Cognitive Profile Patterns Are Affected by Measurement Precision

Sorel Cahan[1], Hendrik Jurges[2] & Jenni Hannin[1]

[1] School of Education, The Hebrew University of Jerusalem, Israel

[2] Schumpeter School of Business and Economics, University of Wuppertal, Wuppertal, Germany

Correspondence: Sorel Cahan, School of Education, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem, 91905, Israel.

Abstract

A necessary, albeit tacit assumption underlying pattern analysis of cognitive profiles is that an examinee's profile pattern is not affected by the level of precision used in measuring the subtest, index or factor scores. We empirically test the truth of this assumption across various precision levels, such as IQ points (1/15SD), T-scores (0.1SD), scaled scores (1/3SD) and stanines (0.5SD). The results clearly refute the pattern stability assumption. They question the very uniqueness of profile patterns as a stable individual characteristic and challenge their use in both clinical practice and scientific research. Possible solutions are suggested and critically examined.

Keywords: cognitive tests, subtest scores, profile pattern, measurement precision level

1. Introduction

1.1 Cognitive Patterns

Cognitive tests, such as the Wechsler Intelligence Scale for Children (WISC-V; Wechsler, 2014) typically yield, in addition to the total, overall IQ score, a series of subtest, index or factor scores, often referred to as the individual's cognitive "profile" (Angoff, 1971). Since these scores are expressed on a common (usually standardized) metric, they invite ipsative comparisons between an individual's scores on different subtests or indexes (e.g., Bolen, 1998; Naglieri & Paolitto, 2005). For any two scores, A and B, the question is whether A>B, A<B, or A=B. Taken simultaneously, the results of these comparisons determine the "pattern" of the individual's cognitive profile, that is, the within-individual rank order of the scaled subtest or index scores. Given the trichotomous nature of the magnitude relation between any two scores (see above), the number of theoretically possible different profile patterns depends on the number of subtest scores included in the profile:

$$g(n) = \binom{n}{0} g(0) + \binom{n}{1} g(1) + \cdots + \binom{n}{n-1} g(n-1) \tag{1}$$

Where: $g(n)$ is the number of possible profile-patterns created by n subtest scores and $g(0)=1$. According to this formula, two subtest scores can yield three distinct profile patterns, three subtests can yield 13 possible distinct profile-patterns[2], four subtests result in 75 potential patterns and five subtests can yield close to 10,000 (!) different profile-patterns.

Importantly, profile pattern is independent of the profile's elevation (the average score across subscales) (Cronbach & Gleser, 1953; Watkins, Glutting, & Youngstrom, 2005). Two profiles can have the same elevation, but different patterns and vice-versa (Watkins et al., 2005). Thus, profile-patterns provide rich and valuable information regarding individuals' relative cognitive strengths and weaknesses, beyond that provided by the total test score. That explains their frequent use in psychological diagnosis, counseling and therapy (e.g., Bergeron & Floyd, 2006; Bolen, 1998; Gati, Landman, Davidovitch, Asulin-Peretz, & Gadassi, 2010; Robinson & Harrison, 2005; Willner, Gati, & Guan, 2015; Wright et al., 2013).

In addition, profile patterns have been the subject of extensive empirical research aimed at the identification of patterns that are shared by the majority of the individual members of a particular group (i.e., "typical" patterns) and vary between groups, where groups are defined by sociodemographic or psychological characteristics. For example, one of the first studies (Lesser, Fifer, & Clark, 1965) reported considerable differences in cognitive patterns between children from different ethnic backgrounds: "The Chinese children were very strong in space

and considerably weaker…on verbal ability…the verbal skills of the Jewish children were clearly superior to…their own abilities in other intellectual areas…the verbal skills of the Puerto Rican children were the weakest of all their mental abilities" (Lesser et al., 1965, pp.71–72).

Distinct cognitive patterns have also been reported for intellectually gifted children (Raiford et al., 2016), for children with low academic performance (Souroulla and Panayiotou, 2017), for children with learning disabilities (McDiarmid, 1993; Toffalini, Giofrè, & Cornoldi, 2017), autism (Nader, Courchesne, Dawson, & Soulieres (2016), and ADHD (attention deficit with hyperactivity; Dickerson-Mayes, Calhoun, Chase, Mink, & Stagg, 2009; Hale et al., 2012; Mayes, Calhoun, Mayes, & Molitoris, 2012). Furthermore, and more important from the perspective of differential diagnosis, different profile patterns have also been reported for children with different pathologies (e.g., Fiorello et al., 2008); for example, between children with ADHD and children on the autistic spectrum (Zayat, Kalb, & Wodka, 2011).

