Journal of Psychoeducational Assessment
Volume 24 Number 1
March 2006 52-68

10.1177/0734282905285225
http://jpa.sagepub.com
hosted at
http://online.sagepub.com

Long-Term Stability of Membership in a Wechsler Intelligence Scale for Children–Third Edition (WISC-III) Subtest Core Profile Taxonomy

Ellen R. Borsuk
Marley W. Watkins
Pennsylvania State University
Gary L. Canivez
Eastern Illinois University

Although often applied in practice, clinically based cognitive subtest profile analysis has failed to achieve empirical support. Nonlinear multivariate subtest profile analysis may have benefits over clinically based techniques, but the psychometric properties of these methods must be studied prior to their implementation and interpretation. The current study posed the following question: Is Wechsler Intelligence Scale for Children–Third Edition (WISC-III) cluster membership based on nonlinear multivariate subtest profile analysis stable over a 3-year period? Membership stability to a subtest taxonomy, including constancy of displaying an unusual profile, was based on data from 585 students. General (.39) and partial (.26 to .51) kappa coefficients either failed to reach statistical significance or indicated poor classification stability, with the exception of two profile types. It was concluded that, with these two possible exceptions, profile-type membership to an empirically derived WISC-III subtest taxonomy should not be used in interpretation or educational decision making.

Keywords: *intelligence test; Wechsler Intelligence Scale for Children; cluster analysis; profile analysis*

Wechsler scales are often used by school psychologists to assess intellectual functioning (Pfeiffer, Reddy, Kletzel, Schmelzer, & Boyer, 2000; Stinnett, Havey, & Oehler-Stinnett, 1995). In school contexts, millions of students have been given the Wechsler Intelligence Scale for Children–Third Edition (WISC-III; Wechsler, 1991) to determine entitlement for special education services (Kamphaus, Petoskey, & Rowe, 2000). Thus, it is vital to determine whether practitioners are making sound clinical decisions based on obtained

Authors' Note: This research was supported, in part, by an Eastern Illinois University Faculty Development Grant and a Pennsylvania State University College of Education Alumni Society Faculty Research Initiation Grant. The authors wish to express their gratitude to the school psychologists who generously responded to our request for data. Correspondence should be directed to Ellen R. Borsuk, Pennsylvania State University, Department of Educational and School Psychology and Special Education, 224 CEDAR Building, University Park, PA 16802; e-mail: erb155@psu.edu.

Wechsler scores. Given that reliability is a prerequisite for validity (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999), it is critical to examine the reliability of scores from Wechsler tests. Because intelligence is a construct that should be stable over time, scores from tests estimating intelligence should remain stable over time. Furthermore, special emphasis should be given to WISC-III score stability given how widespread this measure became among school psychologists and, therefore, its likely continued popularity in the form of the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV; Wechsler, 2003a, 2003b).

Clinically Based Profile Analysis

Although there is evidence for the internal consistency, short-term stability (Wechsler, 1991), long-term stability (Canivez & Watkins, 1998, 1999, 2001), and validity (Reschly, 1997) of global WISC-III scores, many clinicians go beyond these global scores and base intervention decisions on subtest profile analysis. For example, about 89% of the sample of school psychologists surveyed by Pfeiffer et al. (2000) reported that they used index scores and/or subtest profile analysis. This practice is endorsed by prominent authors (Kaufman, 1994; Sattler, 2001); even the WISC-III manual implicitly supports the use of profile analysis in making classification decisions by stating that "intersubtest *scatter* is the variability of an individual's scaled scores across the subtests. Such variability is frequently considered as diagnostically significant" (Wechsler, 1991, p. 177). Like Sattler (2001) and Kaufman (1994), the WISC-III manual outlines procedures for conducting profile analysis; the WISC-IV continues to provide similar guidelines.

Description of clinically based profile analysis techniques. *Profile analysis* refers to the determination of cognitive strengths and weaknesses to come to decisions regarding diagnosis and treatment (Glutting, McDermott, Konold, Snelbaker, & Watkins, 1998). *Hypothesis generation* refers to identification of intellectual strengths and weaknesses, which can in turn be used to guide intervention. Profiles can be defined as an examinee's set of scores on a given assessment occasion, where the *elements* of the profile would be subtest scores, index scores, and the like (Livingston, Jennings, Reynolds, & Gray, 2003). A profile has three dimensions: elevation (i.e., the level, or mean element score), scatter (i.e., a measure of dispersion, such as the range), and shape (Cronbach & Gleser, 1953). The shape of a profile is the residual data in the profile once elevation and scatter information have been removed (Cronbach & Gleser, 1953). Shape can be described as an examinee's unique patterns of high and low element scores on a given test (Watkins & Glutting, 2000). Given that elements are deemed to be high or low relative to an examinee's own mean, shape measurement is ipsative.

