Psychological Reports, 1998, 82, 619-625.

CONCURRENT VALIDITY OF THE SNIJDERS-OOMEN NONVERBAL INTELLIGENCE TEST 2½-7–REVISED WITH THE WECHSLER PRESCHOOL AND PRIMARY SCALE OF INTELLIGENCE–REVISED[1]

CRISTAL MOORE, STEPHEN L. O'KEEFE AND DEL LAWHON

West Virginia Graduate College

PETER TELLEGEN

University of Groningen, The Netherlands

Summary.—This study examined the concurrent validity of the Snijders-Oomen Nonverbal Intelligence Test–Revised compared to the Wechsler Preschool and Primary Scale of Intelligence–Revised. Subjects were 25 4-yr.-olds of lower, lower-middle, and middle socioeconomic status from both urban and rural areas of Appalachia. The SON–R IQs correlated .93 and .87 with the WPPSI–R Performance IQs and Full Scale IQs, respectively. The correlation of .45 with Verbal IQs was not significant.

Interest in the assessment of young children has gone beyond narrow clinical concerns and research interests since the advent of Public Law 99-457, which mandates extension of public education and developmental services to handicapped children between the ages of 3 and 5 years. The law has given added impetus to the need for valid, reliable assessment instruments for young children, particularly those from special populations whose hearing, speech, or language may be impaired, and whose language is different from that of the examiner. Currently available nonverbal tests of preschool intelligence are plagued with such problems as inadequate standardization, outdated norms, and questionable reliability or construct validity, and are not recommended for use in decision-making situations (Sattler, 1992). A common approach to assessing children for whom traditional measures are inappropriate is nonstandard administration (pantomimed as for deaf children) of Wechsler Performance Scale subtests; however, this use is not supported by normative data and gains only equivocal support in the literature (Kaplan, 1993; Lowe, Anderson, Williams, & Currie, 1987).

The original Snijders-Oomen Nonverbal Intelligence Test (Snijders-Oomen, 1943) was devised to assess the spatial abilities, abstract reasoning, concrete reasoning, and memory functioning of deaf children. Subsequent revisions expanded the test, provided parallel norms for hearing and deaf

[1]Address enquiries to second author at 100 Angus E. Peyton Drive, South Charleston, West Virginia 25303-1600.

children, and developed separate tests for preschool and school-age children (Tellegen & Laros, 1993a). The new standardization edition of the Snijders-Oomen Nonverbal Intelligence Test for children age 2 years, 6 months to 7 years (SON–R 2 1/2-7; Tellegen, Winkel, & Wijnberg-Williams, 1996) was used in this study. The purpose of this latest revision was to provide updated norms and to incorporate new test construction, administration, and psychometric methods (Tellegen & Laros, 1993a).

Perhaps the greatest departure from standardized testing is the provision of feedback following each item administered. The examiner scores the error, then attempts to engage the child in making the necessary corrections or demonstrates the correct solution. The intent is to create a "more natural situation" and to provide the child an opportunity to learn (Tellegen & Laros, 1993b). The effect of this type of administration on test performance is unknown. Glutting and McDermott (1989) examined the effect of standardized teaching items on subsequent performance and found no statistically significant differences. However, their standardized teaching items differed in presentation from the method employed in the SON–R in that they taught the task prior to a standard test administration. Every failed item on the SON–R becomes a teaching item, providing an opportunity to adjust strategy and making the test perhaps a measure of learning potential.

Should the SON–R have good concurrent validity, it could provide an attractive alternative means of assessing cognitive functioning of the deaf and others for whom traditional measures are inappropriate given language impairments or cultural/linguistic differences.

The purpose of this study was to investigate the concurrent validity of the SON–R and the Wechsler Preschool and Primary Scale of Intelligence–Revised (WPPSI–R; Wechsler, 1989). The WPPSI–R is generally accepted as a psychometrically sound measure of preschool cognitive abilities (Bracken, 1992; Braden, 1992; Kaufman, 1990; Sattler, 1992). It yields three scores of Performance IQ, Verbal IQ, and Full Scale IQ. The SON–R yields a single IQ score which was expected to correlate positively with all three WPPSI–R IQs. The SON–R IQ was expected to correlate most strongly with the Performance IQ and least with the Verbal IQ.

Method

Subjects

Children age 4 years in private day care, preschool and public preschool programs were identified and given the opportunity to participate in the study. Children with severe handicaps were not included. Also, one girl, who was obviously not motivated, was excluded from the sample. The remaining sample consisted of 25 subjects, 15 boys and 10 girls. Mean age at

the time of the SON–R administration was 52.4 mo., with a standard deviation of 3.6 mo. Of the subjects 14 were from urban areas and 11 were from rural areas. Nine subjects were from families of lower-middle to middle-class socioeconomic status, and 16 from lower-socioeconomic backgrounds.

