

Correlations Among the Quick Picture Reading Test,
The Shipley Institute of Living Scale, and the Slosson Intelligence Test-Revised-R-3

Michael Firmin, Ph.D.

Chi-en Hwang, Ph.D.

Jennifer Evens, B.A.

Shelby Keyser, B.A.

Jolene Bennington, B.A.

Cedarville University
Cedarville, OH

Abstract

The purpose of this study was to appraise the potential interrelatedness of three instruments of potential use to school and educational psychologists. This study examined intercorrelations among scores on the Quick Picture Reading Test-Research Edition (QPRT), the Shipley Institute of Living Scale (SILS)- Research Edition, and the Slosson Intelligence Test Revised-3 (SIT-R-3). A sample of 120 children and adults with ages from 7 to 82 completed each test. The results demonstrate correlations between the SIT-R-3, the QPRT, and the SILS vocabulary standard scores were relatively high, while the other correlation coefficients obtained were slightly lower. These results suggest that while the tests may measure some common aspects of reading ability and intelligence, they are also each measuring some particular dimensions of intelligence, providing unique information about the participant.

Keywords: Shipley Institute of Living Scale, Slosson Intelligence Test, Quick Picture Reading Test, Intelligence Testing.

Shipley Institute of Living Scale

The Shipley Institute of Living Scale (SILS) was developed in 1940 by Walter Shipley for the purpose of assessing general intellectual functioning, primarily in adolescents and adults. While generated almost 65 years ago, the Shipley “survives today as an often-used technique for the brief estimation of IQ” (Weiss & Schell, 1991). In fact, the SILS is one of the few published tests that does not require oral individual administration in order to yield a quick, overall standard score of intellectual functioning. Administering and scoring the test takes relatively little training, although as with all standard instruments measuring intellectual functioning, advanced education is required for interpretation. In sum, because the SILS is one of the few available quick screening IQ tests, it enjoys popular contemporary use.

Retzlaff, Slicner, and Gibertini (1986) found a modest correlation between scores on the Wechsler Adult Intelligence Scale-Revised (WAIS-R) and the SILS ($r=.46$) among non-clinical research participants. Klett and Watson (1986) used a psychiatric population, finding a correlation of .79 between WAIS-R scores and the patients’ scores on the SILS. Weiss and Schell (1991) estimated WAIS-R standard scores from SILS scores in a psychiatric population. Their results indicate that the score on the WAIS-R was highly correlated with the obtained SILS IQ ($r = .86$).

Bowers and Pantle (1998) assessed the comparability of the SILS and the Kaufman Brief Intelligence Test (K-BIT), an instrument used to screen overall intelligence levels in adults. The obtained results indicate that even within two diverse groups, the K-BIT and the SILS yielded comparable scores. Particularly, they found correlations of .77 in a college sample and .83 in a forensic sample.

Tamkin and Jacobsen (1984) found that generally SILS scores tend to decline continuously through lifespan decades, except for an unexplained small spike when patients are in their 70’s. Their sample consisted of veteran psychiatric inpatients. Harnish, Beatty, Nixon, and Parsons (1994) criticized the SILS test for norm deficiencies in middle-aged and older adults. They further indicated that most of the published information about the SILS from non-clinical populations used college students, children, and nursing students. Dalton and Dubnicki (1981) found some relatively minor racial differences in SILS scores among inpatients in alcoholic treatment.

Slosson Intelligence Test

The Slosson Intelligence Test (SIT) was developed in 1963 and is an untimed, individually administered test, intended to be utilized as a brief screening instrument, assessing mental ability in children and adults (Campbell & Ashmore, 1995). In its original edition, the SIT was relatively atheoretical. Questions were comprised based on their psychometric loading vis-a-vis the overall composite score and similar to ones used in previously published IQ tests. Hammil (1969) reported that the SIT could be administered successfully by teachers with no prior course work in testing procedures. Because of its relatively brief time for administration and reported concurrent validity with other IQ tests, as Stone (1975) noted, the SIT has had the reputation of being the busy practitioner’s workhorse.