Many methods of profile analysis are clinically based rather than empirically derived (e.g., Kaufman, 1994; Kaufman & Lichtenberger, 2000). These popular systems reflect the relevance of IQ, index, and subtest score scatter in the interpretation of results. This scatter analysis is followed by an interpretation of the profile shape, which is thought to provide insight into the examinee's underlying set of abilities (Kaufman & Lichtenberger, 2000). Unfortunately, hypotheses about the correspondence between subtests and abilities (e.g., attention span, social comprehension) have little empirical support and, instead, are based on clinical experience. For example, Kamphaus (1998) acknowledged that "most of the presumed abilities that are offered for WISC-III interpretation are just that: Presumptions that are not supported by a preponderance of scientific evidence" (p. 45).

Difficulties with clinically based profile analysis techniques. Practitioner judgment regarding diagnosis and treatment is subject to error (Davidow & Levinson, 1993; Faust, 1986; Spengler, Strohmer, Dixon, & Shivy, 1995). Furthermore, clinicians are instructed to integrate data obtained from formal testing with other relevant knowledge (e.g., background information) in a manner consistent with his or her theoretical perspective to arrive at educational decisions for students (Kaufman & Lichtenberger, 2000). However, this is neglectful of the research that speaks to the inability of practitioners to accurately integrate clinical data in order to arrive at meaningful results (e.g., Faust, 1986).

There are many other difficulties associated with the use of clinically based profile analysis methods. These techniques are univariate in nature and do not incorporate multiple dependent comparisons simultaneously. Multivariate techniques are necessary when conducting profile analysis (Konold, Glutting, McDermott, Kush, & Watkins, 1999) to fully consider all three profile dimensions. In addition to relying on univariate techniques, clinically based profile analysis makes interpretations based on ipsative scores, which are theoretically different (McDermott, Fantuzzo, & Glutting, 1990; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992) and psychometrically inferior (Livingston et al., 2003; McDermott et al., 1992; Watkins & Canivez, 2004) to normative scores. Furthermore, these ipsatized scores may be derived from cognitive subtest scores with weak reliability; more global IQ scores demonstrate better reliability (Canivez & Watkins, 1998; Livingston et al., 2003; Wechsler, 1991). Finally, contrary to the implications made by clinically based profile analysis techniques, significant differences are a frequent occurrence and may not indicate that a child has a problem (Glutting, McDermott, & Konold, 1997; Konold et al., 1999). Given these difficulties, it is not surprising that there is little evidence to support the diagnostic application of clinically based profile analysis (Kavale & Forness, 1984; Watkins, 2003). Furthermore, there has not been support for the premise underlying hypothesis generation that identification of cognitive strengths and weaknesses are indicative of specific aptitudes, which can be translated into appropriate treatments, known as aptitude by treatment interactions (Reschly, 1997).

Nonlinear Multivariate Profile Analysis

Although clinically generated methods have not been fruitful, an empirical approach to profile analysis may engender support for diagnostic and educational decision making. Nonlinear multivariate profile analysis, an empirical method of profile analysis, has certain advantages over clinically based approaches. Through simultaneous examination of multiple subtest scores (Glutting, McDermott, & Konold, 1997), nonlinear multivariate profile techniques take both linear (i.e., level) and nonlinear (i.e., shape) characteristics of the profile into consideration at the same time (Glutting et al., 1998; Glutting, McDermott, Watkins, Kush, & Konold, 1997). Another advantage of nonlinear multivariate profile analysis is that unusual profiles thought to be of clinical interest are defined as those that show a difference that is significant when compared to all typical profiles (Glutting, McDermott, & Konold, 1997); common statistically significant differences are not interpreted. A third way in which nonlinear multivariate profile techniques may be beneficial relates to the fact that these methods do not rely on ipsative scores. Given these advantages, nonlinear multivariate profile analysis warrants further study.

