

Intellectual Disability Assessment: Anomalous Picture Concepts Scores in WISC-IV

Michael G. King*

Sunshine Health and Dental, 429 Ballarat Rd, Sunshine, 3020, Victoria, Australia

Abstract: Attention is drawn to a putative WISC anomaly, specific to the case of Intellectual Disability (ID) IQ assessment. With ID, at best the WISC-IV Picture Concepts (PCn) is not a consistent part of the underlying “ability” represented by Perceptual Reasoning Index (PRI) factor, while at worst the PCn score may be “wrong”. This worst case situation is marked by a “very high PCn score” in an otherwise low profile and this outcome awaits a better interpretation which would be meaningful to the target child with Intellectual Disability. The literature relating to PCn scores in cases of “intellectually different” cohorts shows conflicting published evidence that the sub-test may be either generally lower than the cohort average, or generally above that average, for example in cases of Autism. The present paper proposes that the broad explanation for the PCn anomaly is explained by deriving an answer to each item from a focus on single-feature solution rather than an integrative approach. Future research may clarify the frequency and the meaning of this issue, while the present paper raises a caution in the “standard” interpretation of intellectual disability IQ profiles.

Keywords: WISC-IV, Picture Concepts, Intellectual Disability, Criticism, Sub-test Anomaly, Cohort IQ profile, Perceptual Reasoning, Autism.

INTRODUCTION

The present paper focusses upon anomalous scores on one of the core sub-tests of the current Wechsler Intelligence Scale for Children (WISC-IV). The issue raised is not based upon a wide-ranging and statistically robust study, but upon case by case results in “real world” practice. The purpose of the present report is to firstly stimulate other practising professionals to examine their own results for this possible (otherwise overlooked) distortion – potentially an issue which requires discussion on a case by case basis. The second goal is to suggest a theoretical perspective (a visual information processing model) for this particular anomaly. The third goal of the drawing attention to this issue is to encourage careful and systematic study by those positioned to conduct such projects (universities and the like). Restricted as it is to these three goals, a broader and all-encompassing review of cognitive assessment on the intellectually disabled, or the discussion of syndrome-specific profiles, is consciously not encompassed. The author seeks to make the simple point: *there seems to be a problem, based upon case by case outcomes; a candidate explanation may be found in well-established visual information processing models; further research is encouraged to establish frequency and severity rates of this anomaly among different cohorts of children who generally score low on other IQ tasks.* Finally, noting that there is apparent high face validity to the groupings (“indices”) of sub-test scores, it

is important to pause and reflect on the opportunistic nature of factor analysis, and to remember that for particular groups other than “the norm”, and even more so individual students who are far from “average” in their scores, these people are not necessarily well described by, nor educationally assisted by, population-derived factors.

From the beginnings of mental assessment, the measurement of cognitive functioning among the Intellectually Disabled (ID) was always seen as a core task. However to an increasing extent the statistical and interpretative underpinning of intelligence tests is based upon large cohorts of relatively “normal” subjects. This is of course reasonable, as it is necessary to know precisely the nature and extent of “average” in order to define “disability”, but the present paper raises a caution about the interpretation of intelligence measures that may make sense “in general” but may not be so meaningful in the case of intellectual outliers. Those active in the field have repeatedly shown that although group data show a “flat” profile, it is common to find among individual ID subjects an atypically high result on one or more sub-test scores [1]. As interpretative comment, Bergeron and Floyd [1] comment that only *part of the score variation can be explained by regression toward the mean . . .* and they further elaborate: *It is unclear [if it is the nature of the cognitive demand itself] . . . or some other influence that contributes to this pattern.*

The present paper focuses on one particular test (Picture Concepts, WISC-IV) and (in line with and Bergeron and Floyd [1]) raises a caution about

*Address correspondence to this author at the Sunshine Health and Dental, 429 Ballarat Rd, Sunshine, 3020, Victoria, Australia; Tel: +61 3 93116909; Fax: +61 3 93116819; E-mail: mgking@tpg.com.au

interpreting high scores as evidence of specific strength in the student. The WISC-IV is recognised as the current evolution of a continuously developed and refined measure of the concept of general intelligence. Over the various editions of WISC, some sub-tests remain essentially unchanged (although normative data may be seen to drift slightly, typically towards higher absolute scores for a given standardized level). Other sub-tests may be eliminated altogether (for example Object Assembly from WISC-III) or may be assigned to the “supplemental” category and have their position as “core” components supplanted by others. One such relegated test is Picture Completion (which retains a relatively high correlation with overall intelligence, and also bears a strong relationship with the “visual” factor, Perceptual Reasoning Index). A new test, Picture Concepts (PCn) took the place of Picture Completion. Wechsler notes that the changed name of the Perceptual grouping (and implicitly its content as well as its title) *reflects the increased emphasis on fluid reasoning abilities in this index* [3] (p 6).

