visited 9/6/2019

NIH Public Access
Author Manuscript
Intelligence. Author manuscript; available in PMC 2011 March 1.

Published in final edited form as:
Intelligence. 2010 March ; 38(2): 242–248. doi:10.1016/j.intell.2009.12.005.

Stanford-Binet & WAIS IQ Differences and Their Implications for Adults with Intellectual Disability (aka Mental Retardation)

Wayne Silverman,
Department of Behavioral Psychology, Kennedy Krieger Institute, and Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine

Charles Miezejeski,
Department of Psychology, New York State Institute for Basic Research in Developmental Disabilities

Robert Ryan,
Department of Psychology, New York State Institute for Basic Research in Developmental Disabilities

Warren Zigman,
Department of Psychology, New York State Institute for Basic Research in Developmental Disabilities

Sharon Krinsky-McHale, and
Department of Psychology, New York State Institute for Basic Research in Developmental Disabilities

Tiina Urv
Department of Psychology, Eunice Kennedy Shriver Center, University of Massachusetts Medical School

Abstract

Stanford-Binet and Wechsler Adult Intelligence Scale (WAIS) IQs were compared for a group of 74 adults with intellectual disability (ID). In every case, WAIS Full Scale IQ was higher than the Stanford-Binet Composite IQ, with a mean difference of 16.7 points. These differences did not appear to be due to the lower minimum possible score for the Stanford-Binet. Additional comparisons with other measures suggested that the WAIS might systematically underestimate severity of intellectual impairment. Implications of these findings are discussed regarding determination of disability status, estimating prevalence of ID, assessing dementia and aging-related cognitive declines, and diagnosis of ID in forensic cases involving a possible death penalty.

Current definitions of intellectual disability (ID, mental retardation in previous terminology) have evolved from their predecessors to include three key features: (1) significantly subaverage intellectual functioning (IQ of 70 or below, allowing for measurement error), (b)

Correspondence to: Wayne Silverman, Kennedy Krieger Institute, 707 North Broadway, Suite 222s, Baltimore, MD 21205. silvermanw@kennedykrieger.org.
Charles Miezejeski is now in private practice, Staten Island, New York; Tiina Urv is now at the Intellectual and Developmental Disabilities Branch, Eunice Kennedy Shriver National Institute on Child Health and Human Development, National Institutes of Health, Bethesda, Maryland.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

visited 9/6/2019

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

substantial impairments in adaptive functioning, and (c) onset prior to adulthood (e.g., American Association on Mental Retardation, 2002; American Psychiatric Association, 2000). While considerable debate has focused on characterization of adaptive deficits and associated criteria defining "substantial" impairment (e.g., Clausen, 1972; Reshly, Myers & Hartel, 2002; Widaman & Siperstein, in press), there has been a longstanding consensus regarding best practice for the assessment of intellectual impairment. The gold standard has been a broadly focused and individually administered IQ test that provides a comprehensive measure of overall intelligence (often with other scores reflecting more focused domains of cognitive competency). Historically and currently, the two instruments most widely used in the United States have been the Stanford-Binet (e.g., Roid, 2003) and Wechsler Intelligence Scales (e.g., Wechsler, 2008), although other options are certainly available (see Reshly et al., Chapter 3).

While Flynn (1984; 1985) noted that populations improve over time in performance, suggesting needs for periodic restandardization, the validity of these IQ tests when properly administered has been and continues to be broadly accepted (Harrison, Kaufman, Hickman & Kaufman, 1988.) In practice, both the Stanford-Binet and one of the Wechsler tests are often used for assessing intelligence of children and young adolescents. However, because earlier versions of the Stanford-Binet were not normed for older ages, the various editions of the Wechsler Adult Intelligence Scale (WAIS) became the dominant IQ assessment for adults (see Baroff, 2003; Harrison, et al, 1988). In principle, this fact should have little clinical importance, and results from Stanford-Binet, Wechsler, and other IQ tests with comparable scaling and psychometric properties have been assumed to be largely equivalent as long as floor and ceiling effects are avoided.

Should this assumption of equivalency in IQs be compromised, and should substantive differences exist between these assessment instruments, it could have important implications, especially regarding their use in high stakes contexts (e.g., disability determination and forensics). Of course, no psychological assessment provides an absolutely perfect reflection of the underlying construct(s) it seeks to measure, in this case intelligence, but if two gold-standard procedures consistently produce divergent results, then the validity of test-based inferences would have to be questioned.