1.2 Stability of Cognitive Patterns Across Measurement Precision Levels

A necessary, albeit tacit assumption underlying pattern analysis of cognitive test scores is that an examinee's profile pattern is not affected by the arbitrary level of measurement precision used, that is– the magnitude of the unit used for measuring the various scores. In other words, it is assumed that the pattern obtained using a given measurement unit–usually a specific fraction of the population standard deviation (*SD*), such as $1/15SD$ (IQ point), $0.10SD$ (T scores), $1/3SD$ (scaled scores) or $0.5SD$ (stanines)–will remain invariable if the specific measurement unit used is replaced by a different unit, be it larger or smaller[3].

If this critical assumption does not hold empirically true, then an examinee's profile pattern might be specific to the particular measurement unit used, rather than a stable individual characteristic and, therefore, not generalizable across measurement units. Use of a different, however equally legitimate unit of measurement could yield a different pattern. Pattern invariance across different degrees of measurement precision (or measurement units) is thus a *sine qua non* condition for the meaningfulness of pattern analysis of cognitive score profiles and their valid use in both practice and research.

Surprisingly, however, the literature is silent regarding this crucial issue. An extensive search of the theoretical and empirical literature in the field found no theoretical discussion of the pattern invariance assumption nor empirical studies that examined it. This work aims at filling this gap; that is, to empirically test the invariance assumption at the basis of pattern analysis of cognitive profiles and discuss the implications of our results. For this purpose, we make use of the rich database built in the framework of the recent research project, titled: "Determinants of Cognitive Development in Deprived Environments: Evidence from the West Bank".

2. Method

2.1 Target Population and Sample

The target population consisted of all fifth to ninth grade students attending gender-segregated public schools in the Palestinian West Bank, and included all the above-mentioned grade levels in the 2012–13 school year. A stratified random sample of 100 schools (out of the total population of 463 schools) was selected from the three educational regions in the West Bank (north, center and south). From each sampled school, 60 students were randomly selected (12 students from each grade, 5th–9th). Thus, the planned size of the sample of students was 6000 students, 1,200 per grade. Of these, 5725 (95%) completed the entire battery of tests. The participation rate was stable across grade levels and did not vary considerably between schools.

2.2 The Cognitive Test

Cognitive development was measured by the cognitive ability test battery used by Cahan and Cohen (1989). The battery consists of 12 subtests and a total of 178 items, covering a wide range of content (e.g., analogies, series, sentence completion, vocabulary) and varying in the nature of their items (verbal, numerical, and figural), selected from well-known tests of general ability: The Cognitive Ability Test (CAT; Thorndike & Hagen, 1971), the Lorge-Thorndike Test (L-T; Lorge & Thorndike, 1954), Standard Progressive Matrices (SPM; Raven, 1983), and Cattell and Cattell's Culture Fair Intelligence Test (CFIT; Cattell & Cattell, 1965). All of the verbal tests were translated into Arabic and adapted for the population.

Test administration. School counselors administered the entire test battery to the students as a group, at school, during May 2013. The tests were given in a fixed order in a single, two-hour session with a 15-minute break. Students were given general explanations about the item format and the response sheet. In addition, the administration of each subtest was preceded by a short explanation and an illustrative example of the particular task.

Cognitive test scores. For each participant we computed the 12 subtest scores. In order to allow for between-test comparability, the subtest scores were standardized separately within each grade and expressed in within-grade standard deviation units (i.e., *z*-scores). The inter-correlations between the 12 subtests are substantial: They range from .33 to .64 (median = approx. 50; see Appendix A).

