Journal of Psychoeducational Assessment
1990, *8*, 268-275

GOODNESS OF FIT OF THE WOODCOCK-JOHNSON PSYCHO-EDUCATIONAL BATTERY-REVISED TO THE HORN-CATTELL *Gf-Gc* THEORY

James E. Ysseldyke
University of Minnesota

Empirical evidence is examined on the extent to which the structure of the Woodcock-Johnson Psycho-Educational Battery-Revised is matched to Horn-Cattell *Gf-Gc* Theory. Confirmatory factor analysis is described, and the confirmatory factor analyses used to support the match are examined. It is concluded that the test is a good fit to the model, though some questions and research possibilities are raised.

Most of the tests designed to measure intelligence, scholastic ability, learning aptitude or whatever it might be termed were designed in the absence of theory. Indeed, tests authors sometimes pride themselves in being atheoretical. Intelligence tests are samples of behavior. Interpretation of the performance of individuals on the tests is enhanced when the tests are organized or constructed on the basis of some meaningful theory. When this is not the case, as it most often is not, we have simply an interesting set of results on a set of test items, subtests, scales, clusters, or batteries.

The Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R; Woodcock & Johnson, 1989) clearly is grounded in psychological measurement theory, although the first edition of the test (Woodcock & Johnson, 1977) was not. Rather, it was grounded in a pragmatic decision-making model (McGrew, 1986; Woodcock, 1984). Woodcock is to be commended for having clearly articulated the theory that underlies the test and for having worked so very hard to fit the test to the theory. The WJ-R clearly is based on the Horn-Cattell (Horn, 1986, 1988) *Gf-Gc* theory.

I have been asked in this paper to react briefly to Woodcock's (1990) paper entitled, "Theoretical foundations of the WJ-R measures of cognitive ability." I discuss first the development of the WJ-R Cognitive measures, with special focus on the confirmatory factor analysis procedures used to develop the test. I raise a set of questions about the confirmatory factor analyses that were performed, and a few about the test in general. I then describe a number of possibilities for exciting research with the WJ-R.

I am extremely impressed with the convincing job that Woodcock has done fitting the WJ-R to the Horn-Cattell *Gf-Gc* model of intelligence and with the extent to which he has gone to demonstrate that fit empirically. I am convinced that the test is a good fit to the model. So, the nit picking in which I engage must be understood within a broad, overwhelmingly positive view of the scale and its fit to the theory.

Reprint requests should be addressed to James E. Ysseldyke, Department of Educational Psychology, 350 Elliott Hall, University of Minnesota, Minneapolis, MN 55455.

DEVELOPMENT OF THE WJ-R COGNITIVE MEASURES

The WJ-R Tests of Cognitive Ability is probably the most comprehensive measure of cognitive functioning available, though the DAS may rival it in magnitude of subtests. The WJ-R is a wide age range, comprehensive set of cognitive measures. The 21 Cognitive Tests are organized into seven factor scores (an eighth Quantitative Ability factor is contained in the achievement battery), each based on two measures. Confirmatory Factor Analysis procedures were used to fit the test to the model, or to examine the extent to which the test fits the model. New subtests were added to the original battery (Memory for Names; Delayed Recall: Memory for Names; Delayed Recall: Auditory-Visual Learning; Memory for Words; Cross Out; Incomplete Words; Sound Patterns; Visual Closure; Picture Recognition; and Listening Comprehension), and subtests included in the earlier battery were revised (e.g., Verbal Analogies; Spatial Relations; Visual Matching), in an effort to produce "clean" measures of eight factors. The factors measured are: Long-Term Retrieval; Short-term Memory; Processing Speed; Auditory Processing; Visual Processing; Comprehension-Knowledge; Fluid Reasoning; and Quantitative Ability.

Confirmatory factor analysis is a procedure in which an investigator specifies in advance the subtests expected to load significantly on factors, the correlations that are present (oblique) or absent (orthogonal) between factors, and whether any subtest residuals (error variance and uniqueness) are correlated. The subtests are given to a sample of individuals, and Linear Structural Relations (LISREL) procedures then are used to investigate the extent to which the data support the proposed model. At the risk of possibly offending those who know these procedures well (often called LISREL-ites) a simplified conceptual explanation is offered. This approach is illustrated in Figure 1, which shows that subtests A, B, and D load on Factor 1, subtests C, E, and H on factor 2, and subtests F, G, and I on Factor 3. The absence of arrows from factors to certain tests (e.g., A, B, D do not load on Factor 2) indicates zero-order loadings for the subtests on the factor. Also shown is the fact that Factors 1 and 2 are related/correlated. (Note: the subtest residuals arrows have been omitted from the Figure.)