This very concern has been raised for adults with ID in the past. Bensberg & Sloan (1950) found that WAIS IQs of a group of adults with ID were higher than their Stanford-Binet IQs by from 7 to 20 points, the magnitude of difference increasing with age at test administration. Brengelmann & Kenny (1961) reported an advantage of 8 points for the WAIS compared to the Stanford-Binet, Spitz (1986) an advantage of 12.4 points, and Nelson & Dacey (1999) an advantage of 14.6 points. In fact, the Stanford-Binet Fourth Edition Technical Manual described a 9.3 point advantage for the WAIS for adults with ID (Thorndike, Hagen & Sattler, 1986). While the data have been quite consistent, interpretations of their implications have varied. On one extreme, Spitz (1986) concluded his paper with a plea to the field to address the divergence in findings or "abrogate our responsibilities." In contrast, Thorndike et al. were content with attributing the apparent WAIS advantage to the fact that the Stanford-Binet allowed for an expanded range of scores in the lower tail of the performance distribution, and this seems to have become the generally accepted understanding to date.

In previous eras, differences between WAIS and Stanford-Binet IQ test results would have been predominantly of academic interest. The presence of ID was almost always established during childhood or adolescence, and the diagnosis would have followed the individual into adulthood given its accepted status as a lifelong impairment. However, more recent conceptualizations of ID have changed this substantially. ID is no longer defined as a static

visited 9/6/2019

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

characteristic of an affected individual (e.g., American Association on Mental Retardation, 2002), and diagnosis in adulthood needs to be established (or reestablished) based upon evaluations at the time relevant issues are addressed. This dramatically increases the significance of testing during adulthood. Should different testing instruments provide systematically different outcomes, the nature of the differences needs to be carefully characterized, potential implications need to be broadly appreciated, and consensus needs to be developed for valid interpretation of scores. The current analyses were conducted to examine these issues with respect to the Stanford-Binet and WAIS within the adult population with ID.

Method

Comprehensive clinical record reviews were conducted for 542 adults with ID participating in a larger, multidisciplinary study focused on aging and dementia in adults with ID (see Silverman et al., 2004). All procedures, including the collection of test results from clinical records, were reviewed and approved by appropriate Institutional Review Boards (IRBs) to ensure protection of research participants. Informed consent as well as assent (when consent was provided by a correspondent) was obtained in every case prior to data collection.

Clinical record reviews were standardized employing a 65-page format that recorded all information on demographics, current health status (including all illnesses, medications and lab results), health history, sensory status, and developmental history. This initial chart review was supplemented using an abridged format (31 pages) during follow-up assessments occurring at 14 – 18 month intervals (the number of which varied with length of participant enrollment). Chart reviews included the results of 1,851 psychological assessments explicitly relevant to a diagnosis of ID. These tests had been conducted over the course of many years and provided data for the present analyses. Information recorded included test name, score(s), and date of administration.

An initial inspection indicated that at least one Stanford-Binet or WAIS IQ score was available for 225 individuals. These tests were administered between 1949 and 2005, and covered a wide range of ages at assessment (5 – 81 years of age). All of the various test editions were represented in this sampling except for the Stanford-Binet Fifth Edition and WAIS-IV, and specific format could have varied within as well as between individuals. Seventy-four individuals were tested with both a Stanford-Binet and WAIS on one or more occasions, and these 74 adults were the primary focus for the current analyses. In all cases, knowledge of IQ test results was unavailable at the time of enrollment and inclusion for the present analyses was based simply on score presence without any further consideration.

Many other test results were also recorded, the most frequent other than a WAIS or Stanford-Binet being the Leiter International Performance Scale (Levine, 1993), the Slosson Intelligence Test (Nicolson & Hibpshman, 1991), and the Vineland Adaptive Behavior Scales (Sparrow, Balla & Cicchatti, 1984). All recorded assessment data were entered into a Microsoft Access database, and these data were merged with descriptions of individuals' gender, birthdate, and Down syndrome status (present/absent) for analyses using Statistica 8.0 (Statsoft, 2008). Duplicate entries were possible, given that multiple chart reviews were conducted at 14 – 18 month intervals. These were removed prior to analysis based upon test name, date and score. Because multiple results were sometimes available, median IQs were calculated for Stanford-Binet and/or WAIS IQs for each individual to consider all available data in generating an estimate of his or her true IQ. (Regarding Stanford-Binet assessments, 42 people had one score available, 14 had two, 11 had three, and seven had more than three. For the WAIS, 41 had one score, 25 had two, four had three, and four had more than three.) Median composite scores for Leiter, Slosson, WISC and Vineland Scales, when available,

visited 9/6/2019

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

were also calculated. (Unfortunately age-equivalent Vineland scores were often recorded without a composite, and only composite scores were included in the present dataset).

