Wormeli, C. T.
2013.06.06

"AVERAGE":

a single value (as a mean, mode, or median) that summarizes or represents the general significance of a set of unequal values

from earlier *average* proportionally distributed charge for damage at sea, modification of Middle French *avarie* damage to ship or cargo, from Old Italian *avaria,* from Arabic *'awārīya* damaged merchandise

First Known Use: 1732
(Merriam-Webster On-line Dictionary. http://www.merriam-webster.com/dictionary/)

[syn.] regular, normal, usual, typical, middling, ordinary . . . mean, mode, median, norm, arithmetic mean . . .

Average probably appears as a word in almost every document that a school psychologist writes. It either stands alone or is used as part of phrase that describes a client's performance: high average, low average, below average . . . As is evident from the etymology above, the meaning of average has changed over the last two centuries and more. Our [School Psychologists'] interest in the meaning of average is a reflection of the language we use to describe and organize information that relates to an individual's performance on various scales that we administer and then interpret. We use it, particularly, when writing about the performance of individuals who are diagnosed with a learning disorder or who are recommended for designation as a student with a learning disability. This discussion is intended to supplement **Best Practice Guidelines for the Assessment, Diagnosis and Identification of Students With Learning Disabilities (BCASP, 2007).**

The BC Ministry of Education accepted in 2002 the definition of learning disabilities that was written by the Canadian Association for Learning Disabilities:

Learning disabilities refers to a number of disorders that may affect the acquisition, organization, retention, understanding or use of verbal or nonverbal information. These disorders affect learning in individuals who otherwise demonstrate <u>at least average abilities essential for thinking and/or reasoning</u> . . . (Special Education Services, p. 46).

Since the adoption of this definition for students categorized by the Ministry as "**Q**," the words "thinking," "reasoning," and "average," especially the word "average," have coloured discussions within our profession that involve assessment of students with learning disabilities. Until more recently (2006), our definitions of this particular word varied at least as much as the words that I found in my laptop's thesaurus (above). In DSM-5 (2013) "normal intellectual functioning" is described as -2 SDs and higher (I. Q. $\geq 70 \pm 5$), yet "well below average" scores (for achievement) are described as at least 1.5 SDs below the population mean, although it is possible to diagnose a specific learning disorder with an achievement score that is as high as one SD below the population mean. Test developers, on the other hand, have tried to define "average" with proprietary quantitative definitions that have ranged from 2/3 of a standard deviation to 4/3 of a standard deviation on either side of a mean, e. g.:

110-120 high average
90-109 average
80- 89 low average

But there was not a standard across tests; I have written (and read) reports in which "average" for one test included a range that differed from that of another test that was used in the same assessment. As I began to wonder how parents and teachers made sense of that, one of my mentors, O. A. Oldridge, asked me: "If 90 and 109 are both 'average' scores [that are not much different from each other], would it be important if your own child had a Full Scale IQ of 90 or 109?" Subsequently, I wrote a set of descriptors for scores that owed something to Kaufman's KABC, as well as to my mentor and to my own research, in which I found that, with regard to achievement test scores, students with scores lower than -1 SD from the mean were more likely to have been referred for remedial instruction than those with scores higher than - 1 SD from the mean:

well above average	126 -
above average	116 - 125
high average	106 - 115
average	96 - 105
low average	86 - 95
below average	76 - 85
well below average	- 75

I defined "average" as applying to range of scores (±1 SD) on either side of the mean, and I sliced that range into three equal portions, rather than one large and two smaller portions because that allowed more differentiation of scores across that range. But for me, -1 SD was the important value. I initially parked "85" with "below average" because in my analyses, when I averaged and rounded my results, that value fell in the referred group. I used "75" to mark the highest score of the lowest category because that was the highest score that could be used to establish a diagnosis of intellectual disability, and the score of "126" is among the lowest scores that some writers have used to describe "gifted" performance. My colleagues in my school district subsequently adopted, and adapted, that scale to enhance consistency across our assessments. But that was only in one district.