Description of nonlinear multivariate profile analysis. Nonlinear multivariate profile analysis involves empirically grouping people into profile types based on score configurations that are commonly found in the population. One way of doing this is through cluster analysis. After identifying several difficulties with existing options in cluster analysis, McDermott (1998) developed a three-stage clustering method. His Multistage Euclidean Grouping (MEG) incorporates best practice techniques in cluster analysis, such as application of Ward's (1963) method (e.g., Konold et al., 1999), combining hierarchical and nonhierarchical clustering algorithms, and built-in replications (Milligan & Hirtle, 2003).

Using MEG procedures, or a modification of these procedures, core profiles have been identified for the standardization samples of a number of cognitive tests, including the Wechsler Intelligence Scale for Children–Revised (WISC-R; Wechsler, 1974) (McDermott, Glutting, Jones, Watkins, & Kush, 1989), Wechsler Preschool and Primary Scale of Intelligence (WPPSI; Wechsler, 1967) (Glutting & McDermott, 1990a), McCarthy Scales of Children's Abilities (MSCA; McCarthy, 1972) (Glutting & McDermott, 1990b), Wechsler Adult Intelligence Scale–Revised (WAIS-R; Wechsler, 1981) (McDermott, Glutting, Jones, & Noonan, 1989), Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983a, 1983b) (Glutting, McGrath, Kamphaus, & McDermott, 1992), Differential Ability Scales (DAS; Elliott, 1990) (Holland & McDermott, 1996), and WISC-III (Donders, 1996; Glutting, McDermott, & Konold, 1997; Konold et al., 1999). Full coverage was required in every case for derived taxonomies to be representative of the population.

A taxonomy based on the 10 mandatory WISC-III subtests was developed by Konold et al. (1999) for the 2,200 students in the WISC-III standardization sample. Cluster analysis was conducted using MEG procedures (Konold et al., 1999). The resulting eight core profiles were mainly distinguished by Full Scale IQ (FSIQ) level, but four were also marked by Verbal IQ (VIQ)/Performance IQ (PIQ) discrepancies. These attributes (i.e., a small number of profiles that were defined primarily by level, with some defined by shape) conformed to a more general trend. That is, core profiles from many tests of intelligence revealed these characteristics across experimenters and methods. As such, results are promising in terms of internal structure, but must also be examined for stability over time (reliability) and external validity. The stability of profile membership should be investigated because evidence of reliability is a prerequisite for the valid interpretation and use of test scores (AERA, APA, & NCME, 1999). That is, if profile membership is not stable over time, then relying on nonlinear multivariate profile analysis to make decisions for students would not be defensible.

Temporal stability of membership to core profiles. Initial evidence exists to suggest that membership to empirically derived cognitive subtest profiles possesses some level of stability in the short term. Using a subsample of the WISC-R standardization sample (n = 303), twice administered 11 WISC-R subtests with a retest interval generally ranging from 3 to 5 weeks, McDermott, Glutting, Jones, Watkins, et al. (1989) examined membership stability to the taxonomy that they derived. Initial membership was easily determined as these students had been part of the sample used to generate the taxonomy. Retest membership was established by first calculating Tatsuoka and Lohnes's (1988, pp. 377-378) modifications of Cattell's (1949) coefficient of profile similarity (r_p). Second, an iterative cluster analytic technique reassigned profiles to more appropriate clusters. Both general and partial kappa coefficients (k_m; Fleiss, 1971) were calculated. Considerable agreement was found between profile membership across testing times (McDermott, Glutting, Jones, Watkins, et al., 1989): Overall agreement between test and retest classifications was determined to be 57.5% beyond chance levels. Furthermore, most partial k_m coefficients were found to be statistically significant ($p < .01$).

Comparable results were found for overall short-term stability of profile-type membership for the MSCA (general $k_m = .728$; Glutting & McDermott, 1990b), K-ABC (general $k_m = .497$; Glutting et al., 1992), and DAS (general $k_m = .541$; Holland & McDermott, 1996). In addition, partial k_m coefficients were found to be statistically significant for MSCA core profiles (Glutting & McDermott, 1990b) as well as for K-ABC core profiles and a group of unusual K-ABC profiles (Glutting et al., 1992). The short-term stability coefficient of profile membership for the WPPSI, however, was lower (general $k_m = .216$; Glutting & McDermott, 1990a).

In contrast to short-term stability findings, one study found that empirically derived cognitive subtest profiles may not be sufficiently stable in the long term (Livingston et al., 2003). Scores from 60 students referred due to academic and behavior difficulties were examined. Participants were administered the WISC-R with an average retest interval of 3.09 years. Livingston et al. (2003) concluded that the average subtest profile stability coefficients "indicate an unsatisfactory level of reliability" (p. 504).