Procedure

Graduate students with appropriate training in testing administered the Snijders-Oomen Nonverbal Test of Intelligence–Revised, 2^1/$_2$-7, and the Wechsler Preschool and Primary Scale of Intelligence–Revised in counterbalanced order. Average time between tests was 12 days with a range of 0 to 56 days.

Tests

The SON–R was individually administered according to instructions provided by the test's authors in a prepublication manual (Tellegen, Winkel, & Wijnberg-Williams, 1994).

There are six subtests, each in two parts, with items of increasing difficulty. Mosaics requires the subject to copy a mosaic pattern in a frame using red, yellow, or red/yellow squares or some combination of the squares. The patterns to be copied are pictured in a stimulus booklet and the first three items are demonstrated by the examiner. Categories Part I involves sorting picture cards in a forced-choice format, i.e., flower or candy. Part II is in multiple-choice format, whereby the child chooses among five cards the two that have something in common with three pictured objects. Puzzles in Part I are constructed within a frame to resemble an example. In Part II the frame and examples are not used. The Analogies subtest requires separate test booklets for each part. In the first part the child sorts blocks into two boxes according to their shape, color, size, or some combination of attributes cued by pairs of pictures in the test booklet. Part II is multiple-choice and involves the transformation of geometric figures. Situations items in Part I depict four objects with half of the picture missing. The child matches the other halves which are printed on cards. Part II requires the child to choose a card or cards from several alternatives to complete a drawn situation with part(s) missing. Patterns are geometric forms copied by the child in the test booklet. A time limit of 150 sec. is imposed on items in Parts II of the subtests Mosaics, Puzzles, and Patterns.

The six standardized scores with a mean of 10 and a standard deviation of 3 are combined to give the IQ. Based on principal component analysis separate total scores are presented for the three performance tests Mosaics, Puzzles, and Patterns (Performance Scale) and for the three reasoning tests Categories, Situations, and Analogies (Reasoning Scale). A similar distinction was found in the SON test for older children (Snijders, Tellegen, & Laros, 1989). The normative group consisted of a stratified representative sample of

1124 Dutch and immigrant children in The Netherlands. The reliability of the IQ is .90; the Performance and the Reasoning Scale have a reliability of .85 and the average reliability of the subtests is .72.

A study of the concurrent validity of the SON–R IQ with the WPPSI–R Performance, Verbal, and Full Scale IQs, conducted with 59 nondisabled children in Australia, showed correlations of .74, .53, and .74, respectively (Jenkinson, Roberts, Dennehy, & Tellegen, 1996; Tellegen, 1997).

The ten basic subtests of the WPPSI–R (Wechsler, 1989) were administered and scored according to instructions in the manual. For use with 4-yr.-olds the WPPSI–R is highly reliable. The manual reports subtest reliability coefficients ranging from a low of .63 for Object Assembly to .86 for Similarities, with most falling in the low .80's. Average test-retest reliabilities across the age ranges for the Performance, Verbal, and Full Scale IQs were .92, .95, and .96, respectively. At the 4 and $4^1/_2$ year levels they were .93, .96, and .97. Interscorer agreement for those subtests which are scored subjectively ranged from .88 to .96. Factor analysis yielded a two-factor solution and supports the construct validity. Studies comparing the WPPSI–R and the WPPSI (Wechsler, 1967), the Stanford-Binet Intelligence Scale: Fourth Edition (Thorndike, Hagen, & Sattler, 1986), the McCarthy Scales of Children's Abilities (McCarthy, 1972), and the Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1983) indicate good concurrent validity, with correlations ranging from the .50's to the .80's and score differences between 2 and 7 points (Wechsler, 1989).

Both tests were scored by the examiner according to instructions in the manuals. The protocols were reviewed by a second, trained examiner to verify the scores.

Results

Pearson product-moment correlations of the standardized SON–R scores with the WPPSI–R IQs and the means and standard deviations are presented in Table 1. All correlations of the SON–R scores with the WPPSI –R Performance and Full Scale IQs are significant ($p < .01$). The correlations with the WPPSI–R Verbal IQs are substantially lower and, with the exception of Analogies, not significant. The correlations of the SON–R IQs with the WPPSI–R Performance, Verbal, and Full Scale IQs are .93, .45, and .87, respectively. Scores on the Performance Scale of the SON–R correlated higher than those on the Reasoning Scale of the SON–R with all WPPSI–R IQs. The mean score of the SON–R Performance Scale is significantly lower than that for the SON–R Reasoning Scale ($t = 3.1$, $p < .01$). The SON–R mean IQ is significantly lower than the WPPSI–R Full Scale mean IQ ($t = 3.8$, $p < .01$). The difference between WPPSI–R Performance IQ and Verbal IQ is not significant ($t = 1.21$, $p < .20$).