After considerable discussion, including consultation with BCASP, in 2006 the Special Education branch of the Ministry of Education defined "average" as -1 SD and higher ("above average" or "below average" was not defined or described): "Students with average or above potential will score at or above one standard deviation below the mean for their age on standardized tests (*Special Education Services: A Manual of Policies, Procedures and Guidelines* - March 2011, p. 48)." This provided practitioners for the first time with a provincial standard that could be used to interpret test results. Now, I still see reports in which a publisher's descriptor of 80-89 as "low average" is used, which means that scores of 88 and 82 are described in the same way, and I still wonder if that is confusing to parents or teachers who may not have the same level of training as a psychologist, but I presume that the writers of such reports are able to educate the parents about what a standard deviation is. My latest scale now looks like this:

well above average	126 -
above average	116 - 125
high average	106 - 115
average	95 - 105
low average	85 - 94
below average	76 - 84
well below average	- 75

Recently, Dawn Flanagan and Kevin McGrew, who have operationalized the application of Cattell-Horn-Carroll theory to cognitive assessment and measurement, have promoted the concept of -1 SD as being the cut-off between age-appropriate or not in the *Essentials of Cross-Battery Assessment* editions and in the third edition of *Contemporary Intellectual Assessment* (Flanagan, D. & Harrison P. (Eds), 2012). Their scale is (Table 26.3, p. 650):

Upper extreme	130 -
above average	116 - 129
high average	111 - 115
average	90 - 110
low average	85 - 89
below average	70 - 84
Lower extreme	- 69

One standard deviation below the mean has become a marker that we use to describe the lowest performance that is “average” or “age-appropriate” or “normal” or “typical” or in some other way as likely not needing to be supported. And it makes common sense. The lowest achieving four or five students in a regular classroom of 25-30 students are likely to require more support than their classmates. The reasons for their low achievement and the forms of support needed are separate issues, but it is likely that they will need more intervention, accommodation or adaptation than their peers to deal with the curriculum and the materials in their classroom.

Although we have guidance now on what an average performance ***is***, we have less guidance on what ***average*** is - on how to apply –1 SD to the diagnoses that we make or to the designations that we recommend with regard to children or adults, especially with regard to their eligibility for designation as learning disabled. And our need to do this persists: the latest edition of the *Diagnostic and Statistical Manual* describes specific learning disorder with reference to “well below average” performance that is not “appropriate for their intelligence and age” (DSM5|psychiatry.org, May 3, 2013), and -1 SD for achievement separates individuals for whom a diagnosis of specific learning disorder might be appropriate from those for whom it is not.

Profile 1

Consider, e. g., an instance in which an examinee has obtained cognitive standard scores that are almost two SDs < mean on all the abilities or factors of a level C test (we will use the WISC-IV, since that is the most commonly used level C test; we will use these factor scores: PRI=70; WMI=74; PSI=75 and say all the factors are unitary), except for a score for verbal ability that is in the low average range (e. g., 85). If the individual’s reading comprehension score is 70, is this student eligible for designation as a student with a learning disability or diagnosis as a person with a learning disorder? And how does this relate to the definition of average thinking or reasoning [DSM-5 specifiers for diagnosis of learning disorder no longer include average ability, except to exclude intellectual disability, although there *is* a specifier for low achievement for age (APA, 2013)]?

Our question restated is: “What is this student’s academic potential?” Is it that low average score for verbal ability, or is it the well below average overall (Full Scale) score? Even though the score for verbal ability is significantly higher than the other cognitive factor scores, it is not so different (i.e., the base rate of the differences between the score for verbal ability and the other scores is higher than 10 %) that we should consider *not* reporting the overall score for the test. And if the overall score obtained from the test is 72, and if we decide to use that as the estimate of aptitude, then clearly there is *no* support for designation as a student with a learning disability. If we decide to use the overall score, then we subsume that verbal ability (thinking) score into the overall test result.

The Ministry does not tell us how to decide what a student's academic potential is. That is left to practitioners and local institutions (using language that contains the word "must" can be expensive, apparently). Flanagan, Alfonso and Sotela-Dynega et al. (2012) noted that "Determining an average . . . profile for a child who has a below average cognitive-achievement consistency is not a straightforward task, and there is no agreed-upon method for determining this condition (p. 663)." While much of the time the results of our assessments are as straightforward as the vignette described above, there are occasions when they are more difficult to interpret. Currently, our (slightly dated) *Guidelines* relies upon the use of publishers' factor and composite scores to establish either an overall estimate of aptitude or partial estimates that eschew one or another cognitive ability in favour of a core assembly of measures that are intended to measure either the most important cognitive abilities (e. g. reasoning) or those of one type (e. g., "verbal" ability or "nonverbal" ability; "simultaneous processing" or "sequential processing"). There is merit in this: the WISC-IV's GAI and the DAS-II's GCA are based on the cognitive abilities that contemporary CHC theory, the most widely accepted theory among school psychologists, has ranked as the first, second and fourth most closely correlated with intelligence (below, p. 7): Fluid Reasoning, Crystallized Knowledge, and Visual Processing; the SB-V's composite score is based on the cognitive abilities that are considered to be the four most highly correlated with g: Fluid Reasoning, Crystallized Knowledge, Short-Term Memory and Visual Processing, as well as verbal and nonverbal composite scores.