However, the long-term stability of cognitive profiles must be further examined as Livingston et al. (2003) did not evaluate profile stability by comparison to a core taxonomy. Instead, each student's profile at Time 1 was compared to his or her profile at Time 2. It is possible that some participants' profiles appeared unstable but were stable in the sense that they remained unusual over time compared to core profile types. Furthermore, the change in profiles found across time may not have been large enough for core profile type reassignment. For example, in developing core subtest taxonomies for the WISC-R and WISC-III, McDermott, Glutting, Jones, Watkins, et al. (1989); Glutting, McDermott, and Konold (1997); and Konold et al. (1999) all set the a priori mean r_p between clusters to be $<.40$. Additionally, results showed that the range of average r_p scores across these studies was .20 to .33. Profiles in the Livingston et al. study were more similar to one another at different points in time (mean $r_p = .43$) than were core profile clusters based on the standardization sample cognitive scores. Finally, with only 60 participants, it is quite possible that not all taxonomy categories were adequately represented, possibly resulting in misleading conclusions.

As there is virtually no research examining long-term empirical profile stability, the present study explored the long-term (i.e., 3 year) stability of WISC-III cluster membership based on nonlinear multivariate profile analysis. The research question was, Is WISC-III cluster membership based on nonlinear multivariate profile analysis stable over a 3-year period?

Method

Participants

Participants represented a subset of the sample studied by Canivez and Watkins (1998). Participants in the current study consisted of 585 students who had data available for all 10 WISC-III mandatory subtests at two points in time. WISC-III data was reported by 107 school psychologists in 33 different states. On average, 5.47 cases were reported per psychologist, with a range from 1 to 24 and a standard deviation of 3.84. Table 1 displays the demographic characteristics of this sample.

Participants' average age was 9.16 years at Time 1 (range = 6.00 to 14.60; *SD* = 2.02) and 11.98 years at Time 2 (range = 7.50 to 16.90; *SD* = 2.07). The mean retest interval was 2.82 years (*SD* = 0.54), and the range was 0.50 to 6.00 years. The test-retest interval was less than 1 year for only 1.20% of the sample.

Instrument

The WISC-III is an individually administered test of intelligence that can be administered to children between the ages of 6 years, 0 months and 16 years, 11 months. Altogether, the WISC-III is comprised of 13 subtests; 6 subtests are classified as Verbal, and the remaining 7 are Performance subtests. All Verbal and Performance subtest scores have a mean of 10 and a standard deviation of 3. A child's performance across the 10 mandatory subtests yields an overall, or (FSIQ). In addition, both a VIQ and PIQ composite score can be calculated based on scores from the 5 mandatory subtests found under each scale, respectively. IQ scores have a mean of 100 and a standard deviation of 15. Evidence of score reliability and validity are documented in the WISC-III manual (Wechsler, 1991).

Procedures

Canivez and Watkins (1998) mailed a request to 2,000 school psychologists who were randomly chosen from among members of the National Association of School Psychologists. Contacted members were asked to submit demographic information as well as test and retest data for students who were twice tested with the WISC-III as part of special education eligibility evaluations. No other criteria were specified, such as number of cases to report or age of the children. There was no requirement that intellectual functioning be assessed by the same school psychologist at both points in time. Finally, confidentiality was ensured as students' names were not requested.

Profile similarity measure employed. To determine whether cognitive profile membership remained stable over time, participants' profiles at both Time 1 and Time 2 were compared to the eight core subtest profiles identified by Konold et al. (1999). A measure of profile similarity was used to measure the likeness between a profile and the core profiles and, thus, to establish profile membership. Many profile similarity techniques exist. Similarity measures that disregard elevation and scatter information are generally not preferred (Cronbach & Gleser, 1953). Furthermore, level information should not be ignored when examining stability of cognitive profiles, given that elevation information (i.e., intelligence) is the variable of interest. Finally, a similarity measure representing all profile dimensions was desirable for this study to be consistent with the nonlinear multivariate nature of core profiles (Konold et al., 1999). For these reasons, D^2 (Cronbach & Gleser, 1953; Osgood & Suci, 1952) was chosen as the similarity measure for the current study. D^2 is a measure of dissimilarity that is based on the Euclidean distance between two profiles. That is, the sum of squared differences between each pair of points from the two profiles is calculated. D^2 is sensitive to all three profile dimensions and does not have any restrictive assumptions (Livingston et al., 2003).