Results

Demographic characteristics of the 74 individuals assessed with both Stanford-Binet and WAIS IQ tests are summarized in Table 1. As the table indicates, women were over sampled, as were adults with Down syndrome. (This was a consequence of the priority interests of our broader program, which included studies of aging effects on women's health and the association between Down syndrome and Alzheimer's disease.) Therefore, an initial mixed model analysis of variance was conducted to determine if etiology or sex had differential effects on profiles of IQ scores. Etiology (Down syndrome – Yes/No) and Sex were between subjects variables and Test (WAIS/Stanford-Binet) was the within subjects variable. IQs of males and females were comparable, F (1, 70) = 1.7, p >.1, but adults with Down syndrome had lower IQs than those with ID due to other causes, F (1, 70) = 23.6, p <. 0001. (This latter finding was also a consequence of inclusion criteria for our broader program of research, which excluded adults without Down syndrome with relatively severe ID.) A small but significant Sex by Test interaction, F (1, 70) = 4.02, p < .05, reflected a four point advantage for women on the Stanford-Binet contrasted with a one point advantage for men on the WAIS, while a possible Etiology by Test interaction, F (1, 70) = 3.52, p < . 07, reflected a small relative disadvantage for adults with Down syndrome for the Stanford-Binet. In both cases, these effect sizes were quite small (partial $\eta^2 < 0.06$) and had only marginal effects on the substantial differences observed between the two tests described next (partial $\eta^2 = 0.882$).

Analyses of primary interest focused on contrasts between the WAIS and Stanford-Binet IQs within this group of 74 individuals. Figure 1 illustrates these findings as a scatterplot, with median Stanford-Binet IQ values along the x-axis, median WAIS IQ values along the y-axis, and a diagonal plotted to designate equality. Clearly, WAIS IQ was higher for all individuals, exact probability < 10^{-22}, and the mean difference of 16.7 points was significant, t (73) = 24.4, p < .00001. Interestingly, there is a strong linear correlation between the two measures despite the systematic differences, r (72) = .818, p < .000001, suggesting that, as expected, scores on the two scales reflect the same underlying construct(s).

To examine results in somewhat greater detail, distributions of differences were generated (Table 2). Because adults with Down syndrome have been over-sampled within our broader program, and because Down syndrome is associated with a distinct cognitive phenotype (see Silverman, 2007), findings were examined separately for individuals with and without Down syndrome. As indicated in Table 2, etiology had an insignificant effect on the findings of primary interest, χ^2 (4) = 6.6, p > 0.1. Differences of more than 10 points were present for 85.1% of the group and more than 20 points for 24.3%, with a range of from 4 to 31 points.

Several potential confounds might have contributed to the striking and systematic divergence between Stanford-Binet and WAIS scores, and these will be addressed in turn. One likely factor is the difference in the minimum scores provided by the two scales using routine scoring, the Stanford-Binet extending to a lower bound (e.g., Thorndike et al, 1986). To address this possible confound, the analysis was repeated only for individuals receiving a WAIS-IQ of at least: (a) 55, t (44) = 22.54, p < .00001; (b) 60, t (28) = 18.05, p < .00001; or even (c) 65, t (18) = 13.68, p < .00002. For these higher IQ subsamples, the mean difference between WAIS and Stanford-Binet IQs ranged from 19.0 to 19.8 points.

visited 9/6/2019

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Another possible confound could have been due to differential practice, and this was addressed in two ways. First, the magnitude of practice effects was estimated separately for the two instruments by comparing the first available IQ score of each individual to his or her second IQ score (for individuals tested more than once). Considering all available records, 75 individuals were tested more than once with a Stanford-Binet, and mean scores differed by less than one point between their first and second testing, t (74) = 0.38, $p > 0.7$. A similarly nonsignificant practice effect of less than one point was found for the 32 individuals with at least two Stanford-Binet IQs within the subsample of 74 described above, t (31) = 0.89, $p > 0.3$. For the WAIS, the improvement from the first to second testing was one point for all available data, t (95) = 1.4, $p > .15$, and 1.5 points when only the subset of 74 cases was considered, t (32) = 1.1, $p > .29$. Next, each of the 74 individual's first Stanford-Binet score (rather than median to provide a clearly defined date of testing) was compared to their first WAIS result to determine if testing order had an effect. A mixed model analysis of variance with Order as a between subjects variable and Test as a within subjects variable showed that receiving the WAIS first produced an insignificant overall advantage of just under 3 points, $F(1,70) = 1.8$, $p < 0.18$, while the Test by Order interaction failed to approach significance, $F(1, 70) < 0.1$. A 16.3 point difference between the two IQ assessments was again significant, $F(1,70) = 391.8$, $p < .00001$, and, as was the case for median IQs, every individual scored higher on the WAIS.

