WISC–IV Technical Report #2
Psychometric Properties

June 15, 2003
Paul E. Williams, Psy.D. Lawrence G. Weiss, Ph.D. Eric L. Rolfhus, Ph.D.

Overview

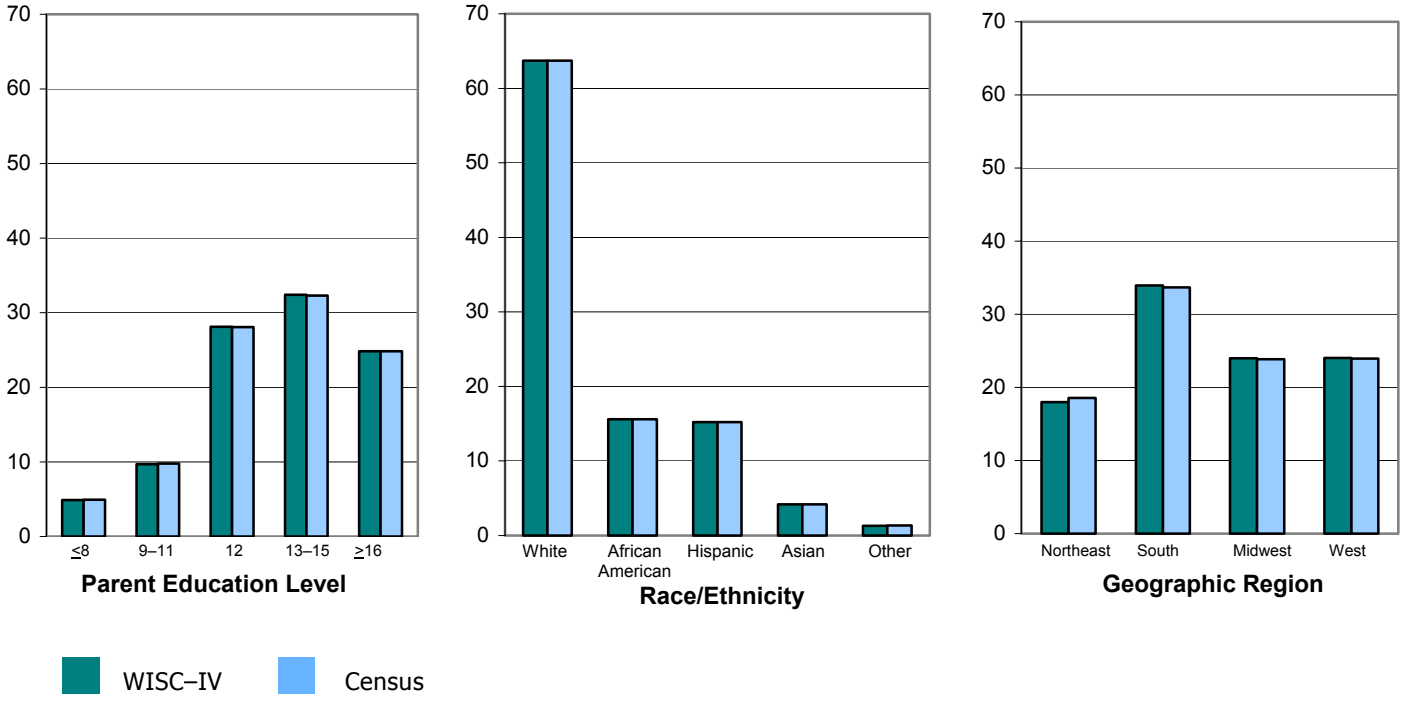
This technical report is the second in a series intended to introduce the Wechsler Intelligence Scale for Children—Fourth Edition (WISC–IV). Technical Report #1 presents the theoretical structure and test blueprint for the WISC–IV, as well as subtest changes from Wechsler Intelligence Scale for Children—Third Edition (WISC–III); Technical Report #2 presents the psychometric properties of WISC–IV; Technical Report #3 addresses the instrument's clinical validity.

Neurocognitive models of information processing provides the basis for the new structure of the WISC–IV, which replaces the traditional Verbal IQ (VIQ)/Performance IQ (PIQ) dichotomy. The index scores that were supplemental in WISC–III are now primary, and each has been enhanced according to contemporary research. The names of two indices have been updated in order to more accurately reflect the content measured by subtests introduced in this revision. Detailed information is provided in the WISC–IV Technical and Interpretive Manual.