According to Harris and Reid (1991), the most recent edition of the SIT is not truly a revised test. It has no new test items because the present items either were selected from or are similar to items from the 1960 Stanford Binet. The SIT has many critics, and Campbell and Ashmore (1995) summarize some of the problems: a lack of clearly stated theoretical rationale, the utilization of an independent standardization sample, and minimal evidence of reliability and

validity. Bohning (1980) adds that the item analysis sheet included with the test has no reported psychometric properties.

Carney and Karfgin (1971) reported a .92 correlation between SIT and WAIS scores in forensic psychology populations. Hammill (1969) showed a .81 correlation coefficient between SIT scores and the Peabody Picture Vocabulary Test (PPVT) among children without mental retardation. Among candidates for special education, however, Covin (1976) reported lower correlations ($r=.41$ and $r=.51$) with the PPVT. Hale, et al. (1978) reported only moderate correlations between composite SIT scores and Wide Range Achievement Test (WRAT) scores for children referred for psychological evaluations. Namely, they reported correlations of .37 for Reading, .32 for Spelling, and .40 for Arithmetic.

Using gold standard IQ tests such as the Wechsler scales and Stanford-Binet, published literature has shown variable results, but overall, that the SIT seems relatively robust. Nicholson (1970) expected correlation coefficients between the SIT and SB to be generally high due to categorical inter-correlations among the SIT's domains by age levels, although differences would be expected to level-out by around age 15. Ritter, Duffy, and Fischman (1973) reported a correlation of .92, using the SIT and SB, in children 4-12 years of age who were referred for evaluations by their teachers. Stewart and Myers (1974) showed correlations of .83 for the Wechsler Intelligence Scale for Children (WISC) and .84 for the Stanford-Binet, L-M (SB) when comparing overall composite scores. Children in their sample were being screened for special education purposes. Using a population of low achieving students, Prewett and Fowler (1992) showed a correlation of .75 between students' scores on the SIT and WISC-R. The results are similar to a more recent study by Blackwell and Madere (2005), showing a correlation of .89 between the scores of 234 children (ages 6-16) when comparing WISC-III scores with the children's SIT-R-3 scores.

In short, three decades of research with the SIT shows a pattern of relatively strong correlations among a number of widely-accepted instruments. Nonetheless, the SIT is not without significant critics. For example, Stewart and Jones (1976) recognized the SIT's value as a general screener, but warned against categorizing students with the test. Other than very broad and general categories, such as below average, average, and above average, they advocate that the Slossen is inappropriate for finer distinctions such as range-levels of mental retardation. Harris and Reid (1991) claim that the SIT overestimates IQ scores, particularly with gifted individuals and use of the test with specific populations—particularly learning disabled students—is questionable. They cite potential ethical concerns in utilizing the instrument with populations outside of typical or psychometrically normal.

Quick Picture Reading Test

The Quick Picture Reading Test (QPRT) is a product being developed by Western Psychological Services (in press) in order to provide a brief screening assessment of reading abilities. Among other potential applications, the test is expected to be useful for psychologists administering test batteries where a general screening level of reading aptitude is desired. The test presents 35 pictures to be matched with 26 phrases. It is intended to assess reading ability in a timely manner, with results being quickly scored and yielding an overall standard score.

Although the QPRT is a newly developed instrument, a precursor to the test, Borgatta and Corsini's (1964) Quick Word Test (QWT), showed apt psychometric properties with respect to reading and general aptitude abilities. Particularly, Martin, Blair, and Vickers (1979) reported respectable correlations among the QRT, SIT, and Wide Range Vocabulary Test (WRVT).

