RIVIER UNIVERSITY

**DIVISION OF EDUCATION**

# **SPECIALIST IN THE ASSESSMENT OF INTELLECTUAL FUNCTIONING PROGRAM**

AND

**ASSOCIATION OF SPECIALISTS IN ASSESSMENT OF**

**INTELLECTUAL FUNCTIONING (ASAIF)**

**Comments on Reports 5/4/16 # 257**

**CONTENT**

**WISC-V comparisons with WISC-IV scores**. If we re-test (or re-re-re-test) with the WISC-V a student who previously took the WISC-IV, I think it is prudent to administer Comprehension, Symbol Search, Information (if the student took it on the WISC-IV), Picture Concepts, Arithmetic (if the student took it on the WISC-IV), and especially Letter-Number Sequencing (so we can compare the WISC-IV WMI with the WISC-V AWMI (or, as Laurie Farr Hanks noticed that it is written in one place on the Record Form, AMWI). I use ±2 scaled score points as an approximation for a 95% confidence interval.

**WISC-IV and WISC-V Scores**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Index/**Subtests | **WISC-IV** | | **WISC-V** | | **Comparison** | |
| **Standard Score/**  Scaled Score1 | **Confidence Interval (95%)2** | **Standard Score/**  Scaled Score | **Confidence Interval (95%)** | **Change** | **Do the Confidence Bands Overlap?** |
| Similarities\*† |  |  |  |  |  |  |
| Vocabulary\*† |  |  |  |  |  |  |
| Comprehension\* |  |  |  |  |  |  |
| *Information* |  |  |  |  |  |  |
| *Word Reasoning* |  |  |  |  |  |  |
| **Verbal Comprehension** |  |  |  |  |  |  |
| Block Design\*† |  |  |  |  |  |  |
| Visual Puzzles |  |  |  |  |  |  |
| **Visual-Spatial** |  |  |  |  |  |  |
| Matrix Reasoning\*† |  |  |  |  |  |  |
| Figure Weights† |  |  |  |  |  |  |
| Picture Concepts\* |  |  |  |  |  |  |
| *Arithmetic* |  |  |  |  |  |  |
| **Fluid Reasoning** |  |  |  |  |  |  |
| **Perceptual Reasoning vs. Visual-Spatial** |  |  |  |  |  |  |
| **Perceptual Reasoning vs. Fluid Reasoning** |  |  |  |  |  |  |
| **General Ability Index** |  |  |  |  |  |  |
| Digit Span\*† |  |  |  |  |  |  |
| Letter-Number Sequencing\* |  |  |  |  |  |  |
| Picture Span |  |  |  |  |  |  |
| Arithmetic |  |  |  |  |  |  |
| **(Auditory) Working Memory** |  |  |  |  |  |  |
| **Working Memory** |  |  |  |  |  |  |
| Coding\*† |  |  |  |  |  |  |
| Symbol Search\* |  |  |  |  |  |  |
| *Cancellation* |  |  |  |  |  |  |
| **Processing Speed** |  |  |  |  |  |  |
| **Cognitive Proficiency Index** |  |  |  |  |  |  |
| **Full Scale IQ** |  |  |  |  |  |  |

1. WISC standard scores have a mean (average) of 100. The middle half of all examinees' standard scores are between 90 and 100. WISC scaled scores have a mean (average) of 10. The middle half of all examinees' scaled scores are between 8 and 12.

2. On even the very best tests, scores can never be perfectly accurate. Lucky and unlucky guesses, finishing just before or beyond the time limit, and other random factors cause scores to vary. This score interval shows how much scores are likely to vary randomly 95% of the time.

\* Included in WISC-IV Full Scale IQ † Included in WISC-V Full Scale IQ

*Subtests in Italics are not included in the index scores.*

***DISKoveries—Literacy for Special Needs: The Impact of New Technologies* by Joan Tanenhaus.** Jonas Taub came upon this article while doing some other related work. It contains a lot of information on technology and apps for reading and literacy for kids with special needs. He kindly provided a link to the article online. It is a pdf and you can easily download a copy to save and print. I have done so.