If these aren't enough, we can create our own (DIY) composites, using subtest classifications based on CHC practice in the latest edition of *Essentials of Cross-Battery Assessment* or other sources, and averaging them. If we know the correlations between the subtests we would like to form into composites, we can use formulas created by Tellegen and Briggs (1967) for the WISC or by Schnieder and McGrew (2011) for any test to create our more accurate composites. Or we can use the spreadsheets on the CD in the third edition of *Essentials of Cross-Battery Assessment*.

But in the instance described above, we might regard these adaptations as insufficient. How are we to interpret such a profile? I think that to answer that question, we must consider some other questions:

- How do we represent an individual's aptitude (or intelligence)?
- If the score for one Broad Cognitive Ability (or factor) is age-appropriate and scores for other Broad Cognitive Abilities (or factors) are not, does that age-appropriate ability take precedence; that is, does it become by default the estimate of potential and does it, if it is unitary, *exclude* other estimates of potential?
- And does it matter which Broad Cognitive Ability it is; that is, does it matter if it is Long-Term Memory/Retrieval, as opposed to Fluid Reasoning?

Kaufman provides a defensible routine for helping us to decide how many and which score(s) we should use to represent an individual's aptitude. One of the latest iterations of this is to be found in appendix E of *Essentials of the WISC-IV, 2nd Edition* (2009) that he and Dawn Flanagan wrote. While this approach is intended for a particular test, it seems to me that it is reasonable to apply it to any cognitive battery that an examiner uses. The routine is intended to identify the most comprehensive score(s) that might be considered to represent a unitary construct. I summarize it here:

> Step 1. Determine if the overall score that you have obtained from your primary level C test battery is appropriate to report. There are two indices to review: a) the range of subtest scores and b) the differences between the factor scores that are specific to whatever test you are using. Base rates are important to both. With regard to the range of subtest scores, a range (highest to lowest) that occurs in more than 10 % of the normative sample may be acceptable to most practitioners. It is generally what I use; in my opinion, base rates of 10 % are the threshold of that which we might find to be unusual. Using a base rate of 15 % seems not terribly uncommon, and using a base rate of 5 % seems to narrow the definition of unusual a little too much. A similar logic may be applied to the second index: differences in composite factor scores that are obtained from amalgamating the subtest scores into the test battery's interpretive matrix. Kaufman and Flanagan rely upon a third metric: a difference of 1.5 SDs between the highest and lowest factor composite scores as a cut-off for excluding the overall test composite from representing the aptitude of an individual. This includes differences that involve (with regard to Canadian norms for the WISC-IV) base rate differences as low as 3.0 % (Table B2, p. 189: VCI<PRI, WISC-IV^{CDN} Canadian Manual). To me this seems to allow too much difference; in distinction, Saklofske et al. (2008) suggest using base rate differences of 10-15 % to determine when to consider using the GAI.
>
> While software that usually accompanies these tests can do all or most of these comparisons for you, not all test publishers provide the base-rate data that are used to support this point, and for these tests, 1.5 SDs might be helpful to use as a maximum cut-off. In the instance of the WISC-IV, there are four factor scores to compare; for the SB-V there are five, as well as verbal and nonverbal scores. For the DAS-II there are three (for the GCA). For the WJ-III there are seven.
>
> If the range of subtest scores and the differences between composite factor scores occur more often than in 10 % of the normative sample, then using the overall score is recommended. We can stop here with regard to estimating aptitude and intellectual potential. The overall score includes measures of "thinking" and "reasoning," is always more reliable than any score that represents a less comprehensive construct, and correlates better with achievement than does a score for any less comprehensive construct. That does not mean that we do not bother to interpret significant differences between factor or Broad Ability composites, but it does mean that we likely have a robust estimate for overall aptitude.
>
> If the range of subtest scores or the differences between composite factor scores occur in less than 10 % (or whatever base rate you choose) of the normative sample, then it is appropriate to consider the use of a less comprehensive estimate of potential (Step 2).
>
> Step 2. Several level C cognitive batteries provide ready-made restricted alternative estimates of aptitude. In the instance of the WISC-IV, this is the GAI, which is based upon two factor scores that are, in turn, based upon three Broad Cognitive Abilities, and two of these Broad Abilities are intended to measure reasoning and thinking. If the difference between the highest and lowest factor/ability scores used for the restricted composite is less than 1.5 SDs or the base rate is