A third possibility has been raised by Flynn's (1984; 1985) observation that populations tend to improve by approximately 0.3 IQ points per year referenced to a fixed format and standardization sample. If WAIS testing was conducted using editions that were many years older relative to the Stanford-Binet, then that might contribute to the observed WAIS advantage. Of course, the effects described here are far larger than any "Flynn effect" could have caused, but the possibility of a partial contribution merits investigation. Therefore, the interval between assessment date of the first Stanford-Binet and WAIS assessment for each individual and publication date was calculated in years. Following Flynn, this value was multiplied by 0.3 and the obtained value was subtracted from the respective IQ to generate an "adjusted" score. All individuals but two continued to have higher scores for the WAIS compared to the Stanford-Binet, $p < 10^{-20}$, and the 15.6 point mean difference between IQs was significant, t (73) = 17.83, $p < .00001$. (For the two cases now failing to show a WAIS advantage, one person had the same adjusted score for both IQs and the other had a two point advantage for the Stanford-Binet.)

These findings represent clear and compelling evidence of substantive differences between Stanford-Binet and WAIS IQs for the adult population with ID, but they do not suggest which of the two scales provides the more valid score. To address this issue, scores on the Stanford-Binet and the WAIS were compared to other assessments. To increase the number of cases considered in these analyses, scores from individuals with either a Stanford-Binet or a WAIS IQ were included rather than restricting the sample to just the 74 individuals tested with both IQ assessments. The first "reference" was the Vineland Adaptive Behavior Scale (Sparrow et al, 1984). Unfortunately, the number of individuals with data available for both the Vineland, expressed as an overall composite score, and a WAIS or Stanford-Binet was limited, but results seemed clear enough. Only one of 17 Vineland scores (5.9%) fell above its respective WAIS score compared to seven of 15 (46.7%) for the Stanford-Binet, exact probability < .01. Further, while there was no difference between Stanford-Binet and Vineland scores, t (14) = 0.22, $p > 0.8$, WAIS scores were significantly higher than their Vineland counterparts, t (16) = 6.74, $p < .00001$. An additional set of analyses employed other IQ tests as a reference (Leiter, Slosson or WISC formats). Stanford-Binet scores were comparable, t (11) = 0.87, $p > 0.4$, while WAIS scores were higher, t (10) = 3.49, $p < .01$.

visited 9/6/2019

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Discussion

It seems fair to say that the Stanford-Binet and WAIS IQ tests are among the most carefully developed and well respected instruments ever developed in the field of psychological assessment. Of course, the history of intelligence testing is not without its controversies and unfortunate episodes (see Kamin, 1974; Neisser, 1998), and debate persists regarding the definition of intelligence and its fundamental nature, both psychological and biological (see Sternberg & Kaufman, 1998). Nevertheless, professional awareness of the many factors that can potentially influence testing results has increased substantially in recent years, and in the vast majority of instances the well-established individually administered IQ tests provide valid (although imperfect) indications of general intelligence. This high regard has been well earned, with decades of experience demonstrating the outstanding characteristics of these scales. That makes the current findings of substantial differences in test outcomes between Stanford-Binet and WAIS IQs for adults with ID all the more surprising, impressive, and alarming. (Flynn, 2009, discussed possible overestimations of IQ specific to the WAIS-III. However, that effect was too small to explain the present findings, and in any case a preponderance of the present WAIS data predated the WAIS-III.)

Certainly, consistent disparities between IQ test results raise important issues for test developers and users. They suggest that intensified efforts need to be devoted to standardization within the tails of population distributions, clearly for the Stanford-Binet and WAIS and perhaps for all assessment instruments measuring a wide range of abilities, but consideration of the relevant technical issues are beyond the scope of this discussion. Here, the focus will be on more immediate implications of substantial practical import. These are: (a) disability determinations for SSA entitlements, (b) estimating the prevalence of ID, (c) assessment of age-associated declines in cognition, and (d) diagnosis of ID in forensic settings.

Disability determinations

People with ID legally residing within the United States are entitled to support through the Social Security Administration (see Reschly et al., 2002). For ID (still referred to as mental retardation in many jurisdictions), eligibility for adults is demonstrated by: (a) a valid contemporaneously established IQ of 59 or less (including people who are so severely impaired that they are untestable), (b) a valid IQ between 60 and 70 and another substantial physical or mental impairment, or (c) a valid IQ between 60 and 70 and substantial functional impairments (Social Security Administration, 2008). While it is acknowledged that test imprecision can allow people with “true” IQs just under 70 to score higher by chance, and scores as high as 75 might be allowed, it is clear that any IQ result in the 60’s and over imposes an increased burden of proof on the individual in question, especially for scores between 70 and 75. A score exceeding 75 would result in rejection of an application for disability benefits unless some other qualifying condition was noted.