<http://www.strategictransitions.com/pdfs/DISKoveries2015-October.pdf>

**Composite Extremity Effects in Test Scores.** This question of composite scores differing from the average of the component test scores comes up from time to time whenever someone stumbles across the phenomenon with a new test or whenever the total score winds up on the other side of an arbitrary classification demarcation from the (sub)tests ("How can Low subtests give us a Very Low composite?"). Readers of *Report Comments* know all the following information, but the explanations might be helpful when explaining the concept to teachers and parents. There is a lot of information available at Guy McBride, Ron Dumont, and my web page, [www.myschoolpsychology.com](http://www.myschoolpsychology.com/). For example,

<http://www.myschoolpsychology.com/testing-information/#mnemonics> includes links to Ron Dumont and my "Luke Composite Effect" (and other Biblical Mnemonics) and to Kevin McGrew's detailed explanation <http://www.iqscorner.com/2015/03/why-does-wj-iv-gia-score-often-appear.html>, which contains links to <http://www.iqscorner.com/2011/03/iap-applied-psychometrics-101-report-10.html> and Kevin and Joel Schneider's <http://www.iapsych.com/iapap101/iap10110.pdf>.

Kevin also links to Joel Schneider's "great explanation (with videos) of this phenomenon at his awesome blog":  <https://assessingpsyche.wordpress.com/2014/04/19/why-composite-scores-are-more-extreme-than-the-average-of-their-parts/>.  Be sure to study this very helpful explanation.  (We, of course, have a link to Joel's awesome blog at <http://www.myschoolpsychology.com/special-education-links-2/#statistics>. Visit early and visit often.)  Joel uses the more useful and descriptive term, "Composite Extremity Effects." Joel's term is better, but Ron and I have long been committed to Luke 8:18.

The trick, of course, is explaining the phenomenon to team members and readers of our reports. Total scores will be more extreme (farther from the mean) than the average of the components.  Suppose Mordred's WISC-V subtest scaled scores are all 5 (statistically equivalent to an index score of 75, percentile rank 5).  We would expect his composite scores all to be 75, wouldn't we?  Nope.  Ecomodine nailed every subtest with a scaled score of 15 (equivalent to an index score of 125).  Her index scores should all be 125, right?  Again, nope.

**Mordred                             Ecomodine**

          all scaled                               all scaled

        scores = 5                            scores = 15

Percentile Rank (PR) 5       Percentile Rank (PR) 95

  (= index score 75)                (= index score 125)

  VCI   73 (PR 4)               VCI   127 (PR 96)

  VSI     72 (PR 3)                 VSI     129 (PR 97)

  FRI       72 (PR 3)                 FRI     128 (PR 97)

  WMI   72 (PR 3)                 WMI     127 (PR 96)

  PSI       72 (PR 3)                 PSI     129 (PR 97)

  FSIQ     67 (PR 1)                 FSIQ   135 (PR 99)

Short answer: most people who score as low as scaled score 5 (percentile rank or PR 5) on a subtest will score higher on other subtests.  Only 5 percent of people score in or below PR 5.  The other 95 percent score higher.  It is unusual for someone to score in the lowest five percent on a subtest.  It is even more unusual to score that low (or lower) on another subtest, so Vocabulary PR 5 and Similarities PR 5 give a lower (more extreme) VCI in PR 4.  It is even more unusualler to score that low (or lower) on three, and so on, so all subtests are in percentile rank (PR) 5, but the total for seven subtests gives a much more extreme FSIQ PR 1!  Same deal for unusually high scores.  Please see Kevin's and Joel's papers for more detailed and accurate explanations.

The Luke or the Composite Extremity Effect is greater for more extreme scores (Billy Bob's and Quathrynne's scores show smaller effects than Mordred's and Ecomodine's).  It is greater when there are more component parts (seven-subtest FSIQs show greater effects than two-subtest composites).

**Quathrynne** **Williamrobert (Billy Bob)**

  all scaled                               all scaled

        scores = 9                            scores = 11

Percentile Rank (PR) 37     Percentile Rank (PR) 63

  (= index score 95)                 (= index score 105)

VCI     95 (PR 37)               VCI     106 (PR 66)

VSI       94 (PR 34)               VSI     105 (PR 63)

FRI       94 (PR 34)               FRI     106 (PR 66)

WMI     94 (PR 34)               WMI     107 (PR 68)

PSI       95 (PR 37)               PSI     105 (PR 63)

FSIQ     93 (PR 32)               FSIQ   107 (PR 68)

It is the same for the DAS-II, WJ IV, KABC-II, RIAS-2, SB5, and all other tests that use some kind of standard score (e.g., Wechsler-type scaled and standard scores, T scores, z-scores, BOT-2 scale scores *v*-scale scores, standard age scores) and have (sub)tests and composites.  It is more obvious with tests that use the same statistic for (sub)tests and composites (e.g., WJ IV, WIAT-III, KTEA-3) than with tests that use different statistics for (sub)tests and composites. It is a statistical fact of life.

