibpshman, 1998). Distribution by size of residential populations included: 1,000,000 or more (1%); 500,000 - 999,999 (3%); 100,000 - 499,999 (8%); 50,000 - 99,999 (13%); 25,000 - 49,999 (14%); 10,000 - 24,999 (18%); 5,000 - 9,999 (14%); 2,500 - 4,999 (14%); 0 - 2,499 (14%) (Nicholson & Hibpshman, 1998). Percentages of occupational
groups reported by fathers of children comprising the standardization sample included: Professional (27%), Technical/Office/Sales (28%), Service (20%), Production/Craft/Repair (8%), Operators/Fabricators (11%), and Farm/Fishing/Forestry (7%) (Nicholson & Hibpshman, 1998). Percentages of educational levels reported by fathers of children comprising the standardization sample included: less than high school (19%), high school (39%), some post secondary (24%), college degree and beyond (19%) (Nicholson & Hibpshman, 1998).

B. **Reliability:** Internal consistency was calculated using the Kuder-Richardson - Formula 20 (KR-20). Reliability coefficients determined by age level ranged from .88 to .97, with a median of .945, indicating a high level of internal consistency (Nicholson & Hipbshman, 1996). Based on a sample size of 41 children and using a one-week administration interval, the test-retest reliability was reported to be .96 (Nicholson & Hipbshman, 1998).

C. **Validity:** Concurrent criterion-related validity has been reported through correlations between the SIT-R Total Standard Score (TSS) and the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) IQ's. In four samples of children aged 6 to 16 (total n = 234). The SIT-R TSS correlated .829 - .914 with the WISC-R VIQ (verbal intelligence quotient; median = .861). The SIT-R TSS correlated .376 - .837 with the WISC-R PIQ (performance intelligence quotient; median = .521), while the SIT-R TSS and WISC-R full-scale correlated .612 - .920 (median r = .794; Nicholson & Hibpshman, 1998). Also, “data do not seem sufficient to have a significant bearing on concurrent validity for testing of individuals over the age of 16 years” (Campbell & Ashmore, 1995, p. 116). In a study by Kunen, Overstreet and Salles (1996) the SIT-R was found to correlate r = .92 with the Stanford-Binet (SBIS-4; Thorndike, Hagan & Sattler, 1986), but the SIT-R did not adequately match the SBIS-4 IQ category assignments at levels other than mental retardation. The SIT-R was intended as an intellectual screening test, specifically measuring the verbal intelligence factor. The six domains from which the SIT-R items were developed are similar to items on verbal subtests by Wechsler (1991) and the Stanford Binet Intelligence Scale - Fourth Edition (Thorndike et al., 1986). The domains are meant to measure primarily global intellect, defined by Wechsler as "The capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment" (Nicholson & Hibpshman, 1998, p. 1) and also include memory and crystallized ability. Memory was defined as the ability to recall information about what one has learned, including facts, visual images and taught experiences. Crystallized ability was defined as intelligence related to culture, life experiences and one's environment and can be broken down into two further components: verbal ability and quantitative reasoning (Nicholson & Hibpshman). No exploratory or confirmatory factor analytic results were provided to support the unidimensionality of the SIT-R. Correlations between the WISC –R verbal IQ and SIT-R were very high (see concurrent validity) and the SIT-R does show high internal
consistency coefficients, a source of evidence which is necessary but not sufficient to verify the construct validity of the SIT-R.

V. Evaluation

A. Comments of reviewers and general evaluation: The SIT-R is a quick, easy to use screening test for verbal intelligence. The test instructions are clear and easy to administer in a short period of time, which makes it a useful tool for any professional administering the test. The manual contains an easy to follow summary of administration and scoring procedures, which is located inside the front cover for quick reference. There is also a checklist for examiners with ideas that help with proficiency in administering the test. The cost is considerably less than comparable screening measures and far less than diagnostic IQ tests. The manual is more comprehensive than its predecessor providing adequate reliability data and some basic validity data related to the WISC-R. When used as a verbal intelligence screener, the SIT-R is an effective tool for identifying individuals needing further evaluation. However, evaluators must be careful to only use the SIT-R for screening, not diagnostic decisions.

Weaknesses of the original SIT included “... the lack of a clearly stated theoretical rationale, the use of an inadequate standardization sample, too little information on reliability and validity, the difficulty of making comparisons between SIT IQs and the scores of other tests... and the encouragement by the authors for use of the instrument by untrained examiners...[and] have been addressed in this revision, with varying degrees of success” (Campbell & Ashmore, 1995, p. 116).

However, numerous concerns still remain. The collection of the normative data is vague. There is no information regarding the percentages of participants with disabilities or identified as gifted. This is somewhat disturbing as the SIT-R is often used in determining IQ for individuals with mental retardation and superior intellectual ability. Although it is an effective tool in screening for mental retardation, it does not provide appropriate information to determine the level of mental retardation. The "Slosson Classification Chart" for the Total Standard Scores may encourage misuse of scores. This interpretive chart seems at odds with the cautions against overinterpretation given elsewhere in the manual (Kamphaus, 1994). The table equates low test scores with various levels of mental retardation, and may suggest the TSS score as a means to diagnose or confirm mental retardation.

There is no mention of age differentiation upon which the norms were constructed. The age range is from 4 years to 18+ years with no differentiation past 18. Kemphaus (1994) noted that there is no indication of the smoothing method used or whether or not one was used to produce reasonable distributions of scores from age to age. In addition, a rationale for changing the basal and ceiling levels from seven in the original SIT to
ten consecutive correct or incorrect responses on the SIT-R was not provided and appears to contribute to a lengthening of administration time.

Regarding construct validity, the SIT-R fails to provide evidence of factorial validity. There were 1,854 subjects who participated in the sample, an adequate number to determine factor structure. The authors should have determined whether the SIT-R was unidimensional as hypothesized at the time of sampling, as it will be costly and time consuming to sample another 1,800+ participants to reach reliable factor conclusions. The standardization sample also over represents small population centers and under represents large population centers (Campbell & Ashmore, 1995). No data regarding the number of individuals per age category was provided.

Although the authors claim to have selected a sample to match the U.S. population, the SIT-R manual fails to provide the U.S. census statistics, making comparisons difficult. Also, racial composition did not closely reflect U.S. population demographics. This has been attributed to the testing protocol requiring the subject’s first language to be English. In addition, the norming methods are poorly described and sample characteristics are not broken down by age category. Given the verbal nature of the SIT-R items and under representation of minorities, the SIT-R may be less useful in an increasingly multicultural society in which English may not be an individual’s primary language, as well as for the assessment of less verbal and undereducated individuals (Kamphaus, 1994).

Watson (1994) pointed out that major difficulties continue to exist in understanding the SIT-R’s reliability and validity because of small sample sizes and restricted populations. Regarding test-retest reliability, a sample size of 41 children is insufficient for determining reliability, particularly given that there was no indication of the demographic composition of the sample. Thus, additional data must be gathered before the stability of the SIT-R can be evaluated. Also, while it was helpful to provide concurrent validity coefficients with the WISC-R (Wechsler, 1974), the lack of coefficients with other commonly used diagnostic IQ tests, including the SBIS–4 (Thorndike et al., 1986) and K-ABC (Kaufman & Kaufman, 1983) was noticeable.

REFERENCES