Table 3 summarizes data recast to show the clear implications for SSA determinations. Based upon Stanford-Binet results, 94.6% of the current sample would be eligible for benefits on the basis of IQ testing alone (assuming that requirements of age at onset and economic need are also met). However, assessments using a WAIS provided sufficient evidence of ID for only 60.8% of this sample, and therefore many individuals within this sample, all of whom have clearly documented ID, would be required to provide additional evidence of their disability. Perhaps most important in this context, only one of seven individuals with a Stanford-Binet IQ of 55 or over (all under 70) had a WAIS IQ under 74 (see Figure 1).

visited 9/6/2019

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Given current conceptualizations of ID that require contemporaneous determinations of IQ, there is strong justification for alerting examiners to this difference. Results from these two tests are clearly not equivalent for this population, and consensus is needed regarding a preferred "gold standard" for SSA disability determination. The comparisons with other measures presented herein suggest that the WAIS may be systematically overestimating capabilities (although these data should be interpreted cautiously given the small samples), and should the Stanford-Binet IQs be valid, practice based on WAIS test results might actually be employing a criterion of three standard deviations below the population mean. While there has always been broad awareness of the arbitrary nature of an IQ-based criterion of 70 for defining ID, such restrictive eligibility has never been envisioned in either past or current policy.

Prevalence of ID

Prevalence estimates of ID have been quite variable in recent decades. This is illustrated perhaps most clearly by differences among various states for school-age populations (e.g., U.S. Department of Education, 2009), with estimates ranging from 0.32% for Maine to 2.48% for West Virginia. Variation in diagnostic labeling policies are one contributing factor, although it is far less clear if relevant assessment practices are involved (e.g., MacMillan, Gresham, Siperstein & Bocian, 1996), but whatever the root causes of differences in prevalence estimates, there seems to be consistency in at least one respect. Prevalence estimates for ID drop precipitously from school age (peaking at 2.03% nationally) to 18 and over (0.52%) (see Larson et al., 2001). Differences in surveillance practices would be one obvious explanation for this phenomenon, but others have been proposed. Some have suggested that adapting to life as an independent adult is associated with less demanding expectations for many individuals compared to those within school settings, and that the drop in estimated prevalence of ID accompanying transition from school to adulthood reflects the true state of functional status (see American Association on Mental Retardation, 1992). Intuitively, though, it seems hard to believe that our technically sophisticated and highly competitive culture imposes lower demands on independent adults than it does on children, and Tymchuk, Laken, & Luckasson (2001) included extensive discussions of the difficulties encountered by adults with relatively mild ID (there called "mild cognitive limitation") struggling to cope with the challenges of every day life. These discussions suggest that the adult prevalence of ID may be systematically underestimated, and a reexamination of criteria and methods used to define and to identify this population would be in order.

While it may be coincidental, it seems worth noting that IQs of school children suspected of having ID are typically assessed using either the Stanford-Binet or WISC (which also seems to underestimate WAIS IQ, e.g., Spitz, 1988), while adults would be tested with a WAIS more often than not. If scores for adult assessments tend to be, on average, 7.5 points higher (0.5 S.D.) simply because the WAIS is the dominant instrument of choice for this population, then straightforward reference to the normal distribution suggests that the overall percentage of adults scoring 70 or below would change from 2.3% to 0.6%, corresponding nicely to the drop in prevalence documented in national surveys (see Larson et al., 2001). Of course, a difference of 0.5 S.D. is considerably smaller than that found here, but whatever the contribution of test selection, the public health significance of ID cannot be addressed effectively unless and until we determine the true size of the affected population and the nature of the associated impairments. The present findings indicate that more thoughtful selection of assessment methods is required to accomplish these tasks.

visited 9/6/2019

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Declines with aging

Like everyone else, adults with ID are now living longer than ever before, and elderly individuals within this population should be at comparable (perhaps greater) risk for the same concerns their peers without ID face regarding changes in health status and cognitive capabilities. One particular concern focuses on old-age dementia, the most prevalent cause being Alzheimer's disease. Assessment of dementia in adults with substantial pre-existing cognitive/intellectual impairments is particularly difficult given that expectations for "nondemented" performance are very different for individuals with true IQs of 40 versus 65 (see Silverman et al., 2004), not to mention under 70 versus 100 and over. Pre-morbid IQ should influence both selection of appropriate instruments for assessment of performance as well as judgments formed from clinical evaluations, and it is abundantly clear that the sets of Stanford-Binet and WAIS IQs described herein cannot both provide valid indications of performance expectations. It is important to determine how best to use IQ test results to inform diagnoses given that commonly used tools to assess dementia in the general population were never designed to distinguish with precision between lifelong impairments versus those of recent onset.