higher than 10 %, then the use of the restricted composite is recommended, and we can stop here. While a restricted composite is less reliable than the overall test score, it is likely more reliable than discrete factor or ability scores and likely correlates better with achievement.

WJ-III software offers several alternatives to the overall extended GIA, and in the SB-V manual several partial composites are to be found. The DAS-II GCA is already a restricted composite, although not quite as restricted as the WISC-IV's GAI: 2/3 of the subtests used to calculate the GCA are intended to represent Fluid Reasoning and Crystalized Knowledge, while 5/6 of the GIA subtests represent these two Broad Abilities.

If the difference in the factor or Ability scores that contribute to the restricted composite is > 1.5 SDs or the base rate less than 10 %, then an examiner is encouraged to explore the individual factor/ability composites that are available (step 3).

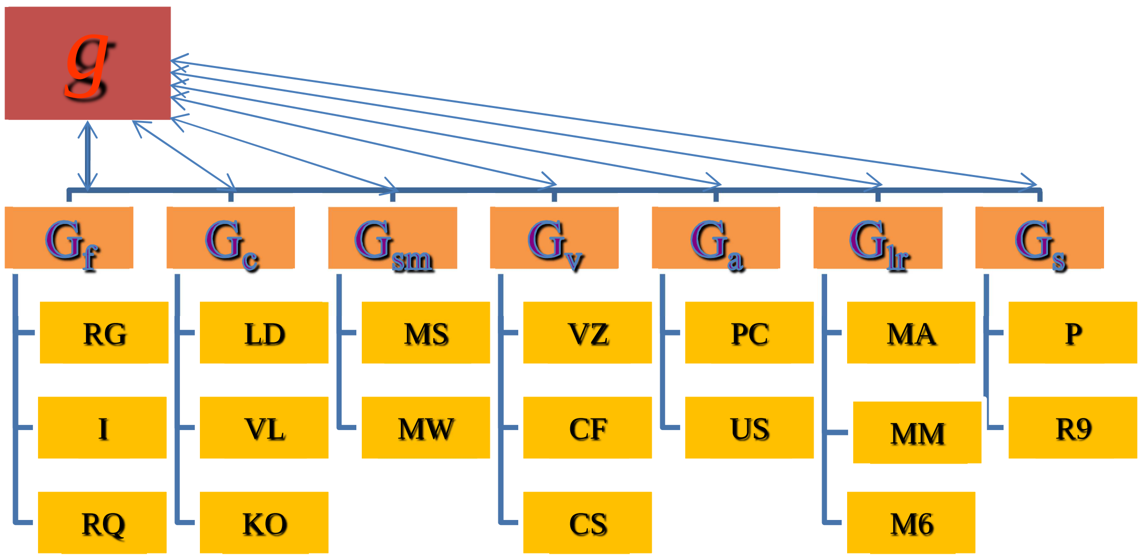
Step 3. Determine if each Broad Ability or factor composite is unitary by comparing the subtest scores that contribute to it. Again, 1.5 SDs is the cut-off criterion recommended by Flanagan and Kaufman. McGrew (2011, June 20) suggests either a formula that includes consideration of subtest intercorrelations or using standard errors of measurement to determine just what is an acceptable range. Examiners can also find base rate data in manuals for some tests for some factors. Depending upon the test, publishers' factors may overlap with the Broad Cognitive Abilities that have emerged from research in CHC theory, and if the examiner has not already begun to do so, consideration of CHC composites is appropriate. It is when we reach step 3 that interpretation begins to be more problematic.