Table 1
Gender, Race/Ethnicity, Disability, Grade Level, and Geographic Region of Participants

Variable	*n*	%
Gender		
Male	394	67.35
Female	191	32.65
Race/ethnicity		
White	447	76.41
Black	86	14.70
Hispanic	33	5.64
Native American	4	0.68
Asian/Pacific	1	0.17
Other	4	0.68
Missing	10	1.71
Disability[a]		
Not disabled	18	3.08
Learning disability	368	62.91
Mental retardation	57	9.74
Emotional disability	42	7.18
Speech and language disability	16	2.74
Other disabilities	38	6.50
Unspecified	46	7.86
Grade[b]		
K	21	3.59
1	109	18.63
2	138	23.59
3	94	16.07
4	76	12.99
5	71	12.14
6	36	6.15
7	26	4.44
8	8	1.37
9	2	0.34
Missing	4	0.68
Geographic region[c]		
West	126	21.54
South	209	35.73
North Central	184	31.45
Northeast	66	11.28

a. Diagnoses made during first testing in accordance with state and federal guidelines.
b. Grades at time of first testing.
c. The country was divided into the regions outlined in the Wechsler Intelligence Scale for Children–Third Edition (WISC-III; Wechsler, 1991).

Core profile membership or designation as unusual. Classification decisions were made for participants at both Time 1 and Time 2. At each time point the D^2 value was calculated between each individual's WISC-III profile and every core profile in Konold et al.'s (1999) subtest taxonomy. Classification to a profile type was based on the lowest D^2 value, as this indicated the greatest similarity.

Participants whose profiles were unlike all core profiles were classified as *unusual*. Consistent with core profile membership, the designation of unusual was also determined by D^2. Using the method and reasoning outlined by Konold et al. (1999), the critical D^2 value used was 98. That is, in the event that every D^2 value calculated between a given participant's profile and each profile in the taxonomy was ≥ 98, the profile was determined to be unusual compared to the general population. This procedure identified a participant as belonging to a subgroup representing the 5% of children having profiles most discrepant from core profiles.

Determination of profile membership stability. Classification stability was calculated across time using Fleiss's (1971) kappa coefficient (k_m), a coefficient that yields the percentage agreement of profile classification between Time 1 and Time 2 corrected for agreement due to chance. Determination of k_m coefficients is consistent with studies that have examined the short-term classification stability of cognitive profiles (Glutting & McDermott, 1990a, 1990b; Holland & McDermott, 1996).

MacKappa (Watkins, 1998) was used to calculate general and partial k_m coefficients. That is, an overall or general k_m coefficient was calculated in addition to nine partial k_m coefficients, representing each core profile type as well as the group of unusual profiles. The level of statistical significance was determined for each k_m coefficient calculated.

Cicchetti (1994) summarized suggested interpretations: A kappa coefficient < .40 indicates poor clinical significance; a kappa value between .40 and .59 is considered fair; good clinical significance is defined as a kappa coefficient ranging from .60 to .74; and a kappa coefficient of .75 and greater is excellent. Based on these guidelines, it was decided a priori that statistically significant general and partial k_m coefficients of ≥ .40 would indicate that future research is warranted to determine whether there is evidence of validity for the interpretation and use of core profile membership information. By choosing this conservative cutoff point, helpful practices in educational decision making based on nonlinear multivariate profile analysis would not be overlooked. On the other hand, for k_m coefficients found to be < .40 there would be no support for conducting further research. As reliability is a prerequisite for validity for stable traits like intelligence (AERA, APA, & NCME, 1999), poor classification stability of core profile membership over time necessarily excludes the possibility that these membership decisions are valid.

Results

IQ and subtest scores for the sample at both Time 1 and Time 2 are displayed in Table 2. Participants were divided into nine different profile types: Profiles 1 through 8 as defined by Konold et al. (1999), and a profile type reserved for those determined to have unusual profiles. The number of children in each profile type is displayed in Table 3. Tables 4, 5, and 6 display the demographic characteristics of each profile type at both Time 1 and Time 2, and Tables 7 and 8 show mean WISC-III IQ and subtest scores at Time 1 and Time 2 across profile types.