Death Penalty Cases

The Supreme Court case of Atkins v. Virginia (2002) established that people with ID convicted of capital crimes are no longer subject to the death penalty. As is the case for SSA eligibility determination, contemporaneous evidence of a "true" IQ of 70 or below needs to be provided as evidence of ID. Of course, evaluators need to be aware of context-specific concerns like malingering, but differences between the WAIS or Stanford-Binet should not be among these concerns. Nevertheless, the present findings indicate that a Stanford-Binet is far more likely to support a diagnosis of ID. Baroff (2003) addressed this issue explicitly within the death penalty context, and consistent with Thorndike et al. (1986), attributed the difference to the lower floor for the Stanford-Binet. Baroff concluded his discussion with a recommendation that the WAIS continue to be the preferred choice for assessing adult intelligence in forensic settings. He also emphasized that IQ tests should never be selected based upon expectations of a higher or lower result if psychologists are to retain credibility with the courts, and that is undoubtedly correct. Nevertheless, psychologists cannot meet their ethical obligations in these cases without knowing which test provides the most valid estimate of true intelligence. The present data for individuals with relatively higher IQs, though sparse, indicate that differences between the Stanford-Binet and WAIS IQ tests can no longer be summarily dismissed as merely reflecting the scales different floors. When test results are informing judgments of literal life and death, any suspected uncertainty regarding the validity of outcomes must be addressed aggressively.

Conclusion

This was far from a perfect study, and an ideal design would have imposed greater controls over assessment procedures and timing. Nevertheless, the present findings reflect real clinical practice over a period of many decades, and the methods employed would not have biased findings for WAIS and Stanford-Binet assessments systematically. The differences found between WAIS and Stanford-Binet IQs of adults with ID described herein are too stunning to ignore and far larger than expected based upon established measurement precision for these two instruments. They call into question the validity of many previous IQ assessments for adults with a developmental history suggestive of ID, and it is particularly troubling that past evidence of substantial divergence among these two gold standard IQ assessments had been so easily dismissed. Efforts now need to focus on determining the reasons for the discrepancies, the validity of these two scales within this lower range of performance, and whether comparable differences persist for the Stanford-Binet Fifth Edition (Roid, 2003) and the WAIS-IV (Wechsler, 2008). More generally, the validity of

visited 9/6/2019

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

scores at substantial distances from their respective population means might need to be verified for other instruments developed to measure a wide range of performance.

Acknowledgments

The authors thank the many participants, agencies and agency staff who have supported our Aging Research Program throughout the many years of its existence. We also thank our staff for many their indispensable contributions: T. Bach, S. Briffa-Mirabella, Anita Camper, D. Conlon-Borsak, M. Dabbene, D. Grandjean, L. Gonzalez, L. Kullman, C. Lawrence, C. Marino, B. Myrhol, K. Olsen, G. Palma, D. Pang, J. Russo, C. Stolzenthaler, D. Swift, A. Trzeciak, M. Tsepilovan, and S. Vietze. This work was supported by funds provided by New York State through its Office of Mental Retardation and Developmental Disabilities, as well as grants PO1 HD035897 and P30 HD024061 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

References

American Association on Mental Retardation. Mental Retardation: Definition, classification, and systems of Support. 10. Washington, D.C: American Association on Mental Retardation; 2002.

American Association on Mental Retardation. Mental Retardation: Definition, classification, and systems of Support. 9. Washington, D.C: American Association on Mental Retardation; 1992 . p. 9-19.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, D.C: American Psychiatric Association; 2000.

Atkins v. Virginia. (2002). U.S. Supreme Court, Case No. 8452.

Baroff GS. Establishing mental retardation in capital cases: An update. Mental Retardation 2003;41:198–202. [PubMed: 12737613]

Bensberg GJ, Sloan W. A study of Wechsler's concept of "normal deterioration" in older mental defectives. Journal of Clinical Psychology 1950;6:356–362.

Brengelmann JC, Kenny JT. Comparison of Leiter, Wais and Stanford-Binet IQ's in Retardates. Journal of Clinical Psychology 1961;17:235–238.

Flynn JR. Wechsler Intelligence Tests: Do we really have a criterion of mental retardation? American Journal of Mental Deficiency 1985;90:236–244. [PubMed: 4083304]

Flynn JR. The mean IQ of Americans: Massive gains from 1932 to 1978. Psychological Bulletin 1984;95:29–51.

Flynn JR. The WAIS-III and WAIS-IV: *Daubert* motions favor the certainly false over the approximately true. Applied Neuropsychology 2009;16:98–104. [PubMed: 19430991]

Harrison PL, Kaufman AS, Hickman JA, Kaufman NL. A survey of tests used for adult assessment. Journal of Psychoeducational Assessment 1988;6:188–198.

Kamin, LJ. The science and politics of I.Q. Potomac: MD: Lawrence Erlbaum Associates, Inc; 1974.

Larson SA, Lakin KC, Anderson L, Kwak N, Lee JH, Anderson D. Prevalence of mental retardation and developmental disabilities: estimates from the 1994/1995 national health interview survey disability supplements. American Journal on Mental Retardation 2001;106:231–252.