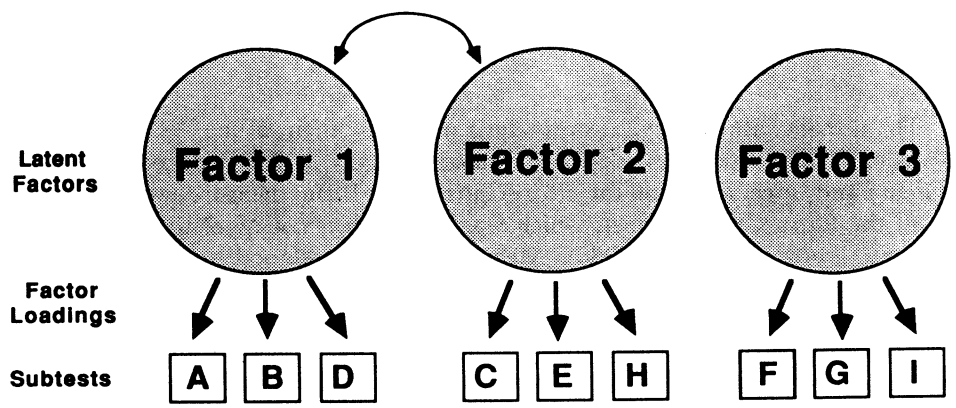

FIGURE 1. Confirmatory Factor Analysis

Confirmatory factor analysis procedures differ from those used in exploratory factor analysis. In exploratory factor analysis, an investigator gives a large number of tests to a large number of individuals and then engages in an exploratory search for interrelationships among subtests, among factors, and between subtests and factors. The structure of the test (factors) emerges from the the data rather than being specified *a priori* by the researcher. This approach is illustrated in Figure 2. In exploratory solutions, all tests load on all factors, although the rotation to simple structure results in certain tests that have strong loadings (the dark arrows), while others have near zero loadings (the lighter lines). This differs from confirmatory factoring, in which actual zero loadings are prescribed (the absence of paths in Figure 1). Also shown is the fact that the factors are correlated (oblique).