Using undergraduate psychology students in their sample, the researchers found the following coefficients: QWT—SILS=.68; WRTV—SILS=.73; and QWT—WRT=.81

The QPRT is highly loaded relative to verbal intelligence. This occurs since students are required to identify the meanings of words as they relate to pictures they readily should recognize, if they know the apt meanings of words. This general type of activity sometimes comprises subtests in a standardized IQ test. Moreover, these activities often highly load into the overall composite IQ score. Consequently, in cases where a general screening or overall gist of students' IQs are desired—rather than a complete, comprehensive assessment of intelligence—using the QPRT might prove to be a useful alternative for those cases. Additionally, the QPRT can be administered by teachers and other personnel without a psychology license. Its use may provide breadth of potential use—in addition to standard IQ tests often utilized by school psychologists.

With respect to construct validity (Thordike, 1997), the three tests share the shared construct of verbal intelligence. In classic conceptualization, IQ tests often focus both on subjects' verbal and non-verbal abilities (Murphey & Davidshofer, 1998). The Wechsler scales, for example, report scores in terms of verbal and performance measures, as well as full scale indexes that combine these domains. From a construct validity standpoint, the tests we selected for the present study, a priori, would be expected to share similar, strong verbal loadings in generating standards scores. While obviously each test would be expected to measure differing elements of verbal ability, we would anticipate them to share correlative measures when administered to the same individuals.

Our present research project builds on this rich history of concurrent validity assessments. Only one study in the literature specifically compares the SILS with the SIT. Martin, Blair, Stokes, and Lester (1977) administered the instruments to 40 undergraduate psychology students. The students were administered the tests twice, with a 45 day interlude separating the administrations. The correlation coefficient between the SILS and SIT for the first administration was .46. The second administration's coefficient between the SILS and SIT was .54. The test-retest reliability coefficients of .83 and .80 (respectively) also were reported.

In short, 30 years have transpired since any correlative measures between the SILS and SIT have been obtained. The tests have continued, of course, to be used with wide popularity among psychologists. Consequently, we believed it was time for reporting updated data on these instruments. The version of the SILS used in the present study is the research edition with an expected publication in the near future by WPS. Since it soon will be the "new Shipley," we utilized it rather than the old (current) version. Although we believe results from the present study may be useful to the new Shipley manual, none of the present researchers are employees of WPS.

The sample of this study was drawn from a non-clinical population. We selected this population since many of the validity and reliability studies previously conducted have taken their samples from atypical or psychiatric populations. The primary purpose of this study was to determine the degree of correlation between scores on the SIT and SILS on a normal population, ages ranging from 7 to 80+ years. A secondary purpose was to assess potential correlation between the scores of the QPRT and the SIT as well as determine the degree of correlations between the QPRT and SILS sub-scores.

Results from our present study are intended to be useful for both clinicians and researchers. That is, we believe our findings help psychologists who use these instruments with clinical utility and ecological decisions when selecting instruments for use in psychological testing batteries. In addition, the findings also contribute to the larger body of research knowledge relative to the instruments studied. The purpose of this study was to appraise the

potential interrelatedness of three instruments of potential use to school and educational psychologists. Since the QPRT, SILS, and SIT-R have similar potential uses in educational settings, educational specialists will benefit from better knowing their correlative relationships. Consequently, our aim was to provide empirical data relating to the tests inter-correlations and to discuss the implications that these relationships have in educational settings.

Method

A total of 120 participants (53 males and 67 females) were included in the comparative study of the QPRT, SILS, and SIT-R-3. Their ages ranged from seven years old to eighty-two years old. For both children and adults, we obtained a relatively even distribution of age ranges across targeted age categories. These included 50 children, 19 teenagers, 19 individuals between ages 20 to 50, and 22 individuals between 51 and 82. The majority of the participants (93% or $n = 112$) were Caucasians. The sample of children who participated in the present study consisted of elementary through high school students. The adults in the study were drawn from volunteers living in a mid-sized Midwest city who agreed to be tested, without compensation or extrinsic incentives. Data were collected during a 15-week time-frame.