When a particular sub-test is downgraded from being “core” to “supplemental” (or totally removed) then it is a credible assumption for the test user to accept in good faith that the change made is for the better. The present proposition for discussion is based upon a combination of real world experience plus perusal of the available statistics, and contrary to the above-mentioned assumption, it is suggested here that Picture Concepts (PCn) can give anomalous results. It is further posited that, rather than illuminating some specific strength (or weakness), the recognition of which may add to our understanding of the student’s intellectual profile or educational potential, the PCn test is simply misleading. That is to say, the PCn score has an unacceptable risk of leading to what I call a type 3 error: *you are wrong*¹.

FACTS ABOUT FACTORS

“Factor analysis was originally invented by Spearman in order to conceptualize the nature of

intelligence. Its underlying assumption is that observed variables are the functions of latent variables called factors and its goal is to identify these factors” [2].

For the present discussion it is necessary to conceptually un-bundle the factorial groupings of sub-tests – the so-called “Indices”. Since these Indices are “factor derived constructs”, straight away then, before teasing out the interpretation of an individual sub-test, it is worth revisiting the nature of factor analysis. There are two broad ways to view the meaning of a “factor” (whether in psychometrics or any other domain):

- (a) there is the hope, the expectation, even the belief that the grouping derived from factor analysis actually relates to some underlying and omnipresent quality².
- (b) there is, however, a real possibility that the factors discovered by factor analysis do not actually “exist” as some sort of latent trait (although that is the way factors are often understood).

Accepting “b”, it could be argued that a factor may be nothing more than the inevitable result of offering to the computer one group of values or scores which are based upon measurements which are in themselves “a bit similar”, and at the same time these “group 1 items” are balanced against another group (or groups) of measures which are credibly “less similar” than the to-be-discovered first group/factor. That is the way to create a good strong factor: put another measure into the mix which is not similar to that first “factor”. And by definition of the process of “explaining the variance” of the whole, you can see what the factor analysis process will inevitably do. This issue of the “real meaning” of psychometric factor analysis is further expanded elsewhere [4].

RECONCEPTUALISING PICTURE CONCEPTS

Now consider the WISC sub-tests and putative factors: the factors in WISC have never existed as “omnipresent” or “ubiquitous” groupings. An extensive trail of studies reaching far back into the past has always shown that the WISC factors are merely groupings of convenience which are relatively robust

¹Type 3 Error: not wanting the credit for this concept, a colleague of mine, several decades ago, used to delight in offending (appropriately knowledgeable) people at social functions when in the middle of a artificially heated dispute on some esoteric topic he would pounce and say “ah ha. In this respect you are making a type 3 error”. The mystified accusee would say quizzically “type 3 error?” to which my colleague would say “yes. A type 3 error – you **are** wrong”. And again a decade or two later I came across a paper written by a Belgian academic (as far as I can remember) which did invoke the concept of the type 3 error as roughly an indication that “you are wrong” and this in a peer reviewed, if minor, journal - but I have since mislaid this paper and its reference. Type 3 errors are a great idea, but I cannot take the credit for this notion.

²an example of “probably genuinely existing underlying traits” can be found in the qualities of “anxiety” and of “introversion/extraversion” which are so regularly isolated in personality measures regardless of the array of items offered, that it is credible to assert that these are indeed two genuine and separable qualities in humans.

Table 1: Intercorrelations for all Ages. (corrected coefficients, removing effect of sub-test from the factor) (Wechsler, 2003, p 51)

Sub-test	BD	PCn	MR	PCm
Picture Concepts PCn	.41			
Matrix Reasoning MR	.55	.38		
Picture Completion PCm	.54	.39	.46	
Perceptual Reasoning Index PRI	.56	.50	.61	.57

Table 2: Scaled Scores: Detail from Table 5.36 (Wechsler, 2003)

Subtest	Mean Asperger's	Mean control	p(difference)
Block design	10.9	11.9	.21
PCn	8.7	10.8	.03
MR	10.7	10.9	.80
PCm	11.5	11.8	.61

for “normal” students, and which are equally “non-robust” (meaning that the standard nominated groupings may not exist) for defined groups of “differently-abled” (or simply “disabled”) students. That assertion is not dismissible statistical heresy – it is published fact. Just flicking through the WISC technical manual, almost at random, the detail in Table 2 shows even to the naked eye (unassisted by factor analysis) that the target of the present discussion (PCn) is not well-related to some implied Perceptual factor – in fact would probably not be located on that factor at all in an even handed analysis of (the referred-to) Asperger's students. Interestingly Wechsler's Asperger's students' impoverished PCn scores [3] were not reflected in a recent study of high functioning Autism students (a group understood to replace the “Aspergers” label in DSM-V): *the new WISC-IV . . . visual reasoning subtests (Matrix Reasoning and PCn) were the highest of the nonverbal subtests* [5] (p 428).