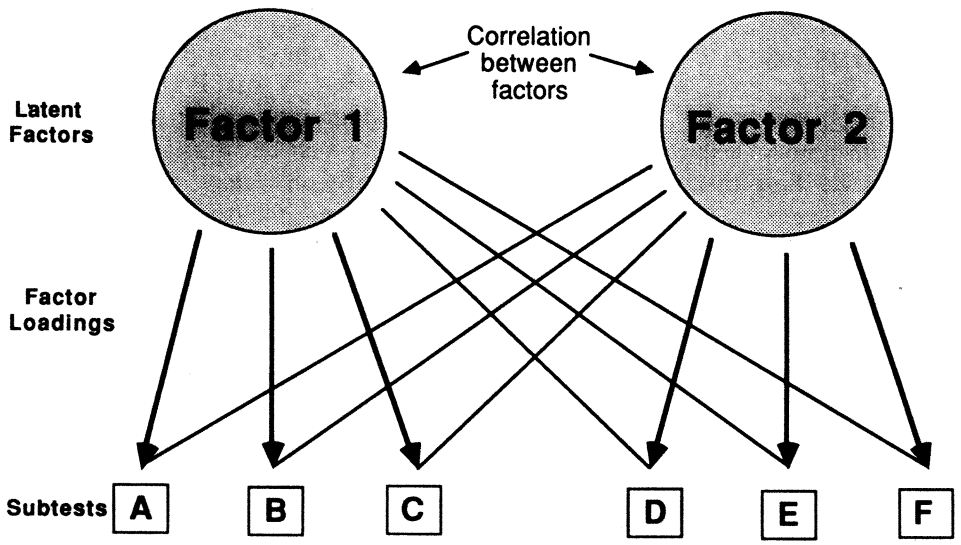

FIGURE 2. Exploratory Factor Analysis

The use of confirmatory factor analysis in development of the WJ-R (McGrew, Werder, & Woodcock, in press) is illustrated in Figure 3. The test is based on *Gf-Gc* Theory and, specifically, on the Horn-Cattell model, and measures were developed to fit the Horn-Cattell model, which was based on a review of the extant factor analytic literature (Horn, 1988). Measures were developed for the WJ-R, and preliminary confirmatory and exploratory analyses were conducted. Subtests were revised, and the tests were administered to the 6,379 individuals whc participated in the standardization. The first model fit was tested with different combinations of tests and with different samples. Woodcock lists the various fit analyses conducted. Analyses were conducted for 16-variable, 27-variable, and 29-variable model fits, and analyses were conducted on samples that included the Stanford-Binet IV, K-ABC, and WISC-R.

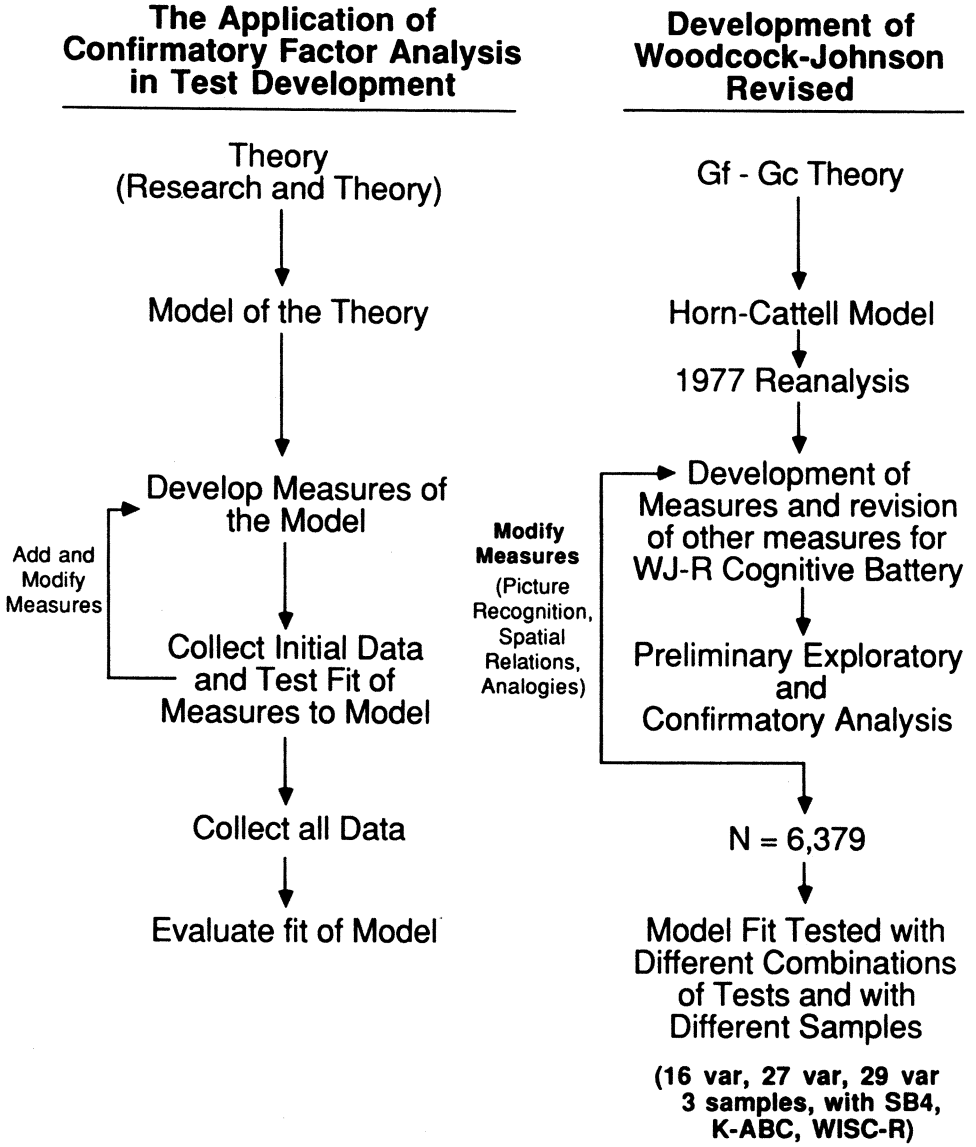

FIGURE 3. The Use of Confirmatory Factor Analysis in Development of the Woodcock-Johnson Revised

There is good evidence for the fit of the WJ-R to the Horn-Cattell *Gf-Gc* model. The author has done a fine job of specifying the model, and it appears that the fit of the model is better for school age than for older individuals, specifically, for older adults. The evidence of a fit suggests simply, however, that the model is plausible, not that it is correct. However, comparisons to other common conceptions of intelligence (e.g., single *g* factor, Verbal/Nonverbal) found the WJ-R Horn-Cattell *Gf-Gc* model to be a much better fit (McGrew et al., in press.)