Analogies to gymnastics, the decathlon, or similar multi-event athletic contests may help us explain the Composite Extremity Effect to our audiences. For example, in the 1976 Montréal Olympic Decathlon, Caitlyn (then Bruce) Jenner ran the 1500-meter final in 4:12.  That was certainly unusual by general population norms, but not mind-boggling (even I managed 4:20 in 1964 [without first competing in nine other events over two days], and John Walker won the open 1500 meters at Montréal in 3:39).  However, Jenner also flung a discus 50 meters, a javelin 69 meters, and a cannon ball more than 15 meters, all of which were pretty unusual, but not world records for the same events outside a decathlon.  Jenner also sprinted and hurdled with unusual speed and jumped and pole vaulted (an event in which I was a danger to self and others) unusual heights and distances. However, the **combination** of unusual, but less-than-world-record, accomplishments in all ten events was, of course, extremely unusual and earned Jenner the Olympic Decathlon gold medal and a picture on Wheaties boxes.

**Don't Throw Away your WJ III.** To the best of my rapidly fading knowledge, the WJ III was the only current (until 2014) test that tested delayed recall over a span of more than about a half-hour.  You could test Story Recall, Visual-Auditory Learning, and Memory for Names and then retest them any time from the usual and customary half-hour to eight (8) days later.  The scoring program took the length of the delay into account when calculating the score. (Yes, of course I did spend an exciting day re-running the same raw scores over and over while changing the date and time of the retest to confirm that assumption. The scores did change with increasing delay spans, but at wider intervals as the span increased until the curve flattened out, as we would anticipate from research data.)  A teacher might tell me that Ecomodine understood what the teacher said at the time, but always forgot the material overnight.  I was able to come back and say either, "Yup, I saw the same thing when she listened to stories, retold them immediately with good accuracy (percentile rank 63 for her age), and barely remembered anything at all the next day (percentile rank 6 compared to students her age who were also retested one day later)." or "Sorry, Mr. Ennuyeux, it's you.  On my test, she listened to stories, retold them immediately with good accuracy (percentile rank 63 for her age), and remembered pretty much everything the next day (percentile rank 66 compared to students her age who were also retested one day later)."

Delayed Recall was voted off the island on the WJ IV.  If the referral questions called for assessment of recall over 24 hours or more, and until the WJ III NU norms get too old or another test offers overnight delayed recall norms, I would still occasionally use one or more the WJ III long-delayed recall tests (but obviously not include the same WJ IV immediate recall tests).

Similarly, my beloved Elithorn Mazes did not make the cut for the WISC-V Integrated, so I am hanging on to my remaining record forms and *WISC-IV Integrated Administration and Scoring Manual.*

# **Norms for Persons with Specific Diagnoses**

A parent asked about IQ test norms for children with Down Syndrome that could be used for their child. The request is uncommon (base rate < 10%) but I have gotten it occasionally with regard to various exceptionalities.  The short answer is "no.

The longer answer is that we do not know enough about the demographics of folks with any particular disability, so the massive effort to assemble a large, random, stratified sample of people with any particular diagnosis would be futile.  A sufficiently motivated (by potential profit) publisher would need to count all of the of children diagnosed with Down Syndrome from each geographic region of the country, and from each racial or ethnic group, including the numbers of children in regular classes, receiving special education in regular classes, receiving part-time pull-out special education services, attending self-contained classes for most of the school day, attending special day programs, and attending residential programs.  They would also need to count the children with various additional conditions or disabilities.

Then they would need to collect a randomized sample of at least 1,000 children stratified to match the proportions of the above variables and also to match census data for parent education, parent income, urban and rural residence, and other potentially germane characteristics.

Most (and all acceptable) test manuals include data on scores of small samples persons in certain broad disability categories, but not, in my experience, narrow specific diagnostic categories.  They might go as far as mild, moderate, and severe intellectual disability, for example, but not as far as Down Syndrome specifically.

We could find journal articles about IQ scores for children with Down Syndrome or another specific diagnosis (probably in medical journals as well as psychology journals), but they would not be based on nationwide stratified random samples matching actual proportions as outlined above.  Articles and books about Down Syndrome might include comments about the distribution of IQ scores, but I would not have much faith in the accuracy of those estimates.  Even in a focused study, a public school sample, for example, might omit children with very severe intellectual impairments who were attending private day or residential programs.  It would also be very easy to miss the small, but significant number of persons with Down Syndrome who had cognitive abilities within the average range and who were not receiving special education services.