2.3 Determination of the Participants' Cognitive Profile Patterns

The first step in our analysis consisted of determining each participant's initial score profile. For the sake of simplicity, the analysis is based on three-score profiles, where the number of theoretically possible patterns is manageable (13). The number (C) of different *k* subtest combinations irrespective of order out of *n* subtests is:

$$C_{(n,k)} = \frac{P_{(n,k)}}{k!} = \frac{n!}{k!(n-k)!} \equiv \binom{n}{k}$$

In our case, where there are *n*=12 subtests and *k*=3, C=220. Consequently, for each participant in the study we determined the test score profile for each of the 220 different three-subtest combinations. For each combination, the initial pattern of each participant's profile was determined based on the magnitude relations between the three infinitesimally precise scores provided by the computer.

In order to empirically examine the stability of the participants' profile patterns across varying degrees of measurement precision, we recoded these initial scores to achieve six decreasing levels of measurement precision: 1/15*SD* (1 IQ point), 0.1*SD* (1 T-score), 1/3*SD* (1 scaled score), 0.5*SD* (1 stanine), 2/3*SD* (10 IQ points) and 0.8*SD* (12 IQ points). The procedure for determining the profile pattern was then repeated for each of these precision levels. For each examinee, the end result was six additional profile patterns of the same three scores, corresponding to the six additional measurement precision levels.

2.4 Within-Individual Stability of the Profile Pattern

For each examinee, the final stage of the analysis consisted of pairwise comparisons between her seven profile-patterns, which only differ in the precision of the scale used to measure the three scores. Based on this comparison, each profile pair was defined as either "identical" or "different". For each of the 220 possible three-subtest combinations, the sample percentage of examinees with identical patterns for any given pair of precision levels indicates the profile patterns' degree of stability or invariance across the respective two measurement precision levels.

3. Results

The results are presented in Table 1. Each cell in Table 1 gives the range (in parentheses) and median (in bold) of the sample percentage of examinees whose profile patterns were identical for the particular pair of measurement precision level across the 220 three-subtest combinations. As evident in Table 1, the profile-patterns' stability decreases the more distant the two precision levels involved in the comparison. Median pattern stability is highest (91%) for the infinitesimal–1/15*SD* pair and minimal (26%) for the infinitesimal–0.8*SD* pair. For measurement precision pairs in the range 1/3*SD*–2/3*SD*, the stability figures range between 53% and 64%.

Table 1. Percentage of stable three-score profile patterns between measurement precision level Pairs

	Measurement precision						
	infinitesimal	1/15*SD*	0.1*SD*	1/3*SD*	0.5*SD*	2/3*SD*	0.8*SD*
infinitesimal	-						
1/15*SD*	**91** (86-96)	-					
0.1*SD*	**87** (82-90)	**92** (89-96)	-				
1/3*SD*	**61** (54-65)	**68** (63-73)	**70** (65-74)	-			

0.5SD	**46**	**53**	**57**	**64**	-		
	(40-52)	(45-60)	(51-64)	(50-75)			
2/3SD	**34**	**40**	**43**	**53**	**61**	-	
	(29-40)	(34-45)	(36-49)	(47-63)	(52-69)		
0.8SD	**26**	**31**	**34**	**48**	**57**	**66**	-
	(19-31)	(24-37)	(26-40)	(40-55)	(50-64)	(58-74)	

Furthermore, pattern stability for the same pair of precision levels is lower, the larger the profile's size (i.e., number of scores). This is illustrated in Figure 1 for three and four-score profiles. The figure presents the joint distribution of the percentage of identical patterns for all pairwise comparisons between the seven precision levels in Table 1, for these two profile sizes. The coordinates of each point in Figure 1 are the percentage of identical patterns found for three-score and four-score profiles (the horizontal and vertical axes, respectively) for the particular pairwise combination of measurement precision levels[4]. As indicated by Figure 1, all the points are below the main diagonal (the "identity line"). That is, for each pair of measurement precision levels, the percentage of identical patterns for four-score profiles is lower than the corresponding figure for three-score profiles.
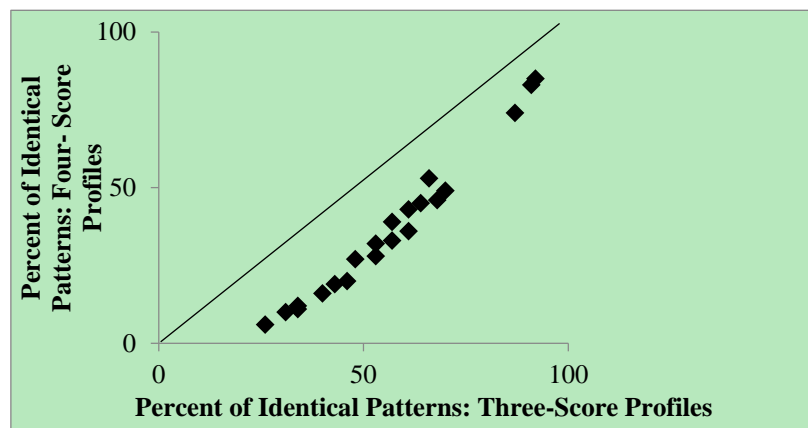