WISC–IV Standardization Sample

The WISC–IV standardization sample was representative of the U.S. population of children age 6–16. The stratified random sampling plan utilized the following variables, based on U.S. Bureau of the Census data from the March 2000 census: age, sex, race, parent education level, and geographic region. The standardization sample for the WISC–IV included 2,200 children who were divided by age into eleven groups, each consisting of 200 children. The figure below shows the demographic characteristics of the total sample.

Figure 1. Demographic Characteristics of the Standardization Sample Compared to the U.S. Population



[a] U.S. Population data are from Current Population Survey, March 2000: School Enrollment Supplemental File [CD-ROM] by U.S. Bureau of the Census, 2000, Washington, DC: U.S. Bureau of the Census (Producer/Distributor).

Evidence of Reliability: Internal Consistency

The evidence of internal consistency reliability for the normative sample was obtained using the split-half method for all subtests except the speeded tasks (Coding, Symbol Search, and Cancellation), for which the test-retest coefficients were used. Table 1 compares the average internal consistency reliability coefficients across ages for the WISC–III and the WISC–IV subtest and composite scales. The reliability coefficients for the WISC–IV composite scales range from .88 (Processing Speed) to .97 (Full Scale). The reliability coefficients of the WISC–IV composite scales are identical to or slightly better than corresponding composite scales in the WISC–III.

That these results can be appropriately generalized is supported by information obtained from the special and clinical samples. Evidence of reliability was obtained utilizing the split-half method from a sample of 661 children from 16 special and clinical groups. Table 1 provides average internal consistency reliability coefficients of subtests for these groups. The majority of the subtest reliability coefficients across special groups are similar to or higher than those coefficients reported for the normative sample, suggesting that the WISC–IV is equally reliable instrument for assessing children who are developing typically and children with clinical diagnoses.

Table 1. Reliability Coefficients of the Subtests and Composite Scales

Subtest/Composite	Average WISC–III r_{xx}[a]	Average WISC–IV r_{xx}[a]	WISC–IV Average With Special Groups r_{xx}
Block Design	.87	.86	.90
Similarities	.81	.86	.90
Digit Span	.85	.87	.87
Picture Concepts	–	.82	.88
Coding	.79	.85	–
Vocabulary	.87	.89	.92
Letter-Number Seq.	–	.90	.93
Matrix Reasoning	–	.89	.93
Comprehension	.77	.81	.87
Symbol Search	.76	.79	–
Picture Completion	.77	.84	.89
Cancellation	–	.79	–
Information	.84	.86	.89
Arithmetic	.78	.88	.92
Word Reasoning	–	.80	.85
Block Design No Time Bonus	–	.84	.89
Digit Span Forward	–	.83	.82
Digit Span Backward	–	.80	.84
Cancellation Random	–	.70	–
Cancellation Structured	–	.75	–
Verbal Comprehension	.94	.94	
Perceptual Reasoning	.90	.92	
Working Memory	.87	.92	
Processing Speed	.85	.88	
Full Scale	.96	.97	

[a] Average reliability coefficients were calculated with Fisher's z transformation.

Evidence of Reliability: Test-Retest Stability

The evidence of the WISC–IV's test-retest stability for subtest and composite scales was evaluated with information obtained from a sample of 243 children. Participants were administered the WISC–IV on two separate occasions, with a mean test-retest interval of 32 days. Table 2 presents the mean subtest scaled scores and composite scores and their standard deviations. As can be seen, the data indicate that the WISC–IV scores are stable across time. The mean retest scores for all subtests are higher than the mean test scores from the first administration, with effect sizes ranging from .08 (Comprehension) to .60 (Picture Completion). In general, the test-retest gains are less pronounced for the verbal subtests than for other subtests. This analysis is presented in five separate age bands that show that the pattern of results is similar.