The conjoined QPRT and SILS, and SIT-R-3 were administered to the individuals during one sitting. The SIT-R-3 was administered in a one-on-one setting to each participant by a test administrator, while the QPRT and SILS were administered collectively to the respective age groups. The lead researcher is a licensed psychologist, experienced in IQ testing, and the research team who administered the instruments were rigorously trained in standardization procedures. This included administering tests under supervision and administering tests flawlessly prior to engaging in data collection.

Results

Table 1 presents the intercorrelations of the QPRT, SILS and SIT-R-3 standard scores. All the Pearson Product-Moment correlation coefficients were significant at the .01 level except for the correlation between the SILS Vocabulary score and the SILS Block Pattern score ($r = .20$, $p < .05$). The correlation between the QPRT and the SILS was considered high for the SILS Vocabulary standard score, with a coefficient of .58, and was slightly lower for the SILS Abstraction and Block Pattern standard scores, with coefficients of .37 and .30, respectively.

Table 1
Intercorrelations of the QPRT, SILS and SIT-R-3 Standard Scores

	SILS Voc.	SILS Abst.	SILS Block	SIT Total
QPRT	.58	.37	.30	.52
SILS Vocabulary		.30	.20	.47
SILS Abstraction			.28	.36
SILS Block Pattern				.29

The correlation between the SILS standard scores and the SIT-R-3 total standard score, like the QPRT, was high for the SILS Vocabulary Test, with a coefficient of .47 ($p < .01$), and slightly lower for the SILS Abstraction and Block Pattern standard scores, with correlations of .36 ($p < .01$) and .29 ($p < .01$), respectively. Compared to the intercorrelations with scores on other tests, the correlation between the SILS subtest standard scores were relatively lower with .30 between the SILS Vocabulary score and SILS Abstraction score, and .20 between the SILS Vocabulary score and Block Pattern score.

Discussion

The results of this study provide updated validity information of the tests involved. Among all three tests, the SILS has the longest history. Research from the literature suggests it as the most apt assessment tool with established validities and reliabilities and can serve as an external criterion for the validation of the other two instruments while both the SILS and SIT can be used as external criteria the more recently developed QPRT. The correlation coefficients of .52 ($p < .01$) between the QPRT and the SIT-R-3, and .58 ($p < .01$) between the QPRT and the SILS vocabulary standard score can be treated as the concurrent validity coefficients for the QPRT. These evidences suggest that the QPRT may be particularly adept for some simple verbal ability purposes. That is, clinicians may be able to obtain a quick and rough estimate of reading ability that moderately correlate to general intellectual ability in situations where taking the time and money to administer an entire SILS or SIT-R-3 is not feasible. The correlation coefficients were not high enough to suggest that the QPRT may be used to obtain the detail and accuracy of the information gained from the SIT-R-3 and the SILS, or that the QPRT accurately can be used in place of either of the two tests.

Although the correlations between the SILS and the SIT-R-3 all were significant at the .01 level, the correlations were not high enough to suggest that these tests measure identical constructs. The SILS and the SIT-R-3 may be measuring different dimensions of general intelligence. The correlation between the SIT-R-3 total standard score and the SILS vocabulary standard score was the highest of the three correlation coefficients obtained for the SIT-R-3 and the SILS. This suggests that both tests may provide similar measures of verbal dimension. The lower correlations between the standard scores for the SIT-R-3 and the abstraction and block pattern standard scores suggest that the SILS may measure several dimensions of intelligence that the SIT-R-3 does not. While we believe these tests are both useful screening tools for estimating intelligence, one test cannot replace the other to obtain the same information.