Figure 1. Joint distribution of the percentage of identical patterns for all pairwise comparisons between the seven precision levels in Table 1, for three and four-score profiles

4. Discussion

The considerable instability of cognitive profile patterns under change of measurement precision, highlighted by our results, questions the very uniqueness of such patterns as a stable individual characteristic and challenges their use in both clinical practice and scientific research. It is important to stress that the concerns it raises are independent of, and additional to other concerns raised in the literature, which focus on the unreliability of profile-patterns, their temporal instability, their low predictive validity and lack of unique informational contribution (e.g., Glutting, McDermott, Watkins, Kush, & Konold 1997; Jensen, 1998; McDermott, Fantuzzo, & Glutting, 1990, 1992; Watkins, 2000; Watkins & Canivez, 2004). In fact, they precede other concerns: The very existence of unique "objective" and invariant cognitive profile patterns is a necessary pre-condition for the discussion of their reliability, validity or usefulness to be meaningful. In the remaining of this section we (1) examine the causal mechanism underlying the patterns' instability across different degrees of measurement precision found by our study; and (2) explore some of its implications.

4.1 The Effect of Decreasing Measurement Precision on Profile Patterns

What explains the pattern instability found in our study? We suggest that the answer to this central question is simple and straightforward; in fact, almost self-evident: Decreasing precision in the scoring of the dimensions measured by the scores comprising the examinees' profiles–that is, increasing the size of the unit of measurement used–increases the chances of equality between any two profile scores. As a result, the initial

inequality between two scores, based on a higher degree of measurement precision (i.e., a smaller unit of measurement) will vanish if the difference between the two scores is smaller than the new and larger unit of measurement. For example, if the difference between two scores is 0.1SD, then it would be counted as an inequality if it is larger than the unit of measurement used (e.g., 1/15SD = 1 IQ point), and as equality if the unit of measurement is increased to 0.2SD (3 IQ points).

According to this explanation, decreasing measurement precision should result in increasingly higher relative frequency of patterns consisting of equality between pairs of scores. For example, in the simple case of profiles consisting of only three scores, decreasing measurement precision should lead to more profiles exhibiting one or two equalities. These equalities substitute the original inequalities obtained with higher measurement precision. As evident in Figure 2, this expectation is strongly supported by empirical results. The figure illustrates the increasing relative frequency of three-score profiles including one or two equalities with decreasing measurement precision in our study. The leftmost column in Figure 2 gives the frequency distribution of the number of equalities when measurement precision is maximal, that is, the unit of measurement used (the default provided by the computer) is infinitesimal. As illustrated by Figure 2, in this extreme case, all the empirical profiles in our study include no ties, that is, no equality relations between subtest scores. Lowering the measurement precision results in increasing frequency of profiles including one and even two equalities, which replace the original inequalities. Thus, for instance, when the measurement unit is increased to 0.2SD (instead of the practically 0SD in the infinitesimal case), the profiles of about 20% of the participants change and include one or even two equalities. Further increase of the measurement unit to 0.5SD (the unit of measurement in the stanine scale) causes profile change for about half the study participants, who now exhibit one or two equalities (see the rightmost column in Figure 2).
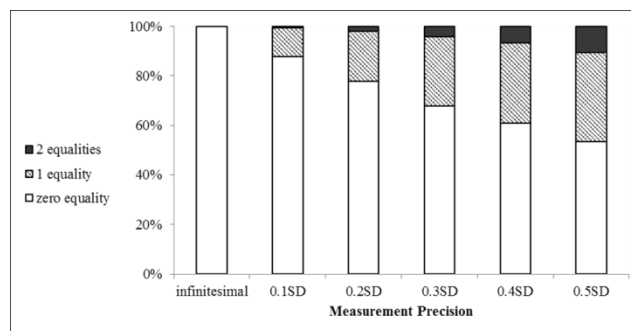