Data would also require decisions about whether to include individuals with additional disabilities or conditions.  A sample of people with Down Syndrome would look very different depending on whether you included, for instance, persons with severe vision or hearing impairments.

**The College Board Elegizes Anachronistic Verbiage with Recondite Panegyric; Celebrates Final Administration of the Extant SAT® on Jan. 23**

# **New York 01/26/16** — Throughout its 100-year history, the abstruse vocabulary words of the SAT® have engendered prodigious vexation in millions of examinees annually. On Saturday, Jan. 23, students across the country participated in the terminal transpiration of the SAT in its habituated gestalt.

# To adumbrate the changes to be manifest in future administrations of the assessment: The new SAT will be more trenchant and pellucid, and the format will no longer pertinaciously reward students who punctiliously engage in the antediluvian praxis of committing idiosyncratic words to memory. College Board President David Coleman promulgated, “Your invectives and maledictions have been heard. Clemency has been granted.” Many within the College Board and the academic community expressed a paucity of maudlin or mawkish emotion in response to the announcement.

# “This is a new beginning for the SAT. Gone are obscure vocabulary words and tricky logic questions that are disconnected from the work students do every day,” said Stacy Caldwell, vice president of the SAT Program at the College Board. “Moving forward, students will encounter a test that focuses on the few things that matter most for college, work, and life. We believe these changes will benefit students and educators alike.” The redesigned SAT will debut on March 5, 2016.

# <https://www.collegeboard.org/releases/2016/college-board-elegizes-anachronistic-verbiage-recondite-panegyric-celebrates-final-administration-extant-sat-jan-23>

# **STYLE**

Don’t write merely to be understood. Write so that you cannot possibly be misunderstood.

– Robert Louis Stevenson

**Changing words.**  Words do change meaning over time, and the disapproval of old fogies such as I is irrelevant. (I do wonder sometimes if "strict constructionist" and "originalist" interpreters of the U. S. Constitution, such as the late Associate Justice Antonin Scalia, regularly consult late 18th century and early 19th century dictionaries to ascertain the then-intended meanings of the words in the Constitution and the first 22 amendments.) When we write reports, I think it is prudent to avoid any word undergoing a change in meaning. Otherwise we may "possibly be misunderstood." *Unique*, for example, has long meant "one of a kind," so constructions such as "very unique" and "the most unique" were utter nonsense. However, English speakers are increasingly using *unique* to mean merely "uncommon." Old fogies such as I may lament the loss of a useful, specific meaning for a word but, like it or not, the meaning is changing. The danger for writers of reports (and other documents) is the risk of being misunderstood. If I wrote that "Siouxanne's pattern of DAS scores is, in my experience, unique," I would mean that I have never seen that pattern in all of the DAS and DAS-II results that I have generated, reviewed, or studied in textbooks since 1990. Some readers would understand my sentence as I intended. However, many readers, probably the majority of parents and teachers too young to receive Social Security benefits, would interpret my sentence as meaning that Siouxanne's pattern of scores was merely moderately uncommon. To be sure of being understood, I should instead write something like, "I have never seen, in the hundreds of DAS scores I have reviewed, a pattern of scores similar to Siouxanne's."

When I was very young, "brain storm" meant a mental breakdown. That term has now shifted completely, so it is safe to use if you really must. Before the total shift, a "brain storming session" would have been more interesting than they usually are. "Bimbo" shifted before my time from referring primarily to men to referring primarily to women.

John O. Willis

Senior Lecturer in Assessment, Rivier University

Assessment Specialist,

Regional Services and Education Center

419 Sand Hill Road

Peterborough, NH 03458-1616

(603) 924-0993 [johnzerowillis@yahoo.com](mailto:johnzerowillis@yahoo.com)

This newsletter goes out intermittently to about 400 people on eleven separate lists (because some mailboxes won't accept mailings to more than 49 recipients). If you wish to contact the entire list, not just your 1/11 of it, please send the message to me, and I will add it (subject to Comstock, Hays, Children's Internet Protection Act, HIPAA, FERPA, copyright, and Homeland Security considerations) to the next mailing. If you wish to be protected from receiving future copies, just email me at [johnzerowillis@yahoo.com](mailto:johnzerowillis@yahoo.com).