Almost all of the time, at the end of level C testing, we have some composite ability or factor scores that represent unitary and, perforce, interpretable, constructs. Assuming that there is more than one construct that is usable, we must choose which one to use to represent aptitude (intelligence, intellectual potential). The results of current CHC theory and research can assist us; below are the Broad Abilities that have been found to be most closely related to *g*. They are arranged in order of importance to *g*, but it is important to remember that that they are correlated not only to *g* but to each other, and that they are based upon narrow abilities that are correlated with each other and which interact to support the manifestation of intelligence. All the Broad Abilities that have been defined are not shown here, but these are the Broad Abilities that appear in the major level C batteries that we use in our assessments.

The issue of which of these Broad Abilities we deem to represent "thinking" and/or "reasoning" becomes salient if we have reached step 2 or step 3 in our interpretation of the results obtained from the major cognitive battery that we are using. If we use less than the entire battery to establish whether or not an individual possesses average thinking or reasoning, what part is essential? Holyoake and Morrison (2012, p. 15) offer a definition of thinking: *Thinking is the systematic transformation of mental representations of knowledge to characterize actual or possible states of the world, often in service of goals.* Thinking and reasoning go hand-in-hand, and reasoning might be described as thinking with a purpose. At the same time, it is possible to reason without language, and they cite reasoning as the sine qua non for our species.

McGrew (March 27, 2011) points out that not only do different Broad Abilities contribute differently to *g*, but test batteries measure these Broad Abilities differently, and the contributions to *g* vary as well across different batteries for abilities that have the same name. The Broad Ability most closely correlated with *g* (aptitude, intelligence, potential) is Fluid Reasoning; with regard to the WISC-IV, measures of this are found in two subtests of the PRI (Picture Concepts and Matrix Reasoning) for students to age 11 and in the same two subtests + Arithmetic for older students. Fluid Reasoning is described as “the deliberate

CHC Theory Operationalized*



Broad Cognitive Abilities in [orange]: G_f =Fluid Reasoning; G_c =Crystallized Knowledge; G_{sm} =Short-Term Memory; G_v =Visual Processing; G_{lr} =Long-Term memory/Retrieval; G_s =Processing Speed.
Common narrow abilities in [yellow]: RG=Sequential Reasoning; I=Inductive Reasoning; RQ=Quantitative Reasoning; LD=Language Development; VL=Lexical Knowledge;KO=General Knowledge; MS=Memory Span;MW=Working Memory; VZ=Visualization; CF=Flexibility of Closure;CS=Closure Speed; PC=Phonetic Coding;US=Speech Sound Discrimination; MA=Associative Memory; MM=Meaningful Memory; M6=Free Recall Memory; P=Perceptual Speed; R9=Rate of Test-Taking. These are only some of the more than 70 narrow abilities researchers have identified so far.

*Adapted from Figure 1.4, p. 11, *Essentials of Cross-Battery Assessment*, 3rd Ed.

but flexible control of attention to solve novel ‘on the spot’ problems that cannot be performed by relying exclusively on previously learned habits, schemas, and scripts (Schnieder, W. & McGrew, K., 2012, p. 112-3)." Fluid Reasoning involves (at least) the narrow abilities of Induction, Sequential Reasoning and Quantitative Reasoning; it is clearly a measure of “thinking” and “reasoning." But thinking might include more than reasoning, especially if we characterize the latter as “purposeful thinking.” Crystallized Knowledge (WISC-IV: VCI) is generally the second most closely correlated Broad Ability to *g*; it is defined as “the depth and breadth of knowledge and skills that are valued by one’s culture (Schnieder, W. & McGrew, K., 2012, p. 122).” It includes narrow abilities of General Verbal Information, Language Development, Lexical Knowledge and Communication Ability. As it involves the transformation of knowledge (to include “last year’s Fluid Reasoning”), it also is clearly a measure of “thinking.”

If we are going to use a restricted estimate of potential, we would clearly want that estimate to reflect those abilities that are most closely related to general aptitude, so if we had to choose between a publisher's factor that included Processing Speed or a factor that included Crystallized Knowledge, to estimate potential, it is not a difficult choice. If the score for Processing Speed or Long-Term Memory/Retrieval were in the average range, but the scores for Fluid Reasoning and Crystallized Knowledge were both well below average, we would likely not conclude that the individual's potential is "average," and we would probably not write a recommendation to designate the client as a person with a learning disability or make a diagnosis of learning disorder. And while all seven Broad Abilities shown above contribute to an individual's aptitude (or to *g*), Fluid Reasoning and Crystallized Knowledge patently involve "thinking" and "reasoning" to a greater extent than do the other Broad Abilities. The other Broad Abilities, while they do not exclude thinking and reasoning, are fundamentally characterized by other attributes: memory/retrieval from memory, visual perception, auditory perception and speediness.