TABLE 1
SON-R AND WPPSI-R MEANS AND STANDARD DEVIATIONS FOR IQ

Measure	*M*	*SD*
WPPSI-R		
Performance IQ	97.4	15.3
Verbal IQ	94.1	8.4
Full Scale IQ	95.0	11.1
SON-R		
Performance Scale	85.2	16.0
Reasoning Scale	93.7	18.5
Total IQ	87.5	17.8

DISCUSSION

It appears that the SON-R has fair concurrent validity with the WPPSI-R and is measuring some similar constructs, particularly in the Performance domain. However, these results must be interpreted cautiously as the sample is small and the characteristics of the sample unrepresentative.

The difference of 7.5 points between the mean SON-R IQ and the WPPSI-R Full Scale IQ raises the question of the validity of Dutch norms for use in the USA. Item bias could be a factor contributing to the large difference between the mean IQs. The SON-R is intended for use with children from varied cultures and, like many nonverbal tests, is presumed to be culture-reduced. Indeed, the subtests Patterns, Mosaics, and Analogies appear to be free of items which are specific to any culture. However, the subtests Puzzles, Situations, and Categories depict objects or social situations which are supposed to be common but may be less familiar to non-Europeans. This interpretation is not supported by the data. Although the lowest mean score was obtained with Puzzles, the scores on Situations and Categories were relatively high while the scores on Mosaics and Patterns were relatively low.

A factor that certainly should be taken into account is the period of approximately seven years between the standardization of the SON-R $2^1/_2$-7 and the WPPSI-R. By comparing IQs on the WPPSI and the WPPSI-R (with an interval of 22 years between the standardizations) the scores on the latter test were 8 points lower (Wechsler, 1989). Assuming the same rate of change, IQs of the WPPSI-R would be 2.5 points lower had the test been standardized at the same time as the SON-R $2^1/_2$-7. In the Australian study, in which U.S. norms were used for computing WPPSI-R IQs, the mean difference in Full Scale IQs with SON-R IQs was 3.3 points (Tellegen, 1997). Whether this difference is larger in the USA is hard to conclude given the size of this U.S. sample; the 95% confidence zone for the difference between the means runs from 3.7 to 11.3.

The high correlation with the WPPSI-R Performance IQ suggests that

the SON–R is a valid measure of nonverbal cognitive functioning and may be useful in the evaluation of young children for whom traditional measures are inappropriate. Concurrent validity with more diverse groups of children should be explored. Interpretation of subtest scores, instructional implications, and the predictive validity of the SON–R await further study.

Post hoc analysis of the WPPSI–R scores showed some interesting, although not statistically significant, differences between this sample and the normative group. The correlation between Performance scaled score totals and Verbal scaled score totals for these children was .45, compared with .55 to .64 for the same age range in the standardization sample. Also, the discrepancy between Performance IQs and Verbal IQs was larger for this sample. Performance IQ-Verbal IQ discrepancies of more than eight points occurred in 50% of the normative group, but 69% or 19 of the present sample had differences of nine or more points. Gyurke, Prifitera, and Sharp (1991) examined the Performance–Verbal IQ discrepancies for the standardization group by ability and for the group as a whole. Over-all, 48.5% had Performance IQs higher than Verbal IQs. In the present study, 64% of the children had Performance IQs higher than Verbal IQs. The mean discrepancy for this sample was 13.6 compared to a mean of 11.5 for the normative group. These findings revive the notion, often raised but seldom supported, that Appalachian children are linguistically different and therefore penalized by intelligence tests with high verbal content (King, 1990). The high incidence of Performance IQ > Verbal IQ as compared to the normative group and the low correlation of the SON–R with the Verbal IQ relative to that found in the Australian study suggest that the notion may not be unfounded. Comparison of first-grade achievement with SON scores might be useful.

REFERENCES

BRACKEN, B. A. (1992) Review of the Wechsler Preschool and Primary Scale of Intelligence–Revised. In J. J. Kramer & J. C. Conoley (Eds.), *The eleventh mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements. Pp. 1027-1029.