Search through the WISC manual and discover further examples of anomalous sub-test scores, implicitly leading to “different” factorial groupings for various categories of “intellectually disabled” students. Since we know that even considering our ID clients as a group these factors do not apply to them, then going further down the logical spiral – what credibility is there to the proposition that the individual student sitting in your room might be “exactly like” the average Asperger's, or Motor Impairment or Reading Disabled, or other grouping of Intellectually Disabled, student? While it is arithmetically possible to compute the mean

score of any set of numbers, including the mean sub-test score of an ID group of test takers, the reverse process of making inferences about an ID individual, based upon group means, is problematic [1].

The inevitable outcome of factor analysis (a point which often escapes test developers) is this:

- a. two strong factors are inevitably formed by putting in two groups of measures which are “relatively different” and thus much of the total variance will be explained by that grouping;
- b. if the items on only one factor were offered to the computer, then in the absence of a “relatively different” batch of measures in the mix, the single “strong factor” structure will likely fall apart.
- c. In a set of data with two (more or less) clear different groups, in addition to your “two strong factors” there will always be a third “weak” factor (and a few more even weaker ones if you let the computer have its own way). These weaker factors clean up (“explain” in statistical parlance) progressively less and less of the variance. Now if you look at the “third” factor, ponder on what it might be talking about, then if you take the next “obvious” step and add in a few more items which seem to relate to this third concept, your repeat analysis will produce three strong factors and a fourth weak one. You could go on forever adding items to strengthen the weak factors,

Table 3: Examples of Apparently Spuriously High PCn Scores

	Male 13y 2m Previous assess age 9y11m	Female 12y10m	Female 11y 6m
	Verbal IQ 67		
	Performance IQ 71		
	Full Scale IQ 67		
WISC-IVTest	Scaled Score		
Block Design	5	5	2
Similarities	5		
Digit Span	3		
Picture Concepts	14	15	12
Coding	10		
Vocabulary	8		
Letter-Number Sequence	7		
Matrix Reasoning	3		2
Comprehension	6		
Symbol Search	6		
Picture Completion		1	3

slightly changing the underlying nature of the already-discovered factors, and thus discovering more factors³.

When this statistical inevitability is applied to the individual student sitting opposite you⁴, the student who is by definition not part of the “general population” upon whom the factors were derived, then it becomes appropriate to consider the sub-tests individually, and the scoring of the indices (based as they are on factor analysis of “normal” students) is merely a hindsight exercise: if the Indices seem coherent then you can use them, if they do not seem coherent then you should assume that they don’t apply to this student. Future essays may discuss other factorial facets of the WISC-IV, but for the present, taking the Picture Concepts (PCn) subtest, it is proposed:

1. PCn is more poorly correlated with the Perceptual Index than is its predecessor
2. PCn is more poorly correlated with overall IQ than is its predecessor

3. The lower correlations are not necessarily an indication that PCn is bringing some “additional quality” into the mix – rather it is of concern that the weaker correlation is due to increased error variance.
4. The examples listed in Table 3 show that some students of very low intelligence (meaning that their other scaled scores more or less uniformly low, including those of the Perceptual grouping), some of these students can score anomalously high on PCn, and (in the writer’s experience) this can happen sufficiently often as to cast doubt upon the validity of PCn as a useful measure unless an entirely different interpretation is offered.
5. In the absence of a recognised “savant” quality among students for whom the task of PCn becomes easy, then an alternative explanation for the anomalously high scores is sought.
6. It is tentatively proposed here that all PCn items, including the supposedly more difficult ones later in the test, can be solved by “single feature detection” (a visual information processing modality long since described [6]) and that the later items in the task are statistically “difficult” due to confusing or distracting qualities (see for example Kent *et al.* [7]), and that these additional confusing features may only be obvious to those students of medium level intelligence. The underlying task – the search for a simple single