Levine, MN. Leiter International Performance Scale: A handbook. Los Angeles: CA. Western Psychological Services; 1993.

MacMillan DL, Gresham FM, Siperstein GN, Bocian KM. The labyrinth of IDA: School decisions on referred students with subaverage general intelligence. American Journal on Mental Retardation 1996;101:161–174. [PubMed: 8883670]

Neisser, U., editor. The rising curve: long-term gains in IQ and related measures. Washington, D.C: American Psychological Association; 1998.

Nelson WM III, Dacey CM. Validity of the Stanford-Binet Intelligence Scale-IV: Its use in young adults with mental retardation. Mental Retardation 1999;37:319–325. [PubMed: 10463026]

Nicolson, CL.; Hibpshman, TH. Slosson Intellligence Test-Revised (manual). E. Aurora, NY: Slosson Educational Publications; 1991.

visited 9/6/2019

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Reschly, DJ.; Myers, TG.; Hartel, CR., editors. National Research Council, Division of Behavioral and Social Sciences and Education. Washington, D.C: National Academy Press; 2002. Mental Retardation: Determining Eligibility for Social Security Benefits.

Roid, GH. Stanford-Binet Intelligence Scales. 5. Rolling Meadows, IL: Riverside Publishing; 2003.

Silverman W. Down syndrome: cognitive phenotype. Mental Retardation and Developmental Disabilities Research Reviews 2007;13:228–236. [PubMed: 17910084]

Silverman W, Schupf N, Zigman W, Devenny D, Miezejeski C, Schubert R, Ryan R. Dementia in adults with mental retardation: Assessment at a single point in time. American Journal on Mental Retardation 2004;109:111–125. [PubMed: 15000666]

Sparrow, SS.; Balla, DA.; Cicchetti, DV. Vineland adaptive behavior scales. Circle Pines, MN: American Guidance Service; 1984.

Social Security Administration. Disability Evaluation under Social Security. 2008. SSA Publication No. 54-039, section 12.05. www.ssa.gov/disability/professionals/bluebook

SoftStat, Inc. Statistica (data analysis software system). 2008. version 8, www.softstat.com

Spitz HH. Disparities in mentally retarded persons' IQs derived from different intelligence tests. American Journal of Mental Deficiency 1986;90:588–591. [PubMed: 3953693]

Spitz HH. Inverse relationship between the WISC-R/WAIS-R score disparity and IQ level in the lower range of intelligence. American Journal on Mental Retardation 1988;92:376–378. [PubMed: 3342139]

Sternberg RJ, Kaufman JC. Human Abilities. Annual Review of Psychology 1998;49:479–502.

Thorndike, RL.; Hagen, EP.; Sattler, JM. Stanford-Binet Intelligence Scale. 4. Chicago, IL: Riverside; 1986. Correlations between Stanford-Binet Intelligence Scale: Fourth Edition SAS's and Wechsler Adult Intelligence Scale – Revised (WAIS-R) IQs for Examinees Designed by Their Schools as Mentally Retarded.

Tymchuk, AJ.; Lakin, KC.; Luckasson, R. The Forgotten Generation: The Status and Challenges of Adults with Mild Cognitive Limitations. Baltimore, MD: Brookes; 2001.

U.S. Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs. 28th Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act, 2006; Washington, D.C.. 2009.

Wechsler, D. Wechsler adult intelligence scale. 4. San Antonio, TX: Pearson; 2008. technical and interpretive manual

Widaman KF, Siperstein G. Assessing adaptive behavior of criminal defendants in capital cases: A reconsideration. American Journal of Forensic Psychology. in press.