BRADEN, J. P. (1992) Review of the Wechsler Preschool and Primary Scale of Intelligence–Revised. In J. J. Kramer & J. C. Conoley (Eds.), *The eleventh mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements. Pp. 1029-1031.

GLUTTING, J. J., & MCDERMOTT, P. A. (1989) Using "teaching items" on ability tests: a nice idea, but does it work? *Educational and Psychological Measurement*, 49, 257-268.

GYURKE, J. S., PRIFITERA, A., & SHARP, S. A. (1991) Frequency of Verbal and Performance IQ discrepancies on the WPPSI–R at various levels of ability. *Journal of Psychoeducational Assessment*, 9, 230-239.

JENKINSON, J., ROBERTS, S., DENNEHY, S., & TELLEGEN, P. (1996) Validation of the Snijders-Oomen Nonverbal Intelligence Test–Revised 2 1/2-7 for Australian children with disabilities. *Journal of Psychoeducational Assessment*, 14, 276-286.

KAPLAN, C. (1992) Ceiling effects in assessing high-IQ children with the WPPSI–R. *Journal of Clinical Child Psychology*, 21, 403-406.

KAUFMAN, A. S. (1990) The WPPSI–R: you can't judge a test by its color. *Journal of School Psychology*, 28, 387-394.

KAUFMAN, A. S., & KAUFMAN, N. L. (1983) *Kaufman Assessment Battery for Children: interpretive manual*. Circle Pines, MN: American Guidance Service.

KING, J. K. (1990) Comparison of the language development of rural Appalachian preschoolers and urban Appalachian preschoolers. Unpublished master's thesis, West Virginia Graduate College, Institute, WV.

LOWE, J. D., JR., ANDERSON, H. N., WILLIAMS, A., & CURRIE, B. B. (1987) Long-term predictive validity of the WPPSI and the WISC–R with black school children. *Personality and Individual Differences*, 8, 551-559.

MCCARTHY, D. (1972) *Manual for the McCarthy Scales of Children's Abilities*. San Antonio, TX: Psychological Corp.

SATTLER, J. M. (1992) *Assessment of children*. (Rev. & updated 3rd ed.) San Diego, CA: Author.

SNIJDERS, J. TH., TELLEGEN, P. J., & LAROS, J. A. (1989) *Snijders-Oomen Non-verbal Intelligence Test, SON–R 5 1/2-17, manual and research report*. Groningen: Wolters-Noordhoff.

SNIJDERS-OOMEN, N. (1943) *Intelligentieonderzoek van doofstomme kinderen* [The examination of intelligence with deaf-mute children]. Nijmegen: Berkhout.

TELLEGEN, P. (1997) An addition and correction to the Jenkinson, *et al.* Australian SON–R 2 1/2-7 validation study. *Journal of Psychoeducational Assessment*, 15, 67-69.

TELLEGEN, P. J., & LAROS, J. A. (1993a) The construction and validation of a nonverbal test of intelligence: the revision of the Snijders-Oomen Tests. *European Journal of Psychological Assessment*, 9(2), 147-157.

TELLEGEN, P. J., & LAROS, J. A. (1993b) The Snijders-Oomen Nonverbal Intelligence Tests: general intelligence tests or tests for learning potential? In J. H. M. Hamers, K. Sijtsma, & A. J. J. M. Ruijssenaars (Eds.), *Learning potential assessment: theoretical, methodological and practical issues*. Amsterdam: Swets & Zeitlinger.

TELLEGEN, P. J., WINKEL, M., & WIJNBERG-WILLIAMS, B. J. (1994) *Snijders-Oomen Nonverbal Intelligence Test: SON–R 2 1/2-7 Instructions*. Groningen: Psychologisch Instituut Heymans.

TELLEGEN, P. J., WINKEL, M., & WIJNBERG-WILLIAMS, B. J. (1996) *Snijders-Oomen Niet-verbale intelligentietest, SON–R 2 1/2-7, Handleiding* [Snijders-Oomen Nonverbal Intelligence Test, SON–R 2 1/2-7, manual]. Lisse: Swets & Zeitlinger.

THORNDIKE, R. L., HAGEN, E. P., & SATTLER, J. M. (1986) *Guide for administering and scoring the Stanford-Binet Intelligence Scale: Fourth Edition*. Chicago, IL: Riverside.

WECHSLER, D. (1967) *Manual for the Wechsler Preschool and Primary Scale of Intelligence*. New York: Psychological Corp.

WECHSLER, D. (1989) *Manual for the Wechsler Preschool and Primary Scale of Intelligence–Revised*. San Antonio, TX: Psychological Corp.

Accepted February 16, 1997.