Table 2. Stability Coefficients of the Subtests, Process Scores, and Composite Scales

All Ages	First Testing		Second Testing				
Subtest/Process Score/Composite	Mean	*SD*	Mean	*SD*	r_{12} [b]	Corrected r [a]	Standard Difference
Block Design	10.0	3.0	11.2	2.9	.81	.82	.41
Similarities	10.1	2.6	10.7	2.5	.81	.86	.24
Digit Span	9.9	2.9	10.4	2.7	.81	.83	.18
Picture Concepts	10.1	2.7	10.9	2.8	.71	.76	.29
Coding	10.4	2.7	11.8	3.1	.81	.84	.48
Vocabulary	10.1	2.3	10.4	2.4	.85	.92	.13
Letter–Number Seq.	10.3	2.5	10.7	2.6	.75	.83	.16
Matrix Reasoning	10.2	2.5	10.8	2.7	.77	.85	.23
Comprehension	10.1	2.5	10.3	2.4	.72	.82	.08
Symbol Search	10.4	2.5	11.5	2.8	.68	.80	.41
Picture Completion	10.3	2.9	12.1	3.1	.82	.84	.60
Cancellation	10.2	3.0	11.3	3.0	.78	.79	.37
Information	10.0	2.5	10.4	2.5	.83	.89	.16
Arithmetic	10.1	2.8	10.7	2.5	.75	.79	.23
Word Reasoning	10.2	2.5	11.0	2.6	.75	.82	.31
Block Design No Time Bonus	10.1	2.9	11.2	2.7	.76	.78	.39
Digit Span Forward	9.9	2.9	10.3	2.8	.72	.76	.14
Digit Span Backward	10.1	2.7	10.5	2.7	.67	.74	.15
Cancellation Random	10.0	2.9	11.0	2.9	.68	.72	.34
Cancellation Structured	10.2	2.8	11.0	2.8	.73	.76	.29
Verbal Comprehension	100.0	11.7	102.1	11.7	.89	.93	.18
Perceptual Reasoning	100.7	13.1	105.9	13.9	.85	.89	.39
Working Memory	99.8	13.1	102.4	13.3	.85	.89	.20
Processing Speed	102.4	12.6	109.5	15.2	.79	.86	.51
Full Scale	101.0	11.7	106.6	12.5	.89	.93	.46

[a] Correlations were corrected for the variability of the standardization sample (Allen & Yen, 1979; Magnusson, 1967).
[b] Average stability coefficients across the five age bands were calculated with Fisher's z transformation.

Factor-Analytic Studies

Exploratory and confirmatory factor analyses were conducted to evaluate the internal structure of the WISC–IV. Because the revision retained ten subtests from the WISC–III, as well as including five new subtests designed to measure similar constructs, WISC–IV was expected to measure the same four cognitive domains as the WISC–III (i.e., Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed; Wechsler, 1991).

The initial step in examining the factor structure of the WISC–IV was to determine if the pattern of obtained results matched the hypothesized four factor structure. The stability of the factor structure was then examined across different age groups. Finally, the predicted model was tested against alternative models using confirmatory factor analytic methods.

Table 3. Exploratory Factor Pattern Loadings for Core and Supplemental Subtests

Ages 6:0–16:11 (N=1525)	Four Factor Model			
Subtest	Verbal Comprehension	Perceptual Reasoning	Working Memory	Processing Speed
Similarities	**.71**	.13	.02	-.02
Vocabulary	**.87**	-.05	.06	.00
Comprehension	**.78**	-.13	.06	.07
Information	**.71**	.08	.11	-.06
Word Reasoning	**.73**	.09	-.07	-.01
Block Design	-.06	**.78**	.04	-.02
Picture Concepts	.16	**.40**	.06	.02
Matrix Reasoning	-.03	**.64**	.19	-.04
Picture Completion	.32	**.60**	-.26	.02
Digit Span	.00	-.03	**.67**	-.05
Letter-Number Seq.	.11	-.04	**.62**	.00
Arithmetic	.14	.18	**.51**	.03
Coding	-.02	.01	.05	**.70**
Symbol Search	.01	.17	.08	**.54**
Cancellation	.01	-.09	-.11	**.65**

Note: See WISC–IV Technical and Interpretive Manual for complete table and further discussion.