A sophisticated variation of the process of determining whether an individual has average ability (and is eligible for designation as learning disabled) is provided in the third edition of *Essentials of Cross Battery Assessment* (Flanagan, D., Ortiz, S. and Alfonso, V., 2013). In this latest edition, the authors provide a spreadsheet for practitioners to which weighted values for the Broad Abilities, along with a formula that reflects the intercorrelations between measures, have been attached to create an alternative composite: the Intact Ability Estimate (IA-e):

> . . . this composite is called the *Intact Ability estimate* (IA-e). For example, if the evaluator indicated that five out of the seven scores entered reflect abilities that contribute meaningfully to the individual's overall cognitive functioning . . ., the IA-e is then generated using the five "sufficient" scores. Because actual norms for such a composite are not available (or even feasible within the context of true cross-battery evaluation), a mathematical formula is used to calculate it. The formula used in the program includes use of median reliabilities of CHC broad ability scores and median inter-correlations among them and, therefore, is psychometrically sound (Appendix H).

The IA-e is obtained after a "g-value" is calculated that provides guidance as to whether or not an individual's restricted cognitive scores might be considered as representing average ability. This paradigm can be applied to all of the major cognitive batteries that school psychologists use:

> . . . the *g*-Value was created to answer this question: Is the individual's overall cognitive ability at least average when the cognitive deficit is not included in the estimate? Because individuals with SLD have specific cognitive weaknesses or deficits in combination with a number of intact cognitive abilities or cognitive strengths, total test scores, like the FSIQ, often provide a misleading estimate of the individual's intellectual capacity . . .
>
> The computed *g*-Value reflects the likelihood that the individual has at least average overall cognitive ability when the potential attenuating effects of the specific area(s) of weakness are removed . . . an individual whose *Gf*, *Gc*, *Glr*, and *Gsm* were judged to be sufficient would have a higher *g*-Value than an individual with sufficient *Gsm*, *Gv*, *Ga*, and *Gs*, even though each had the same *number* of sufficient abilities (i.e., 4). The reason for the different *g*-Values in the latter example is because abilities such as *Gf* and *Gc* have higher *g-weights* than abilities like *Ga* and *Gs*, meaning that they contribute more to overall *g* (or general intelligence) . . .(Appendix H)

So, in our example, would we rely on the overall score ($FSIQ^{CDN}$: 72) or the score for Crystallized Knowledge (85)? A composite of 85 for the VCI on the WISC-IV means that possibly 30 % of the subtest scores are "age-appropriate." Is that sufficient to declare that this individual has "average abilities necessary for thinking and/or reasoning?"

Seventy percent of the WISC scores in our example are likely less than age-appropriate. If all the other factor scores are 1.5 - 2 SDs < mean, that means that scores for Fluid Reasoning, Short-Term Memory, Visual Processing, and Processing Speed are low. If we apply Kaufman's logic to our scores, we would use the FSIQ as our estimate of aptitude; the index scores differ by only one SD, not 1.5 SDs, and, while that difference between the VCI and PRI is significant, it is not unusual (the base rate difference would be at least 14.1 % (Base Rate for FSIQs ≤ 79, Table B2, p. 188, Pearson Canada, 2004). Finally, intellectual disability is considered an exclusionary factor to consider in the determination of learning disability (Flanagan et al., 2012; DSM-5, 2013).

Profile 2

What if we had a score of 85 for Crystallized Knowledge (Wechsler: VCI) and scores for other factors that were 25 points lower? In such an instance, the differences would exceed 1.5 SDs, and base rates would all be < 10 %, even < 5 % for some differences. For this group of scores, it would be inappropriate to report either the FSIQ or the GAI. Would we use the score of 85 for Crystallized Knowledge as the estimate of potential? In this instance, the lack of cohesion across index scores or Ability scores is so large that we might conclude that the *variation* in development of this individual's cognitive abilities is unusual, and the emergence as "average" of one of the two major cognitive abilities that involve thinking and reasoning suggests that the individual from whom we obtained these scores should be considered as having average ability in at least one important feature of his or her cognitive development and should be considered as eligible for designation as an individual with a learning disability or learning disorder. Or should we?

