Psychological Assessment
1995, Vol. 7, No. 1, 115-117

1040-3590/95/$3.00

BRIEF REPORTS

Establishing Covariance Continuity Between the WISC–R and the WISC–III

Wallace E. Dixon, Jr.
Heidelberg College

Timothy Anderson
Vanderbilt University

Much of the research based on Wechsler Intelligence Scale for Children–Revised (WISC–R; D. Wechsler, 1974) subtest covariances may not validly be applied to the Wechsler Intelligence Scale for Children–Third Edition (WISC–III; D. Wechsler, 1991) without ascertaining whether the covariance structures of the normative samples for the two versions are equivalent. The covariance matrices for the WISC–R and the WISC–III were tested for equality by age group using the EQS Structural Equations Program (P. M. Bentler, 1989). Chi-square tests revealed significant differences between the covariance matrices for 10- and 16-year-olds; however, the more stable normed fit index revealed essentially no difference across all age groups. It can be concluded that, for all practical purposes, covariation among WISC–III subtests does not differ from covariation among WISC–R subtests.

Researchers and clinicians interested in applying the vast Wechsler Intelligence Scale for Children–Revised (WISC–R; Wechsler, 1974) literature to investigations and applications incorporating the Wechsler Intelligence Scale for Children–Third Edition (WISC–III; Wechsler, 1991) can benefit by knowing the extent to which the WISC–III overlaps with or differs from the WISC–R. For example, the WISC–III has been found to cluster more people into the lower IQ ranges (Graf & Hinton, 1994) than the WISC–R. Because of the multidimensional nature of the WISC, it is important to evaluate the extent to which the subtests of the WISC–III covary as they did in the WISC–R. With this information, researchers and clinicians interested in relationships among WISC–III subtests could use the WISC–R literature as a foundation and would not be required to start from scratch.

The WISC–III has incorporated both exploratory and comfirmatory factor analytic approaches in the interpretation of subtest scores. The test differs from the WISC–R by the addition of the Symbol Search subtest, the addition and omission of various items from the remaining subtests, and the expansion to a four-factor interpretation. The new factors include Verbal Comprehension, Perceptual Organization, Freedom From Distractibility, and Processing Speed. The first three factors are those originally found by Kaufman (1979) to describe the WISC–R factor structure. Although the factor loadings for the WISC–R and the WISC–III are identical for the Verbal Comprehension and Perceptual Organization factors, the loadings for the Freedom From Distractibility factor are different. The WISC–R Freedom From Distractibility factor was made up of the Arithmetic, Digit Span, and Coding subtests; in the WISC–III version, however, it is made up of only the Arithmetic and Digit Span subtests. The Coding subtest now loads with the Symbol Search subtest on a new Processing Speed factor.

Wechsler (1991) reported that this four-factor structure was superior to all other structures for the standardization sample as a whole, and for each of the 11 age-group subsamples. This may be surprising given the large body of literature arguing in favor of other solutions with respect to the previous version, the Kaufman three-factor solution in particular (see Kaufman, 1979 for discussion; Anderson & Dixon, in press; O'Grady, 1989). However, at least two possibilities may account for the different factor structures for these two WISC versions. It may be that in the WISC–III, relations between the subtests were modified because of the addition and omission of items within the subtests. Consequently, the subtests may no longer covary as they did in the older version. On the other hand, the simple addition of a new subtest may have been sufficient to alter the factor structure, with covariation among the original subtests remaining unchanged.

Because WISC–R scores have been so widely used for such varied purposes as prediction, diagnosis, and outcome, and because the WISC–III is likely to be used similarly, we believe it is essential to evaluate the degree of continuity between subtest covariances. Differences in the covariance structures between the two versions would imply that some findings from the WISC–R could not be extended to the WISC–III. The purpose

Wallace E. Dixon, Jr., Department of Psychology, Heidelberg College; Timothy Anderson, Department of Psychology, Vanderbilt University.

We wish to acknowledge the assistance of Patrick Harris and the helpful comments of Edward Rigdon and Keith Smolkowski during preparation of this article.

Correspondence concerning this article should be addressed to Wallace E. Dixon, Jr., Department of Psychology, Heidelberg College, Tiffin, Ohio 44883. Electronic mail may be sent via the Internet to dixon@nike.heidelberg.edu.

of the present investigation is to compare the subtest covariances of these two versions of the WISC, excluding the Symbol Search subtest. Because covariance continuity may vary as a function of age (O'Grady, 1989), the present investigation evaluates covariance equality at each of the 11 ages to which the WISC applies.

Method

The subtest covariances for both the WISC–R and WISC–III at each of the 11 ages (6–16 years) were obtained directly from the correlation matrices and standard deviations published in the respective manuals (Wechsler, 1974, 1991). Comparisons were made at each age (e.g., 6.5 years in the WISC–R versus 6 years in the WISC–III). The testing for equality of covariance structure was completed using Bentler's (1989) EQS Structural Equations Program. To assess the degree of similarity between the covariance matrices at each age, the chi-square and Bentler-Bonett normed fit index (NFI) were used. These indices were chosen because they are the most popular and are likely the most familiar to the readership.

