

Handbook of Understanding and Measuring Intelligence

2005

Edited by

OLIVER WILHELM • RANDALL W. ENGLE

Humboldt-University, Berlin, Germany

Georgia Institute of Technology

ASSESSING INTELLIGENCE

p. 333 - 360

Past, Present, and Future

RICHARD D. ROBERTS

PIPPA M. MARKHAM

GERALD MATTHEWS

MOSHE ZEIDNER

the scales have a mean of 100 (standard deviation [SD] = 15), with test taker's performance compared with scores earned by individuals in his or her own respective age group.

Space precludes detailed consideration of all three Wechsler scales, so in the interests of brevity and consistency (i.e., the group test discussed later in this chapter is for adult populations), we focus the rest of this review on the WAIS-III. Notably, earlier versions of this test have been claimed to be the standard against which other adult tests can be compared (e.g., Gregory, 1999; Lyman, 1972). Table 19.2 lists each of the subtests, along with capsule descriptions, and each of the composites that they form. In this revision, the well-known Verbal, Performance, and Full-Scale IQ scales have been retained. In addition, responding to emerging data from differential psychology, cognitive science, and their integration, an attempt has been made to incorporate constructs from these disparate subdisciplines by providing so-called Index scores, which reflect the purportedly theoretically meaningful cognitive constructs of

PK

WALSLEY

P. 335

Perceptual Organization, Verbal Comprehension, Working Memory, and Processing Speed (Wechsler, 1997). Notably, to ensure theoretical, political, and sociological relevance (in the process, likely catering to market concerns), the test includes a measure of (pure) fluid intelligence for the very first time (i.e., Matrices), whereas individual items have addressed the needs of minority groups and the aged by being constructed in larger font, colorized, or else more sensitive in depicting "significant others" in various item stems.

T. 11. 10. 2 provides internal consistency and

enterprise of intelligence assessment per se.

In defense of the technical merits of the WAIS-III (especially, the construct validity) might be invoked commonality between this instrument and its ancestor, the WAIS-R. For example, conceptual comparisons of item content and statistics indicate much in common between the instruments. Moreover, correlations between overlapping IQ scores tend to be high

p 340

(rs, based on a reasonable sample size of 40, range between .82 and .85). However, this defense of the WAIS-III rests, in part, on demonstrating that the WAIS-R is relatively impervious to meaningful scientific criticism. However, consider the following two quotes, emanating from different research disciplines, separated by a meaningful time period (i.e., one decade), and based on very different forms of analysis:

The Wechsler tests are like the dinosaur, too large, cumbersome and ill-fitted and awkward in the age in which they developed, unable to remain viable in a psychometric age which has passed it by in conceptualization. As with the dinosaur it is time for the Wechsler test to become extinct. (Frank, 1983, p. 126)

Although Frank's views may be somewhat extreme, I would regard them as essentially correct.

Presently available technology would permit the development of tests and scales that would be much more adequate for their purpose than the Wechsler scales. (Carroll, 1993, p. 702) β 34)

Table 19.4 shows how subtests of the WAIS-III correspond to Gf-Gc theory and how they map onto both the IQ scale and Index scores provided by the test publishers. Noting frequent anomalies, the reader can make a determination for herself or himself whether it is likely that changes to the most recent revision of the Wechsler scale have the type of criticisms raised in the preceding passage. Moreover, the construct of Perceptual Organization finds no ready evidence in the psychometric literature, nor indeed is the construct of Working Memory identified as an independent construct in Carroll's (1993) analysis (see p. 303). Further still, as discussed in the future of intelligence

1
testing, it is an open, empirical question whether paper-and-pencil versions of a construct purportedly measuring "Processing Speed" has efficacy, though it may serve as a rough approximation of constructs that clearly deserve greater attention in theoretical discourse and attempts to model intelligence quantitatively (see Chapters 3 and 8, this volume). In sum, if the WAIS-III is a yardstick by which other contemporary individual tests of intelligence should be evaluated meaningfully, available standards for doing so have been set lower than rigorous, unbiased scientific analyses would seemingly dictate.

A Brief Comment on Other Individual

1993; Roberts et al., 2000), it might be concluded that the Stanford-Binet-IV is a sounder instrument, certainly from a theoretical perspective (despite the dominance of the WAIS-III in clinical applications). Even so, in each test, far too many broad abilities of plausibly important significance (e.g., auditory reception [Ga], especially as it pertains to reading competencies [see McGrew & Flanagan, 1999]) are simply not represented.

A noteworthy exception to the preceding near status quo in individual intelligence testing is the Woodcock-Johnson Psychoeducational Battery, which has recently gone through its third revision (WJ-III; Woodcock, McGrew, & Mather, 2000). Consisting of two conormed batteries—Tests of Achievement and Tests of Cognitive Abilities—the test has been designed for use with persons as young as 2 and as old as 90+ years of age (Woodcock et al., 2000). Importantly, the theoretical basis of the Tests of Cognitive Abilities, in particular, is Gf-Gc

theory (Woodcock, 1990). Table 19.6 provides the subtests (which may be divided into a standard [i.e., the first 10 numbered subtests] and extended battery), along with the broad and primary abilities assessed. Each subtest has high levels of reliability, with the standardization sample ($N = 8,818$) large and representative (Cohen & Swerdlik, 2002). It is our contention that the battery is one that both researchers and practitioners should consider using more often, though even one of the test authors has acknowledged the need to supplement these tests with others for comprehensive intelligence test assessment (McGrew & Flanagan, 1999). Note also that the WJ-III may actually oversample fluid (rather than crystallized) intelligence concepts, given that some of its supposed markers of SAR (e.g., Auditory Working Memory) and, arguably, TSR are thought to be highly related to reasoning at the first order and fluid intelligence at the second order of broad cognitive abilities.