A caveat: in neither instance have I mentioned the importance of information that is ancillary but essential to consider along with the results obtained from a level C test: exclusions for sensory issues, social-emotional issues, etc. and supporting data. For example, we would want to administer adaptive behaviour scales to parents and teachers in both of these instances. Through the histories that we gather, we determine if there are influences on an individual that would cause us to reject the notion of a designation as a student with a learning disability, and through examination of both historical and supplementary assessment data, we can determine if there is any support for using Crystallized Knowledge as the estimate of potential. Such data might involve an average score on a reading comprehension subtest or another vocabulary test or evidence from classroom assignments or scores from previous assessments. We would want to have some evidence of age-appropriate development from at least one other reliable source that encompasses or at least involves some of the same abilities that form the Broad Ability that we are considering using as our estimate of potential. If we have a single solitary ability that is average – and no other cognitive ability, on any other test, including measures of achievement and classroom performance that would be considered "average" and no exclusions (mental health, medical, social, etc.), then I suggest that it would likely be inappropriate to use the single composite score as evidence of average aptitude. But if we have uncovered a rationale for using Crystallized Knowledge, such as a history of motor or visual impairment, or vice-versa, if we are considering using the PRI, a history of language impairment, then using a single one of these Broad

Abilities is justifiable, in my opinion. The process of SLD assessment is articulated in the BCASP Best Practice Guidelines (2007) as well as in Flanagan et al. (2013).

Profile 3

If characterization of a client as having average ability in thinking and reasoning is rejected in the first instance (Profile 1) because of an overall score that supports a designation/diagnosis of intellectual disability (assuming that the rest of the information that we gathered supports that conclusion) and accepted in the second instance (Profile 2) because of very large differences in cognitive scores, how might we deal with instances in which an overall score on our primary level C battery is > 75 but < 85 and in which one of Fluid Reasoning and Crystallized Knowledge is ≥ 85? What do we do with a situation in which the overall score is in this range but the score for one of the Broad Abilities is ≥ 85 <u>and</u> *significantly* higher than other factor or ability scores? If we exclude cognitive profiles that resemble those in the first or second instance (that is, score differences or ranges are smaller than those that would occur in 10 % or less of the normative sample), the residual permutations of scores are substantially constrained. But consider a student who obtained a WISC-IV GAI of 76 and VCI of 85 and PRI of 74. For a 14 or 15 year-old, these composites (VCI and PRI) are significantly (at the .05 level) but not unusually different. Would we use the VCI of 85 to characterize the examinee as "average?" The base rate for this difference is not unusual, but an argument might be made that the significant difference means that these scores represent cognitive constructs that differ with regard to their development and that the score for Crystallized Knowledge indicates average ability/development with regard to thinking. We might be tempted to conclude that the individual has average ability. Before we would accept that conclusion, however, I suggest that we need to have some confirmatory evidence in the form of some other average performance on a measure that is influenced by or related to Crystallized Knowledge (Flanagan et al.'s IA-e might be helpful here in organizing this information). In the absence of such, I would likely be more persuaded that the GAI is the better index of ability.

In the absence of either significant differences or differences that are larger than those that reflect base rates of 10 % across composite scores for factors or abilities, it seems to me that it is best to use overall composites. In the instance of an examinee with a VCI of 87, a PRI of 87, a WMI of 80 and a PSI of 80, the FSIQ is 80. Although the GAI is 85, it seems to me that the FSIQ would be a more appropriate estimate of aptitude.

Summary

After we consider exclusions and have gathered comprehensive and sufficient data, we are drawn back to our question. Does the individual we are assessing possess **at least average abilities essential for thinking and/or reasoning**? It is clear that thinking and reasoning involves Fluid Reasoning and Crystallized Knowledge. **And/or** seems to mean that this could involve **one <u>or</u> both** of these Broad Cognitive Abilities. Restricted composites, such as the Wechsler GAI or the DAS-II GCA or the SBV Composite include both. The IA-e from Flanagan et al. (2013) is a new and possibly useful alternative. It seems to me that as we work our way through the interpretation of a cognitive battery (or batteries), we should allow ourselves to be guided by logic and theory and parsimony in the identification of aptitude. In the majority of instances, "average" *is* a comprehensive composite. Occasionally, it is a restricted composite. Sometimes, in unusual circumstances, Fluid Reasoning *or* Crystallized Knowledge might appropriately be reported as the estimate of potential. And sometimes a significant difference in scores might support that conclusion. In such an instance, it is important to consider the decline in reliability (and the necessity to increase our caution) as we narrow the number of abilities that we undertake to establish as the estimate of an individual's potential and to be as certain as we can that, if we choose to use the score of a single one of these two Broad Abilities, this emerges from a logical and defensible paradigm and that we can support the choice of a particular Broad Ability composite with reasonable (external-to-the-test battery) evidence.

