**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?"). 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>.)  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 Composite 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.

John O. Willis

Rivier University

Regional Services and Education Center

3.18.16