visited 9/6/2019

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

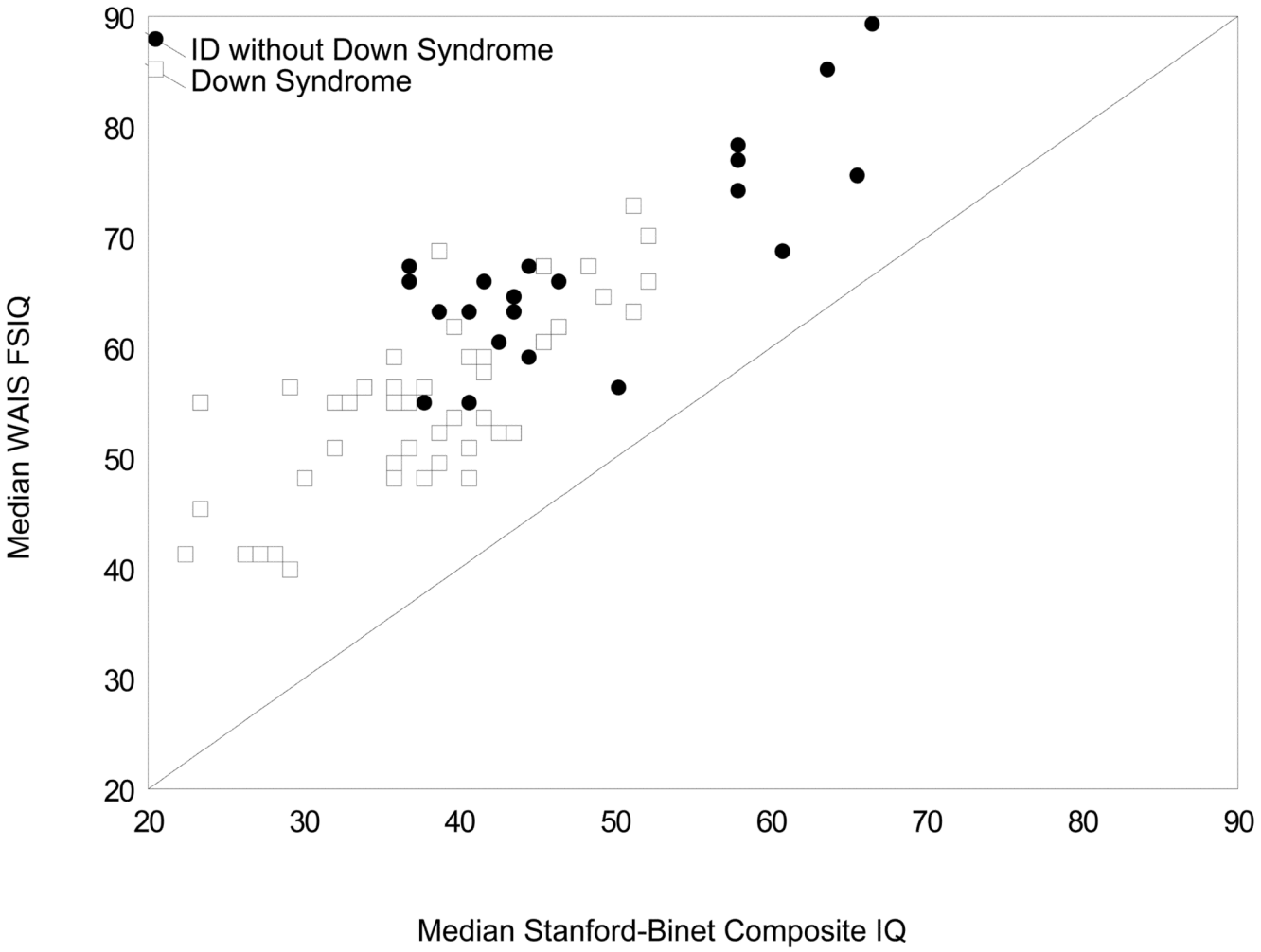

Figure 1.
Scatterplot of individual Stanford-Binet and WAIS IQs of 74 adults with intellectual disability, with a diagonal to indicate equality. (Data for adults with and without Down syndrome are indicated by different symbols.)

visited 9/6/2019

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Characteristics of 74 Individuals with Both WAIS and Stanford-Binet IQs Documented in Clinical Records.

Gender	
Male	32.4%
Female	67.6%
Etiology	
Down syndrome	70.3%
Other Causes	29.7%
Birth Year	
Median	1947
Range	1914 – 1973
Stanford-Binet IQ	
Mean (SD)	41.3 (9.33)
Range	23 – 67
Mean Test Age (SD)	45.3 (17.9)
WAIS IQ	
Mean (SD)	58.1 (9.80)
Range	39 – 88
Mean Test Age (SD)	44.3 (15.6)

visited 9/6/2019

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 2

Distribution of WAIS/Stanford-Binet IQ Differences for Adults with Down syndrome (DS) and Intellectual Disability (ID) Due to Other Causes.

	Magnitude of Difference (IQ points)					
	1 – 5	**6 – 10**	**11 – 15**	**16 – 20**	**21 – 25**	**>25**
Percent of DS (n = 52)	0	15.38	34.62	30.77	13.46	5.77
Percent of ID (n = 22)	4.55	9.09	9.09	40.91	22.72	13.64
Percent of Overall	1.35	13.51	27.03	33.78	16.22	8.11

visited 9/6/2019

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 3

Comparison between Stanford-Binet and WAIS IQs for Categories Relevant to Disability Determination under U.S. Social Security Administration Guidelines.

	Stanford-Binet IQ				
WAIS IQ	**< 60**	**60 – 70**	**71 – 75**	**>75**	**% of Total**
< 60	45	0	0	0	60.8%
60 – 70	21	1	0	0	29.7%
71 – 75	2	1	0	0	4.1%
>75	2	2	0	0	5.4%
% of Total	94.6%	5.4%	0%	0%	