Designation as an individual with a learning disability with regard to Ministry of Education criteria in British Columbia still requires consideration of average ability in relation to some form of underachievement. In text in DSM-5, while a requirement for "average" ability (except for excluding a diagnosis of intellectual disability) is not included with regard to diagnosis, "unexpected academic underachievement" is cited as a characteristic of learning disorders, and comprehensive assessment is encouraged. It seems to me that practitioners will continue to address aptitude as they designate, diagnose, and explain the difficulties that individuals may experience as they confront difficulties in their learning.

References

American Psychiatric Association. (2013) Diagnostic and statistical manual of mental disorders, fifth edition. Washington, D. C., Author.

British Columbia Association of School Psychologists. (2007). Best practice guidelines for the assessment, diagnosis and identification of students with learning disabilities. http://www.bcasp.ca/sites/default/files/uploads/LD-Guidelines-2007-Official-Version.pdf

DSM5|psychiatry.org. (May 3, 2013). http://www.psychiatry.org/dsm5

Flanagan, D. & Kaufman, A. (2009). Essentials of WISC-IV assessment, Second Edition. John Wiley & Sons, Hoboken, NJ.

Flanagan, D., Alfonso, V., Mascolo, J. & Sotela-Dynega, M. (2012). Use of ability tests in the identification of specific learning disabilities within the context of an operational definition. In D. Flanagan & P. Harrison (Eds). *Contemporary Intellectual Assessment* (pp. 643-669). Guildford Press, New York.

Flanagan, D., Ortiz, S. & Alfonso, V. (2013). Essentials of cross-battery assessment, Third Edition. John Wiley & Sons, Hoboken, NJ.

Holyoak, K.J., & Morrison, R.G. (in press). Thinking and reasoning: A reader's guide. In K.J. Holyoak & R.G. Morrison (Eds.), *Oxford Handbook of Thinking and Reasoning.* New York: Oxford University Press. http://reasoninglab.psych.ucla.edu/wp-content/uploads/2010/09/holyoakmorrison_ohtr_chap1.pdf

McGrew, K. (2011, Mar 27).IAP101 Brief #12: Use of IQ component part scores as indicators of general intelligence in SLD and MR/ID diagnosis. http://www.iqscorner.com/2011/03/iap-applied-psychometrics-101-report-10.html

McGrew, K. (2011, June 20).IAP 101 Psychometric Brief # 9: The problem with the 1/1.5 SD SS (15/22) subtest comparison "rule-of-thumb." http://www.iqscorner.com/2011/06/iap-101-psychometric-brief-problem-with.html

Pearson Canada Assessment. (2004). WISC-IV Canadian manual. Pearson Canada Assessment, Inc., Toronto.

Saklofske, D.,Weiss, L, Zhu, J., Rolfus, E., Raiford, S., & Coalson, D. (2008). Technical report #4.1.2 general ability index Canadian norms. Harcourt Assessment. http://pearsonassess.ca/hai/images/sample-and-technical-reports/wisc-iv-technical-report-number-4-4-1-gai-with-canadian-norms.pdf

Schnieder, J. & McGrew, K. (2011). “Just say no” to averaging IQ subtest scores. Institute for Applied Psychometrics, Applied Psychometrics 101, 2011-03-24, #10. http://www.iqscorner.com/2011/03/iap-applied-psychometrics-101-report-10.html

Schnieder, J. & McGrew, K. (2012). The Catell-Horn-Carroll model of intelligence. In D. Flanagan & P. Harrison (Eds). *Contemporary Intellectual Assessment* (pp. 99-144). Guildford Press, New York.

Special Education Services: A Manual of Policies, Procedures and Guidelines - March 2011. Ministry of Education of British Columbia. http://www.bced.gov.bc.ca/specialed/ppandg.htm

Tellegen, A. & Briggs, P. (1967). Old wine in new skins: grouping Wechsler scales into new scales. *Journal of Consulting Psychology*, 31 (5), 499-506.