The profile membership agreement of all profile types across time was .39 ($p < .0029$). Partial k_m coefficients for each individual profile ranged from .26 to .51 (see Table 3). Further analyses comparing students with unstable profiles ($n = 303$) to those with stable ones ($n = 282$) revealed no significant differences between the groups in terms of age, retest interval, gender, ethnicity, grade, disability, or geographic region ($p < .0029$).

Table 2
Means and Standard Deviations of Wechsler Intelligence Scale for Children–Third Edition (WISC-III) IQ and Subtest Scores at Times 1 and 2

	Time 1		Time 2	
Score	*M*	*SD*	*M*	*SD*
FSIQ	88.68	16.12	88.41	17.00
VIQ	88.71	15.84	88.18	15.79
PIQ	90.79	16.75	90.83	17.91
PC	8.70	3.35	9.05	3.38
IN	7.75	3.13	7.94	3.16
CD	8.30	3.42	7.68	3.25
SM	8.20	3.40	8.39	3.23
PA	8.47	3.58	8.67	3.92
AR	7.25	3.08	7.18	2.95
BD	8.41	3.72	8.30	4.03
VO	8.02	3.22	7.49	3.13
OA	8.42	3.38	8.52	3.62
CM	8.66	3.72	8.40	3.53

Note: FSIQ = Full Scale IQ; VIQ = Verbal IQ; PIQ = Performance IQ; PC = Picture Completion; IN = Information; CD = Coding; SM = Similarities; PA = Picture Arrangement; AR = Arithmetic; BD = Block Design; VO = Vocabulary; OA = Object Assembly; CM = Comprehension.

Table 3
Number of Students Belonging to Each Profile Type at Both Time 1 and Time 2 and Agreement Beyond Chance (kappa) for Each Profile

Profile Type	Time 1	Time 2	k_m Coefficient
Profile 1	15	12	.43
Profile 2	25	40	.40
Profile 3	15	11	.37
Profile 4	58	50	.35
Profile 5	74	81	.32
Profile 6	127	127	.43*
Profile 7	92	75	.36*
Profile 8	116	123	.51*
Unusual profile	63	66	.26

*$p < .0029$ (with Bonferroni correction adjusting for 17 comparisons; experiment-wise error rate = .05).

Discussion

The purpose of this study was to examine the long-term stability of profile-type membership based on an empirically derived taxonomy of WISC-III core subtest profiles (Konold et al., 1999). Given that reliability is a prerequisite for the valid interpretation and use of test results (AERA, APA, & NCME, 1999), it is imperative that the stability of results based on profile analysis be investigated. Due to the fact that clinically based methods of profile analysis have many limitations (e.g., McDermott et al., 1992) and little empirical support (e.g.,

Table 4
Mean, Standard Deviation, and Range of Ages in Years Across Profile Types at Both Time 1 and Time 2

	Time 1			Time 2		
	M	*SD*	Range	*M*	*SD*	Range
Profile 1	8.7	2.2	6.0-13.3	11.4	2.1	8.8-16.2
Profile 2	9.0	1.9	6.0-12.9	11.7	2.0	8.9-16.0
Profile 3	8.3	0.9	7.0-9.6	10.7	1.4	9.2-14.0
Profile 4	8.6	2.0	6.0-13.9	11.6	1.6	9.0-15.5
Profile 5	9.2	1.7	6.0-13.2	12.1	1.8	8.0-16.8
Profile 6	9.3	2.0	6.0-13.9	12.0	2.2	8.3-16.9
Profile 7	8.8	1.8	6.0-12.7	11.8	2.2	7.9-16.6
Profile 8	9.7	2.3	6.0-14.6	12.1	2.2	7.5-16.6
Unusual profile	9.2	2.3	6.0-14.0	12.4	2.3	7.6-16.8

Table 5
Percentage of Participants at Time 1 Distributed Across Gender, Race/Ethnicity, Disability, and Geographic Region for Each Profile Type