Exploratory Factor Analysis

Table 3 presents the factor analysis results for the core and supplemental subtests for all ages. The four factor structure is clearly observed and the primary loading of each subtest is found on the expected factor. On the Verbal Comprehension factor, a small secondary loading across ages is observed for Picture Completion. To further examine the stability of the factor structure, the data in four age bands (6–7, 8–9, 10–11, 12–13, 14–16) were then analyzed separately. With minor variations, each age band supported the overall four factor structure. The factor loading for Picture Concepts for ages 6–7 was evenly split between Verbal Comprehension and Perceptual Reasoning. The response processes of younger children on this task may require more verbal mediation then in older children. Above age 11, Arithmetic—clearly a working memory subtest at all age bands—has a small secondary loading on Verbal Comprehension and on Perceptual Reasoning. At age 10 and below, Information exhibited a small secondary loading on the Working Memory factor.

The results from the Confirmatory Factory Analysis of the total sample verified that, as compared to reasonable alternative models, the four-factor model is clearly fits the data best; this finding was also consistent across the four age bands.

Evidence of Validity: Relationships to Other Measures

Evidence of validity was examined by determining the relationship between the WISC–IV and the following measures: WISC–III, WPPSI–III, WAIS–III, WASI, WIAT–II, Children's Memory Scale (CMS), Gifted Rating Scale (GRS), BarOn Emotional Quotient–Inventory: Youth Version (Bar-On EQ–I:YV), and Adaptive Behavior Assessment System–Second Edition (ABAS–II). This report presents details of the WISC–III/WISC–IV study.

Table 4. Correlations Between Subtest and Composite Scores on the WISC–IV and WISC–III

Subtest/Composite	WISC–IV			WISC–III			r_{12} [b]	Corrected r_{12} [b]	Standard Difference [c]
	Mean [a]	*SD*	*N*	Mean [a]	*SD*	*N*			
Block Design	10.6	2.8	242	11.6	3.5	242	.77	.77	.32
Similarities	10.6	2.7	243	11.3	2.9	243	.75	.76	.25
Digit Span	10.4	3.1	241	10.5	3.3	241	.79	.77	.03
Picture Concepts	10.8	3.0	244						
Coding	10.4	3.1	240	11.1	3.3	240	.77	.76	.22
Vocabulary	10.6	2.5	243	10.7	2.9	243	.78	.82	.04
Letter–Number Seq	10.4	3.1	244						
Matrix Reasoning	10.5	2.9	244						
Comprehension	10.6	2.6	241	11.0	3.1	241	.60	.62	.14
Symbol Search	10.6	2.8	237	11.8	3.8	237	.68	.67	.36
Picture Completion	11.1	3.0	244	11.8	2.8	244	.62	.64	.24
Cancellation	9.6	3.2	239						
Information	10.7	2.6	244	11.0	3.1	244	.82	.83	.10
Arithmetic	10.0	2.9	133	9.8	3.2	133	.74	.74	.07
Word Reasoning	10.7	2.6	244						
VCI/VIQ	103.0	12.3	239	105.4	13.8	239	.83	.87	.18
PRI/PIQ	103.9	14.0	242	107.3	14.9	242	.73	.74	.24
WMI/FDI	101.5	15.3	240	103.0	15.9	240	.74	.72	.10
PSI/PSI	102.7	15.1	232	108.2	16.3	232	.81	.81	.35
FSIQ/FSIQ	104.5	14.0	233	107.0	14.4	233	.87	.89	.18
VCI/VCI	102.9	12.3	238	106.0	13.6	238	.85	.88	.24
PRI/POI	103.9	14.0	241	106.9	14.6	241	.70	.72	.21

Note: Correlations were computed separately for each order of administration in counterbalance design and corrected for the variability of the WISC–IV standardization sample (Guilford & Fruchter, 1978).

[a] The values in the Mean columns are the average of the means of the two administration orders.

[b] The weighted average across both administration orders was obtained with Fisher's *z* transformation.

[c] The Standard Difference is the difference of the two test means divided by the square root of the pooled variance, computed using Cohen's (1996) Formula 10.4.

Correlations With the WISC–III

Both the WISC–IV and the WISC–III were administered in counterbalanced order to 244 children from ages 6–16; the test-retest interval was 5 to 42 days. Table 4 presents the means, standard deviations, corrected and uncorrected correlations, and standard differences.