QUESTIONS

Educators and psychologists long have argued about the merits and limitations of factor analysis in developing models of mental ability and, more specifically, in developing measures of mental ability. Factor analysts disagree on approaches that ought to be used in test construction, and many who oppose factor analytic procedures argue that individuals simply select approaches that will enable them to make the kind of sense they want to make out of their tests/models.

Carroll (1983) addresses the use of factor analytic procedures, stating:

> If factor analysis is given its due and used 'correctly' and appropriately, researchers can arrive at a reasonable and confirmable picture of the structure of mental abilities (p. 9).

Interestingly, Carroll indicates in this same article that the one approach to understanding human abilities that seems to make the most sense is the Horn-Cattell *Gf-Gc* approach. A question that I initially asked myself as I undertook this assignment was whether the use of confirmatory factor analysis was appropriate. I believe it is. It clearly is to be preferred over a causal armchair analysis. I also believe that the Horn-Cattell *Gf-Gc* theory provides the most salient and promising model of intelligence and the one around which cognitive measures ought to be developed.

A second question that I raised concerned the number of tests that were used as markers for factors. Woodcock (1990) states that there must be a sufficient number (usually three or more) of reasonably clean measures, or markers, for each of the factors present so that factor can be clearly identified. The WJ-R certainly contains a large number of subtests. I do not know whether it includes enough subtests to serve as markers for eight factors. Carroll (1983) addresses this issue and specifies that two or more subtests per factor are necessary.

> Finding the 'correct' number of primary factors for a set of data is one of the more difficult problems in factor analysis . . . The ultimate criterion is probably to be found in the pattern of loadings in a rotated solution. Normally, the number of factors can be taken to be the largest number of factors such that there are at least two dominant or salient loadings on each factor in a rotated solution (p. 8).

Woodcock also indicates that confirmatory factor analyses of other intellectual batteries are limited when the variables in those studies have routinely been restricted to those subtests included within the battery itself. In three of the confirmatory factor analyses Woodcock has described in his paper, analysis 1A was an analysis of the fit of 16 subtests to eight factors. Analysis 1B consisted of a fit of 27 subtests (6 from the achievement battery) to the eight factors, while analysis 1C consisted of an analysis of the fit of 29 subtests to the model. In the first investigation, there are only 2 subtests per factor. In all three analyses, the measures are drawn from a single instrument, unless the achievement battery is considered a separate test. I do not know whether there are enough subtests per factor for the 16-variable analysis, and I question the limitation of input variables to just the WJ-R subtests (as Woodcock questions the same thing with other batteries).

Woodcock (1990) does report three investigations in which extra measures were used. These investigations, however, had very small *N*s for confirmatory factor analytic investigations. In addition, Woodcock does not report the fit statistics for these analyses.

A question I have with regard to the structure and use of the WJ-R is not restricted to the WJ-R, and I have raised it in connection with batteries (e.g., Differential Ability Scales and K-ABC). The issue concerns the treatment relevance or relevance to instructional intervention of the test, its subtests, and clusters. I kept asking myself where the evidence was that knowledge of an individual's performance on the many subtests and clusters was important to planning instruction for that individual. I'd feel better about the structure of the WJ-R if I knew that it would provide me with more than heuristic information about the structure of a student's abilities; that it was going to provide me with information I could use to design an appropriate intervention. Although this is a problem with all current tests of intelligence, the WJ-R may help with research in this area because of its strong construct validity. That is, better aptitude-treatment interaction (ATI) research might now be possible with the WJ-R. (The research would be less confounded by poor measures that do not measure adequately the major components of ability.)