³“with the publication of the WISC-R, **factor analyses revealed the presence of a third factor** separate from the verbal and performance factors, calling into question the two-part model of intelligence. During the development of the WISC-III, therefore, an additional subtest was developed to attempt to clarify the factor structure of the test, leading to the emergence of four factors, though the venerable Verbal and Performance IQs remained. The most recent, fourth edition (WISC-IV; 2003) has embraced and expanded **the four factors**, leading to the abandonment of the Verbal and Performance dichotomy in favour of four scores.” [2] (p 588, **emphasis added**)

⁴“sitting opposite? despite the structure of the administrative tools, I actually prefer the Rorschach-suggested model of sitting beside the student.

quality which is present in all three rows – remains childishly simple, which is not the expectation of the more intelligent test-taker.

CONCLUSION

Although not intended as a criticism of the WISC-IV, its subtests, its indices or its authors, the present essay raises a credible note of caution about the process of “discovering” new factors in any area of endeavour – making the tautological (but sadly ignored) observation that factor analysis of any data set will always produce at least two strong factors if there are two “somewhat different” groups of items, and there will always be an additional “weaker” factor which seems to beg the addition of new items in the next edition. The Intelligence Measurement community should be very cautious about any fifth factor coming out in future IQ editions.

The second point made is that the recently added Picture Concepts sub-test seems, even by the figures presented by the test maker, to have a somewhat erratic relationship with the presumed underlying construct. The “true meaning” of PCn may await further research, but the alert is raised that on some occasions spuriously high scores can be attained. It has already been noted that for ID students there may often be one or another relatively higher score, but the suggestion made here (based upon real world experience) is that the anomalous high score happens far too regularly on PCn. A very technical explanation of these apparently anomalous high scores may be found in the musings of visual search-detection processes, but regardless of the accuracy of that particular explanation (point 6, above) it remains the empirical experience of the writer that anomalously high scores on PCn do occur. The really important question is whether this sub-test has explaining power in relation to real-life experiences of children with generally diminished IQ diminished scores. And finally it is of particular importance that the assessor does not arithmetically combine a high score

on PCn with other, much lower, Perceptual Reasoning scores to arrive at an “average” figure which is reflective of neither the high nor the low sub-test scores.

Although it is accepted that occasional “unexpected” scores occur across the WISC array of tests, it is not evident that a systematic anomaly with PCn has been previously mooted, much less “explained”. Since the anomaly flagged by the present paper is that of “higher than expected” scores, and particularly that these higher scores may be most evident (and potentially most damaging in terms of their classification by overall IQ) in the case of ID clients, it arguably behoves our community to at least discuss this proposition, to share supportive (or non-supportive) experiences, and to develop a useful interpretation of this sub-test in the area of understanding the cognitive processes of children with intellectual disability.

REFERENCES

- [1] Bergeron R, Floyd RG. Individual part score profiles of children with Intellectual Disability: A descriptive analysis across three intelligence tests. *School Psychol Rev* 2013; 42(1): 22-38.
- [2] Cohen A, Fiorello CA, Farley FH. The Cylindrical structure of the Wechsler Intelligence Scale for Children — IV: A retest of the Guttman model of intelligence. *Intelligence* 2006; 34: 587-91.
<http://dx.doi.org/10.1016/j.intell.2006.05.003>
- [3] Wechsler D. WISC-IV Technical and Interpretative manual. The Psychological Corporation. San Antonio 2003.
- [4] King MG. Est-ce que c'est vous-même dans la glace? - the Phenomenology of Rorschach Expressive Constraint. *Bull Soc Sci Med* 2008; (2): 295-305.
- [5] Mayes SD, Calhoun SL. WISC-IV and WIAT-II profiles in children with High-Functioning Autism. *J Autism Dev Disord* 2008; 38: 428-39.
<http://dx.doi.org/10.1007/s10803-007-0410-4>
- [6] Treisman A, Gelade G. A feature integration theory of attention. *Cognitive Psychol* 1980; 12: 97-136.
[http://dx.doi.org/10.1016/0010-0285\(80\)90005-5](http://dx.doi.org/10.1016/0010-0285(80)90005-5)
- [7] Kent C, Howard CJ, Gilchrist ID. Distractors slow information accumulation in simple feature search. *J Vision* 2012; 12(1): ArtID 13.
<http://dx.doi.org/10.1167/12.1.13>

Received on 20-06-2013

Accepted on 12-09-2013

Published on 28-10-2013

DOI: <http://dx.doi.org/10.6000/2292-2598.2013.01.01.1>

© 2013 Michael G. King; Licensee Lifescience Global.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.