The corrected correlation between the WISC–III VIQ and WISC–IV VCI is .87 and .74 between the WISC–III PIQ and the WISC–IV PRI. The lower correlation between PIQ and PRI reflects important changes made to this composite in WISC–IV. Tasks that were primarily visual and spatial (Object Assembly, Picture Completion, Picture Arrangement, and Coding) were replaced with fluid reasoning tasks (Matrix Reasoning, Picture Concepts), making the PRI a stronger measure of fluid reasoning than the PIQ; for this reason, a moderate correlation was expected. The WISC–III FSIQ and the WISC–IV FSIQ correlate highly (r = .89).

As anticipated, the older WISC–III norms provided slightly inflated estimates for today's children. The overall difference between the WISC–III and WISC–IV FSIQ scores is 2.5 points, with WISC–III scores the higher of the two. Of the WISC–III scores, processing speed tasks showed the most inflation; the least inflation was observed on the working memory tasks. As Table 4 shows, the WISC–III PSI mean is 5.5 points higher then the WISC–IV PSI mean and the WISC–III FDI mean is 1.5 points higher than the WISC–IV WMI mean. These results are consistent with the Flynn Effect (Flynn, 1984, 1987).

WISC–IV in Comparison to WISC–III

Table 5 provides the expected ranges of the WISC–IV Composite scores for selected WISC–III IQ and Index scores. These ranges are relatively narrow near the middle of the IQ score distribution (i.e., 100) and wider at the upper and lower score levels. The WISC–IV and WISC–III PRI and PIQ scores can be expected to differ more than VCI and VIQ due to the changes in the construct measured by PRI as compared to PIQ. Note that these are 95% confidence intervals based on a non-clinical sample; special education and other clinical children may fall outside of these ranges when they are retested.

Practitioners should keep in mind that these are average differences, and that an individual child who has been administered the WISC–III and is retested with the WISC–IV may score more or less than 2.5 points lower on the WISC–IV FSIQ, as compared to his or her previous WISC–III scores. When retesting clinical or special education students, many factors can contribute to score differences (for example, the compound effects of the disorder or disability with increased educational and environmental demands as the child ages).

Table 5. Ranges of Expected WISC–IV Composite Scores for Selected WISC–III Composite

WISC–III IQ Score	WISC–IV Composite Score Range			
	VCI	PRI	FSIQ	
55	50 – 55	47 – 56	49 – 56	
70	65 – 70	63 – 70	65 – 70	
85	81 – 84	79 – 84	81 – 84	
100	97 – 98	95 – 98	96 – 98	
115	112 – 113	110 – 113	112 – 113	
130	126 – 129	125 – 129	126 – 129	
145	140 – 145	139 – 145	140 – 145	
WISC–III Index Score	**WISC–IV Index Score Range**			
	VCI	PRI	WMI	PSI
55	49 – 55	47 – 57	50[a] – 58	50[a] – 54
70	65 – 69	63 – 70	65 – 72	61 – 68
85	81 – 83	80 – 84	92 – 86	77 – 82
100	96 – 98	95 – 98	97 – 100	93 – 96
115	111 – 113	111 – 113	112 – 115	108 – 111
130	126 – 128	125 – 129	126 – 131	123 – 126
145	140 – 144	139 –145	140 – 147	137 – 142

Note: Ranges are 95% confidence intervals based on linear equating of data (Angoff, 1984, Design II.B) for 244 children administered both tests in counterbalanced order.

[a] The range is truncated due to minimum obtainable Index scores.

Summary

This technical report presents some of the basic psychometric properties of the WISC–IV, including information about the demographically-representative standardization sample, internal consistency reliability, test-retest stability, factor structure, and the correlations with and score differences from the WISC–III. The interested reader is referred to the WISC–IV Technical and Interpretive Manual for further information.

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory.* Monterey, CA: Brooks/Cole.

Angoff, W. H. (1984). *Scale, norms, and equivalents scores.* Princeton, NJ: Educational Testing Service.

Cohen, J. (1996). *Explaining Psychological Statistics.* Pacific Grove, CA: Brooks/Cole.

Flynn, J. R.(1984). The Mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*(1), 29–51.

Flynn, J. R.(1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*(2), 171–191.

Guilford, J. P., & Fruchter, B. (1978). *Fundamental statistics in psychology and education* (6th ed.). New York: McGraw-Hill.

Maggusson, D. (1967). *Test theory.* Reading, MA: Addison-Wesley.

Wechsler, D. (1991). *Wechsler intelligence scale for children—Third edition.* San Antonio, TX: The Psychological Corporation.