The last question that I feel compelled to raise concerns the use and interpretation of discrepancy scores obtained from the WJ-R. This measure is constructed in such a way that one can identify ability achievement discrepancies and intra-achievement and intra-cognitive discrepancies. On the positive side, these scores are probably the most psychometrically sound discrepancy scores available for psychoeducational assessment (i.e., they are actual discrepancy norms that include regression effects and are based on co-normed measures). However, given the overwhelming criticisms of the use of discrepancy scores in the literature, and the widespread use of such scores in the identification of students as learning disabled (in the absence of wide agreement on the concept of LD), I hope the author has described precisely and in a defensible manner the use of such discrepancy analyses. But, alas, this is not the topic of this paper, nor is the use of discrepancy scores specific just to the WJ-R.

POSSIBILITIES

I am excited about a number of possibilities for use of the WJ-R in empirical investigations of important issues in psychology, education, and, specifically, in special education. Given the evidence that the WJ-R subtests provide relatively clean measures of specific factors, there is now an opportunity to use this test in ATI investigations. Such investigations have been plagued in the past by poor measures of both aptitudes and achievement. Given the encouraging findings reported by Woodcock, we may now be able to investigate the extent to which knowledge of pupil performance on the various factors is prescriptively predictive of relative success in school. That is, we may now begin to address treatment relevance.

Second, we now have a wonderful opportunity to engage in research on the developmental characteristics of the factors measured by the WJ-R. If the WJ-R fits the Horn-Cattell *Gf-Gc* model in the way in which it appears to do, then we should be able to see measures of fluid abilities peaking or reaching an asymptote at an earlier age than measures of crystallized abilities; and we may begin to examine correlates of declines in fluid intelligence in later years. Longitudinal research, with a common set of measures, is now possible with the publication of the WJ-R.

Third, we now have a great opportunity to investigate the extent to which status on a particular factor is modifiable through training. It would be anticipated that performance on measures of *Gc* are more modifiable than performance on measures of *Gf*.

Finally, we have an opportunity to use this measure in investigations of differential performance by different groups. McGrew et al. (in press) present some encouraging discriminant function analyses, which show how the WJ-R can be used to discriminate between different populations (normal, LD, gifted, MR). Students who are handicapped sometimes have acculturations that differ from those of other students. Those who are disadvantaged have acculturations that differ from those of others. There are theoretical predictions that can be made with regard to the test performance of individuals who differ in acculturation. Newland (1980), who espouses a variation of the Horn-Cattell model, has specified some of these.

IN CLOSING

The WJ-R represents a significant milestone in the applied measurement of intellectual abilities. The manner in which theory and prior research were used in the development of the WJ-R and in the analysis of the other major measures of intelligence should serve as a model for future research and development in the fields of applied psychometrics and psychoeducational assessment. Based on the information provided by Woodcock (1990) and his colleagues (McGrew et al., in press), it appears that the WJ-R adequately represents many of the major components of human ability as suggested by the extant factor analytic research in intelligence. However, in the final analysis, the value of the WJ-R rests on the degree to which it contributes to resolving many of the significant issues in the fields of psychology and special education (e.g., ATI research, identification of students with learning disabilities, to name a few). Past research in these areas with other measures has been less than encouraging. The WJ-R will assist in resolving many of these complex issues. However, for once research in these areas may have a fighting chance because we now have a collection of satisfactory measures with which to begin to investigate more accurately these most important issues.

REFERENCES

Carroll, J. B. (1983). Studying individual differences in cognitive abilities: Through and beyond factor analysis. In R. F. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition* (Vol. 1, pp. 1-32). New York: Academic Press.

Horn, J. L. (1980). Thinking about human abilities. In J. R. Nesselroade (Ed.), *Handbook of multivariate psychology* (pp. 645-856). New York: Academic Press.

Horn, J. L. (1986). Some thoughts about intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definition* (pp. 91-96). Norwood, NJ: Ablex.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology, Second edition.* New York: Plenum Press.

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (in press). *WJ-R technical manual.* Allen, TX: DLM.

Newland, T. E. (1980). Psychological assessment of children and youth. In W. Cruickshank (Ed.), *Psychology of exceptional children and youth* (pp. 74-135). Englewood Cliffs, NJ: Prentice Hall.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8*, 231-258.

Woodcock, R. W., & Johnson, M. (1977). *Woodcock-Johnson Psychoeducational Battery.* Allen, TX: DLM.

Woodcock, R. W., & Johnson, M. (1989). *Woodcock-Johnson Psychoeducational Battery-Revised.* Allen, TX: DLM.