Figure 2. The relative frequency of ties (equality between pairs of a profile's scores) as a function of measurement precision in the three-subtest scores case

Figure 3 illustrates the mechanism of this process. The figure provides the cumulative frequency distribution of the three categories of profile patterns in terms of number of equalities (0,1, and 2) that can be obtained in the three-subtest scores case, across the seven illustrative measurement precision levels in Table 1.
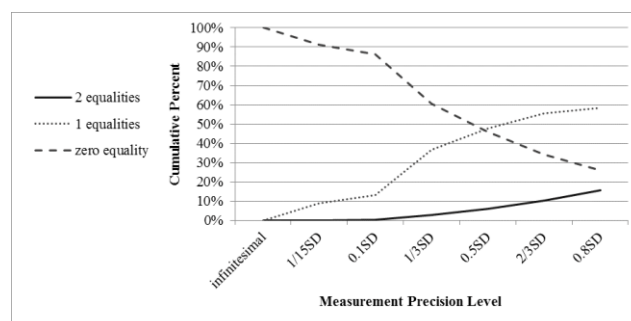

Figure 3. Percentage of zero, one and two equalities between the profile's three scores as a function of measurement precision

As evident in Figure 3, the lower the precision, the higher the relative frequency of patterns including one or two equalities, at the expense of patterns including two inequalities. We interpret these results as supporting evidence for the validity of the suggested explanation of the effect of decreasing measurement precision on the stability of cognitive profile patterns.

4.2 Conclusions and Implications

The instability of the profile patterns of a significant fraction of examinees under change of measurement precision may lead to the conclusion that pattern analysis of cognitive test scores should be abandoned, a conclusion that has been frequently voiced on other grounds, such as subtest scores' unreliability (Bolen, 1998; Watkins et al., 2005), patterns' lack of temporal stability (Livingston, Jennings, Reynolds, & Gray, 2003; Watkins & Canivez, 2004), predictive validity (Watkins and Glutting, 2000; Watkins, Kush, & Glutting, 1997) and informational contribution (McDermott et al., 1992; Moses & Pritchard, 1996). The continuing use of pattern analysis in both clinical practice and research, despite the repeated arguments for its abandonment, however, justifies the call for a pragmatic and partial solution of the patterns' dependence on the measurement precision level, a solution that might have better chances of being adopted. We suggest that such a solution exists, at a price. It consists of avoiding the instability problem by "begging the question", that is, by dispensing with the requirement of pattern stability across measurement precision levels and stipulating instead an arbitrary, however agreed-upon, mandatory precision level (or unit of measurement). The many problems associated with this solution are clear and self-evident. Therefore, we only focus on possible arguments in its favor. A key argument involves the concept of a minimal psychologically significant difference. According to this argument, not any non-zero difference between two scores (such as $0.000001SD$) is psychologically significant. Rather, in order to qualify as psychologically significant, a difference between two profile scores has to exceed a critical threshold (Δ_C), such as $0.1SD$, $0.25SD$ or $0.5SD$. The logic of this argument is identical to that underlying the "effect size" argument in experimental research, instead of or in addition to statistical significance testing (Cohen, 1988). According to this argument, in order to assess the scientific significance of (a statistically significant) observed sample mean difference, one has to consider its magnitude (i.e., the effect size). Indeed, Cohen (1988) suggested several benchmarks (or thresholds) to guide such assessment: $0.20SD$ (small), $0.50SD$ (medium) and $0.80SD$ (large), to which Rosenthal (1996) added a fourth threshold: $1.30SD$ (very large) (Ellis, 2009, Table 1).