In sum, the SIT-R-3 and SILS have uses for clinicians wishing to obtain quick screening of general intellectual ability. The milieus (e.g. forensic or school settings) for the tests' uses often are similar, but we advise clinicians to select the SIT-R-3 or SILS deliberately, and with the clinicians' particular objectives in mind. Verbal intelligence seems to show significant overlap between the two instruments, although enough difference exists to indicate the two tests should not be considered interchangeable for practical use.

In school settings, the findings from the present study have particular implications for educators. First, teachers may find the QPRT to be a useful measure of verbal abilities in students whom they suspect struggle with reading difficulties. Obviously, teachers should not use the QPRT for diagnosing students with reading deficiencies, but it might provide useful data on which potential referrals to school psychologists might be based. In other words, having some objective data to supplement classroom observations and impressionistic hunches can be particularly useful to educators.

The SILS does not require the professional to individually administer the instrument to students. As such, the test can be included as part of a larger test battery. The Slosson, in contrast, requires individual administration by a trained professional (e.g., school psychologist, special education teacher, etc.). Consequently, since the tests reasonably correlate, the selection of either instrument in educational settings might depend somewhat on pragmatic factors—such as the amount of time a given teacher or school psychologist has for testing the particular student at the time when the data is needed. However, knowing the significant psychometric similarity between the two tests can help provide needed confidence that school personnel can achieve similar measures from either instrument.

Limitations and Future Research

All good research possesses limitations and researchers do well to report them when known (Murnan & Price, 2004). First, our sample consisted mostly of Caucasian individuals. Obviously, minorities were not screened out of the study; they simply were not available (including signed parental permissions for children) during the 15 period available for data collection. Future endeavors should replicate the present study, including larger proportions of minority individuals. Additionally, future researchers may wish to replicate the study using all minority students—comparing the present those findings with the present ones representing Caucasians. External validity of the overall construct under investigation in this study would be significantly enhanced (Johnson & Christensen, 2004).

Second, we used the research editions of the QPRT and the SILS. At the present moment, these instruments are not available to the general public for purchase. However, given the amount of time likely for the present article to see print and the publication intentions of WPS, we believe that this data will be very timely. Using a version of the Shipley that likely would be out of date shortly after the article's publication simply did not seem to be the best use of our research efforts.

A third limitation involves our sample's size and normalcy of participants. While we believe that 120 individuals provided enough power to sufficiently detect differences (Kraemer & Thiemann, 1987), obviously having more participants in our sample would have increased the study's power. The decision to use 120 subjects was driven by the trade-off between apt sample size and time allotted for data collection. In short, having the data collection completed in a 15 week period enabled us make the study reasonably feasible to conduct and reduce historical effects as a potential threat to internal validity. Additionally, a limitation relating to our sample was that we used normal (non-clinical) population group. The research literature suggests that variability may exist between clinical and non-clinical groups—so our findings may not generalize as strongly to atypical populations as well as they do to non-clinical ones.