The chi-square and the NFI indices are used somewhat differently in multisample covariance comparisons than in traditional confirmatory procedures. With single samples, a structural model is specified that is hypothesized to generate a given covariance matrix. The model can be rejected if it inadequately reproduces the covariance matrix, as determined by a significantly large chi-square value, or an NFI that is less than the conventional .90 criterion. In multisample analyses, however, a particular model is applied to data from both samples simultaneously (Bentler, 1989). In this case, the goodness-of-fit chi-square test is used to characterize the adequacy of the model applied simultaneously to the multiple samples. The relevant model in the present investigation is one that (a) specifies that each subtest is uniquely identified by a single factor and (b) permits the factors to covary while (c) constraining subtest covariances to be equivalent across the two versions of the WISC. This model is then simultaneously imposed on the standardization data from both versions of the WISC.

Mulaik et al. (1989) have described a number of additional fit indices, including those that adjust for parsimony and the so-called Type 2 indices, but neither of these types of index is warranted for the present investigation. Adjustment for parsimony would be useful if we were to compare the fit of one model with many degrees of freedom to the fit of a different model with few degrees of freedom (cf. Anderson & Dixon, in press). However, in the present study, all models are equivalent with respect to the degrees of freedom. Likewise, a Type 2 index would be useful if we wished to compare the respective fits of a given model in samples of different sizes. However, in our case, the standardization samples at each age for each test are equivalent ($n = 200$). In our study, then, to adjust for parsimony or to convert to a Type 2 version of the NFI would effectively result in multiplying the obtained NFIs by a constant.

Results

Results are presented in Table 1. Based on the chi-square, with alpha set at $p = .05$, it is apparent that the covariance structures are equivalent between the two tests across all age groups, except those of the 10 and 16 year olds. Differences between test covariance structures are marginal for the 6, 8, 13, and 14-year-old groups. This finding implies that there are significant or marginally significant differences between the subtest covariances for over half of the age groups. However, it is widely recommended that additional indices of fit be used to supplement or supplant interpretations based on the chi-square alone (see Mulaik et al., 1989, for discussion). Bentler and Bonett (1980) noted that, as sample size increases, model rejection increases, even with only trivial differences between the model-generated data and the actual data. Thus, it may be that, even though some of the covariance structures at some ages are not equivalent, these differences may be minimal. To supplement the chi-square results, then, we next considered the corresponding NFIs as a function of age group. Contrary to chi-square results, examination of the NFI makes it clear that the covariance matrices are equivalent, across all ages, in terms of practical similarity. In fact, there is very little difference between the covariance matrices as a function of age: All NFI values are .96 ± .01.

Table 1
Relative Fit Indices of WISC Covariance Comparisons by Age

Age (years)	χ^2	p	NFI
6	81.20	.10	.96
7	80.19	.11	.96
8	83.98	.07	.96
9	71.53	.30	.97
10	91.88	.02	.95
11	60.82	.66	.97
12	79.75	.12	.97
13	81.65	.09	.96
14	81.95	.09	.96
15	78.05	.15	.97
16	89.06	.03	.96

Note. For all comparisons, $df = 66$. WISC = Wechsler Intelligence Scale for Children; NFI = normed fit index.

Discussion

Although the WISC–III covariances may be statistically different from those of the WISC–R on the basis of chi-square analysis, the similarity and magnitude of the NFIs across age lead us to conclude that there are no practical differences. The observed differences are likely trivial, as the chi-square test has been shown to be sensitive to relatively large sample sizes (Bentler & Bonett, 1980; Marsh, Balla, & McDonald, 1988). Thus, we argue that there is structural continuity between the WISC–R and WISC–III subtest covariances. Subtests that covaried in the WISC–R covaried similarly in the WISC–III. The fact that a fourth factor emerged with the WISC–III can be attributed to the addition of the Symbol Search subtest and not to changes in the original subtest covariances. We believe that the extant literature pertaining to the WISC–R as a predictive, diagnostic, and outcome tool can be usefully applied to the WISC–III.

References

Anderson, T., & Dixon, W. E., Jr. (in press). Confirmatory factor analysis of the WISC–R with normal and psychiatric adolescents. *Journal of Research on Adolescence.*

Bentler, P. M. (1989). *EQS: Structural Equations Program Manual.* Los Angeles: BMDP Statistical Software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Graf, M. H., & Hinton, R. N. (1994). A comparison study of WISC–R and WISC–III IQ's. *Educational and Psychological Measurement, 54,* 128–133.

Kaufman, A. S. (1979). *Intelligence testing with the WISC–R.* New York: Wiley.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391–410.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105,* 430–445.

O'Grady, K. E. (1989). Factor structure of the WISC–R. *Multivariate Research, 24,* 177–193.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children–Revised.* New York: Psychological Corporation.

Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children-Third Edition.* San Antonio, TX: Psychological Corporation.

Received April 25, 1994
Revision received August 26, 1994
Accepted August 31, 1994 ■