In a similar vein, we suggest that an empirical difference between two scores has to exceed a minimal threshold in order to be considered psychologically significant. While this suggestion might be accepted, in principle, by both practitioners and the scientific community[5], these communities are much less likely to agree on the critical threshold's magnitude. The tradeoff involved is clear: The higher the threshold, the more psychologically significant the differences exceeding it and the lower their relative frequency (and vice-versa)[6]. Unfortunately, there is no objective way to determine the "right" magnitude of the critical threshold. Nor can stipulation of this magnitude be based on statistical significance testing of intra-individual subtest score differences (Cahan, 1986; Cahan & Cohen, 1988). Hence the inherent arbitrariness of the threshold's magnitude and the need for consensus.

A last remark regards the implementation of the threshold approach in pattern analysis. The problem in this respect stems from the possible intransitivity of the magnitude relation (> or =) in this case. The following numerical example illustrates the problem. Let $A=1SD$, $B=0.7SD$ and $C=0.4SD$ be the three scores comprising an examinee's cognitive profile and let the critical threshold Δ_C be $0.5SD$. Here $A-B< \Delta_C$, $B-C< \Delta_C$, and $A-C >\Delta_C$. Hence $A=B$, $B=C$, and $A>C$. The problem can be prevented by using a larger unit of measurement, which equals or exceeds the predetermined threshold, instead of implementing the threshold approach on the original, smaller unit. In our example this can be done by expressing the three scores on the stanine scale, the unit of which equals $0.5SD$. Substituting the stanine scale, which ranges from 1 to 9, for the original z-score scale, which in our numerical example has a precision of $0.1SD$, we get: $A=B=7$ and $C=6$. Hence the profile pattern is $A=B>C$.

Acknowledgements

References

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, D.C.: American Council on Education.

Bergeron, R., & Floyd, R. G. (2006). Broad cognitive abilities of children with mental retardation: An analysis of group and individual profiles. *American Journal of Mental Retardation, 111*, 417-432.

Bolen, L. M. (1998). WISC−III score changes for EMH students. *Psychology in the Schools, 35*, 327-332.

Cahan, S. (1986). Significance testing of subtest score differences: The rules of the game. *Journal of Psychoeducational Assessment*, *4*, 273-280. https://doi.org/10.1177/073428298600400403

Cahan, S. (1989). A critical examination of the "reliability" and "abnormality" approaches to the evaluation of subtest score differences. *Educational and Psychological Measurement*, *49*, 807-814. https://doi.org/10.1177/001316448904900404

Cahan, S., & Cohen, N. (1988). Significance testing of subtest score differences: The case of nonsignificant results. *Journal of Psychoeducational Assessment*, *6*, 107-117. https://doi.org/10.1177/073428298800600202

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum.

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin, 50*, 456-473.

Dickerson-Mayes, S. D., Calhoun, S. L., Chase, G. A., Mink, D. M., & Stagg, R. E. (2009). ADHD subtypes and co-occurring anxiety, depression, and oppositional-defiant disorder: Differences in Gordon diagnostic system and Wechsler working memory and processing speed index scores. *Journal of Attention Disorders, 12*, 540-550. https://doi.org/10.1177/1087054708320402

Ellis, P. D. (2009, September). *Thresholds for interpreting effect sizes*. Retrieved July 29, 2018, from http://www.polyu.edu.hk/mm/effectsizefaqs/thresholds_for_interpreting_effect_sizes2html

Fiorello, C. A., Hale, J. B., Holdnack, J. A., Kavanagh, J. A., Terrell, J., & Long, L. (2008). Interpreting intelligence test results for children with disabilities: Is global intelligence relevant?. *Applied Neuropsychology, 14*, 2-12.

Gati, I., Landman, S., Davidovitch, S., Asulin-Peretz, L., & Gadassi, R. (2010). From career decision-making styles to career decision-making profiles: A multidimensional approach. *Journal of Vocational Behavior, 76*, 277-291. https://doi.org/https://doi.org/10.1016/j.jvb.2009.11.001

Glutting, J. J., McDermott, P. A., Watkins, M. M., Kush, J. C., & Konold, T. R. (1997). The base-rate problem and its consequences for interpreting children's ability profiles. *School Psychology Review, 26*, 176-188.

Hale, J. B., Yim, M., Schneider, A. N., Wilcox, G., Henzel, J. N., & Dixon, S. G. (2012). Cognitive and neuropsychological assessment of Attention-Deficit/Hyperactivity Disorder: Redefining a disruptive behavior disorder. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (3rd ed., pp. 687-707). New York, NY: Guilford Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger Publishers.