References

- Blackwell, T. L. & Madere, L. N. (2005). Test review. *Rehabilitation Counseling Bulletin*, 48, 183-184.
- Bohning, G. (1980). Item analysis of Slosson Intelligence Test: A review. *Psychology in the Schools*, 17, 339.
- Borgatta, E. & Corsini, R. (1964). *Manual for the Quick Word Test*. New York: Harcourt, Brace and World.
- Bowers, T. L. & Pantle, M. L. (1998). Shipley Institute for Living Scale and the Kaufman Brief Intelligence Test as screening instruments for intelligence. *Assessment*, 5, 187-195.
- Campbell, C. A. & Ashmore, R. J. (1995). The Slosson Intelligence Test-Revised (SIT-R). *Measurement & Evaluation in Counseling & Development*, 28, 114-116.
- Carney, F. L. & Karfgin (1971). Using the Slosson Intelligence Test with a criminal population. *Correctional Psychologist*, 4, 233-237.
- Covin, T. M. (1976). Comparison of Slosson and Peabody IQs among candidates for special education. *Psychological Reports*, 39, 814.
- Dalton, J. E. & Dubnicki, C. (1981). Sex, race, and education variables in Shipley-Hartford scores of alcoholic inpatients. *Journal of Clinical Psychology*, 37, 885-888.
- Hale, R. L., Douglas, B., Cummins, A., Rittgarn, G., Breed, B., & Dabbert, D. (1978). The Slosson as a predictor of Wide Range Achievement Performance. *Psychology in the Schools*, 15, 507-509.
- Hammil, D. (1969). The Slosson Intelligence Test as a quick measure of mental ability. *Journal of School Psychology*, 7, 33-37.
- Harnish, M. J., Beatty, W. W., Nixon, S. J., & Parsons, O. A. (1994). Performance by normal subjects on the Shipley Institute of Living Scale. *Journal of Clinical Psychology*, 50, 881-883.
- Harris, K. R. & Reid, R. (1991). A critical review of the Slosson Intelligence Test. *Learning Disabilities Research & Practice*, 6, 188-191.
- Johnson, B. & Christensen, L. (2004). *Educational research*. Boston: Allyn & Bacon.
- Klett, W. G. & Watson, C. G. (1986). The Henmon-Nelson and Slosson Tests as predictors of WAIS-R IQ. *Journal of Clinical Psychology*, 42, 343-347.
- Kraemer, H. C. & Thieman, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park, CA: Sage.
- Martin, J. D., Blair, G. E., Stokes, E. H., & Lester, E. H. (1977). A validity and reliability study of the Slosson Intelligence Test and the Shipley Institute of Living Scale. *Educational and Psychological Measurement*, 37, 1107-1110.
- Martin, J. D., Blair, G. E., & Vickers, D. M. (1979). Correlation of the Quick Picture Word Test and Wide Range Vocabulary Test with the Shipley-Institute of Living Scale. *Educational and Psychological Measurement*, 39, 935-937.
- Murnan, J. & Price, J. H. (2004). Research limitations and the necessity of reporting them. *American Journal of Health Education*, 35, 66-67.
- Murphy, K. R., & Davidshofer, C. O. (1998). *Psychological testing* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Nicholson, C. (1970). Analysis of functions of the Slosson Intelligence Test. *Perceptual and Motor Skills*, 31, 627-631.
- Prewett, P. N. & Fowler, D. B. (1992). Predictive validity of the Slosson Intelligence Test with the WISC-R and the WRAT-R Level 1. *Psychology in the Schools*, 29, 17-21.

- Retzlaff, P., Slicner, N., Gibertini, M. (1986). Predicting WAIS-R scores from the Shipley Institute of Living Scale in a homogeneous sample. *Journal of Clinical Psychology, 42*, 357-359.
- Ritter, D., Duffy, J., Fischman, R. (1973). Comparability of Slosson and S-B estimates of intelligence. *Journal of School Psychology, 11*, 224-227.
- Stewart, K. D. & Jones, E. C. (1976). Validity of the Slosson Intelligence Test: A ten year review. *Psychology in the Schools, 13*, 372-380.
- Stewart, K. D. & Myers, D. G. (1974). Long-term validity of the Slosson Intelligence Test. *Journal of Clinical Psychology, 30*, 180-181.
- Stone, M. (1975). An interpretative profile for the Slosson Intelligence Test. *Psychology in the Schools, 12*, 330-333.
- Tamkin, A. S. & Jacobsen, R. (1984). Age-related norms for the Shipley Institute of Living Scale. *IRCS Medical Science, 12*, 459-460.
- Thordike, R. (1997). *Measurement and evaluation in psychology and education* (6th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Weiss, J. L. & Schell, R. E. (1991). Estimating WAIS-R IQ from the Shipley Institute of Living Scale: A replication. *Journal of Clinical Psychology, 47*, 558-562.
- Western Psychological Services. (in press). *Quick Picture Reading Test*. Los Angeles: WPS.