Lesser, G. S., Fifer, G., & Clark, D. H. (1965). Mental abilities of children from different social-class and cultural groups. *Monographs of the Society for Research in Child Development, 30*, 1-115.

Livingston, R. B., Jennings, E., Reynolds, C. R., & Gray, R. M. (2003). Multivariate analyses of the profile stability of intelligence tests: High for IQs, low to very low for subtest analyses. *Archives of Clinical Neuropsychology, 18*, 487-507.

Mayes, S. D., Calhoun, S. L., Mayes, R. D., & Molitoris, S. (2012). Autism and ADHD: Overlapping and discriminating symptoms. *Research in Autism Spectrum Disorders*, *6*, 277-285. https://doi.org/10.1016/j.ra*SD*.2011.05.009

McDermott, P. A., Fantuzzo, J. W., & Glutting, J. J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. *Journal of Psychoeducational Assessment, 8,* 290-302. https://doi.org/10.1177/073428299000800307

McDermott, P. A., Fantuzzo, J. W., & Glutting, J. J. (1992). Illusions of meaning in the ipsative assessment of children's ability. *Journal of Special Education, 25*, 504-526. https://doi.org/10.1177/002246699202500407

McDiarmid, M. (1993, May). Assessment profiles of children and adults identified as Learning Disabled. *Paper presented at the John Wesley Powell Student Research Conference, Michigan, Illinois*. Retrieved from http://digitalcommons.iwu.edu/psych_honproj/70

Moses, J. A. Jr., & Pritchard, D. A. (1996). Modal profiles for the wechsler adult intelligence scale-revised. *Archives of Clinical Neuropsychology, 11*, 61-68.

Naglieri, J. A., & Paolitto, A. W. (2005). Ipsative comparisons of WISC-IV index scores. *Applied Neuropsychology, 12*, 208-211. https://doi.org/10.1207/s15324826an1204_4

Raiford, S. E., Zhang, O., Drozdick, L. W., Getz, K., Wahlstrom, D., Gabel. A., … Daniel, M. (2016). *WISC–V Coding and Symbol Search in digital format: Reliability, validity, special group studies, and interpretation*. Bloomington, MN: Pearson.

Robinson, B. R., & Harrison, P. L. (2005). Wise-III core profiles for students referred or found eligible for special education and gifted programs. *School Psychology Quarterly, 20*, 51-65.

Rosenthal, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research, 21*, 37-39.

Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly, 15*, 465-479.

Watkins, M. W., & Canivez, G. L. (2004). Temporal stability of WISC-III subtest composite: Strengths and weaknesses. *Psychological Assessment, 16,* 133-138.

Watkins, M. W., & Glutting, J. J. (2000). Incremental validity of WISE-III profile elevation, scatter, and shape information for predicting reading and math achievement. *Psychological Assessment, 12,* 402-408. https://doi.org/10.1037/1040-3590.12.4.402

Watkins, M. W., Glutting, J. J., & Youngstrom, E. A. (2005). Issues in subtest profile analysis. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (pp. 251-268). New York, NY, US: Guilford Press.

Watkins, M. W., Kush, J. C., & Glutting, J. J. (1997). Discriminant and predictive validity of the WISC-III ACID profile among children with learning disabilities. *Psychology in the Schools, 34*, 309-319.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V)*. San Antonio, TX: Psychological Corporation.

Willner, T., Gati, I., & Guan, Y. (2015). Career decision-making profiles and career decision-making difficulties: A cross-cultural comparison among US, Israeli, and Chinese samples. *Journal of Vocational Behavior, 88,* 143-153. https://doi.org/10.1016/j.jvb.2015.03.007

Wright, A. G. C., Hallquist, M. N., Morse, J. Q., Scott, L. N., Stepp, S. D., Nolf, K. A., … Pilkonis, P. A. (2013). Clarifying interpersonal heterogeneity in Borderline Personality Disorder using Latent Mixture Modeling. *Journal of Personality Disorders, 27*, 125-143. https://doi.org/10.1521/pedi.2013.27.2.125

Zayat, M., Kalb, L., & Wodka, E. L. (2011). Brief report: Performance pattern differences between children with autism spectrum disorders and Attention Deficit-Hyperactivity Disorder on measures of verbal intelligence. *Journal of Autism and Developmental Disorders, 41*, 1743-1747. https://doi.org/10.1007/s10803-011-1207-z

Notes

Note 1. The formula incorporates ties. That is, profiles for which two or more subtests receive equal ranking.

Note 2. Denoting the three subtest scores by A, B, and C, the 13 possible patterns are:

A=B=C, A=B > C, A=B < C, A > B=C, A< B=C, A=C > B, A=C < B, A > B> C, A > C>B, B >A>C, B > C> A, C > A>B, C > B > A.

Note 3. One may wonder why not always use the highest available precision–which begs the stability question, and prefer instead a lower precision level. A major reason for preferring lower precision levels when higher precision is available, especially in the context of pattern analysis of test score profiles, is that a greater precision level might yield somewhat negligible differences between test scores (see the Discussion section for further elaboration of this point).

Note 4. Appendix B, the four-score profile analogue of Table 1, provides the percentage of stable four-score profiles patterns between measurement precision level pairs.

Note 5. A case in point is the use of the "abnormality" criterion for evaluating the psychological significance of subtest score differences (Cahan, 1989).

Note 6. An inevitable consequence of decreasing measurement precision through categorization of a continuous scale is that it can yield strange results. Assume that you have a continuous score and recode to stanines, then if

A = 0.24*SD*, B = 0.26*SD* and C = 0.74*SD*, A<B=C, while any reasonable person would say it is A=B<C. Widening the classes a bit would eventually yield the "right" comparison.

Appendix A

The inter-correlations between the 12 subtest scores

Subtest												
Subtest	1	2	3	4	5	6	7	8	9	10	11	12
1. Word classification	1	0.49	0.51	0.45	0.44	0.53	0.42	0.42	0.42	0.33	0.43	0.45
2. Figural classification	0.49	1	0.50	0.58	0.49	0.48	0.42	0.45	0.39	0.41	0.43	0.41
3. Word comparison	0.51	0.50	1	0.54	0.50	0.54	0.44	0.45	0.44	0.38	0.45	0.46
4. Figural analogies	0.45	0.58	0.54	1	0.63	0.50	0.46	0.54	0.42	0.43	0.47	0.45
5. Figural completion	0.44	0.49	0.50	0.63	1	0.50	0.48	0.55	0.43	0.44	0.47	0.46
6. Vocabulary	0.53	0.48	0.54	0.50	0.50	1	0.48	0.47	0.51	0.41	0.51	0.55
7. Number series	0.42	0.42	0.44	0.46	0.48	0.48	1	0.59	0.53	0.48	0.56	0.52
8. Figural series	0.42	0.45	0.45	0.54	0.55	0.47	0.59	1	0.55	0.53	0.53	0.52
9. Verbal oddities	0.42	0.39	0.44	0.42	0.43	0.51	0.53	0.55	1	0.52	0.54	0.57
10. Figural oddities	0.33	0.41	0.38	0.43	0.44	0.41	0.48	0.53	0.52	1	0.51	0.50
11. Verbal arithmetic problems	0.43	0.43	0.45	0.47	0.47	0.51	0.56	0.53	0.54	0.51	1	0.64
12. Sentence completion	0.45	0.41	0.46	0.45	0.46	0.55	0.52	0.52	0.57	0.50	0.64	1

Appendix B

Percentage of Stable Four-score Profile Patterns between Measurement Precision Level Pairs

	Measurement precision						
	infinitesimal	1/15*SD*	0.1*SD*	1/3*SD*	0.5*SD*	2/3*SD*	0.8*SD*
infinitesimal	-						
1/15*SD*	83 (77-89)	-					
0.1*SD*	74 (69-80)	85 (80-91)	-				
1/3*SD*	36 (30-41)	46 (41-52)	49 (43-55)	-			
0.5*SD*	20 (17-25)	28 (21-34)	33 (26-40)	45 (32-58)	-		
2/3*SD*	11 (8-14)	16 (12-20)	19 (15-24)	32 (27-41)	43 (37-53)	-	
0.8*SD*	6 (4-8)	10 (6-13)	12 (8-16)	27 (20-32)	39 (33-45)	53 (43-62)	-