	Profile Type								
	1	2	3	4	5	6	7	8	Unusual
Gender									
Boys	60	84	60	71	78	68	65	58	68
Girls	40	16	40	29	22	32	35	42	32
Race/ethnicity									
White	100	92	100	79	81	80	70	67	70
Black	0	0	0	7	11	11	21	24	21
Hispanic	0	0	0	9	3	5	9	7	6
Native American	0	0	0	2	0	2	0	0	2
Asian/Pacific	0	0	0	0	0	0	1	0	0
Other	0	4	0	0	1	1	0	1	0
Missing	0	4	0	3	4	2	0	1	2
Disability									
Not disabled	7	4	0	0	0	3	3	7	2
LD	67	64	73	91	82	77	63	30	41
MR	0	0	0	0	0	2	1	29	32
ED	13	16	7	2	8	5	5	8	13
SLD	0	0	0	2	1	2	3	6	2
Other disabilities	7	4	13	0	3	2	8	16	5
Unspecified	7	12	7	5	5	9	16	3	6
Geographic region									
West	0	20	27	34	23	22	26	14	19
South	33	28	40	40	36	33	33	41	35
North Central	40	28	27	24	30	36	29	33	32
Northeast	27	24	7	2	11	9	12	13	14

Note: LD = learning disability; MR = mental retardation; ED = emotional disability; SLD = speech and language disability.

Table 6
Percentage of Participants at Time 2 Distributed Across Gender, Race/Ethnicity, Disability, and Geographic Region for Each Profile Type

	Profile Type								
	1	2	3	4	5	6	7	8	Unusual
Gender									
Boys	83	88	73	62	81	73	60	50	68
Girls	17	13	27	38	19	27	40	50	32
Race/ethnicity									
White	100	88	82	82	77	77	76	63	85
Black	0	5	9	12	16	11	17	27	6
Hispanic	0	3	9	0	4	9	7	7	5
Native American	0	0	0	2	0	1	0	1	2
Asian/Pacific	0	0	0	0	0	0	0	1	0
Other	0	3	0	0	0	1	0	1	2
Missing	0	3	0	4	4	2	0	1	2
Disability									
Not disabled	0	8	0	6	5	5	11	8	2
LD	75	70	73	72	70	76	57	36	41
MR	0	0	0	0	0	2	0	20	38
ED	8	13	0	4	11	4	11	6	6
SLD	0	0	0	0	0	2	3	7	0
Other disabilities	8	0	0	4	7	2	9	15	6
Unspecified	8	10	27	14	6	9	9	9	8
Geographic region									
West	0	10	27	24	25	31	16	14	29
South	33	35	45	36	35	31	36	44	30
North Central	50	33	18	32	33	28	31	34	29
Northeast	17	23	9	8	7	10	17	8	12

Note: LD = learning disability; MR = mental retardation; ED = emotional disability; SLD = speech and language disability.

Watkins, 2003), the long-term stability of profile-type membership based on an empirically derived WISC-III taxonomy (i.e., nonlinear multivariate profile analysis) was examined.

Results of this study indicated that cluster membership based on nonlinear multivariate profile analysis was generally not stable over a 3-year period: The agreement coefficient across profile types was poor according to Cicchetti's (1994) guidelines. Although statistically significant, this coefficient did not reach the a priori value that would warrant further research into the potential utility of nonlinear multivariate profile type membership information when making educational decisions. Furthermore, given that those with unstable profiles did not differ as a group from those with stable profiles with respect to reported demographic data, there is no reason to expect a different result for any given demographic subgroup (e.g., males).

Of the agreement coefficients that were calculated for each profile type, only three reached statistical significance: Coefficients from Profiles 6, 7, and 8 likely represented true stability over time. However, only the coefficients for Profiles 6 and 8 exceeded the a priori magnitude deemed necessary to justify future validity research (Cicchetti, 1994).

Table 7
Mean Wechsler Intelligence Scale for Children–Third Edition (WISC-III) IQ and Subtest Scores at Time 1 Across Profile Types

	Profile Type								
Score	1	2	3	4	5	6	7	8	Unusual
FSIQ	126	112	108	102	97	88	86	71	82
VIQ	124	110	111	96	100	84	91	72	84
PIQ	123	114	104	109	95	96	83	73	83
PC	13	13	10	11	10	10	8	6	7
IN	14	11	12	8	10	7	8	5	7
CD	12	9	13	12	7	8	8	7	7
SM	14	12	11	9	11	7	9	5	8
PA	14	13	11	12	9	9	8	5	7
AR	13	10	11	9	9	7	7	5	5
BD	15	13	10	11	10	10	6	5	7
VO	15	12	12	9	10	7	9	5	8
OA	13	12	10	11	9	10	7	6	7
CM	14	13	13	11	11	7	10	6	7

Note: FSIQ = Full Scale IQ; VIQ = Verbal IQ; PIQ = Performance IQ; PC = Picture Completion; IN = Information; CD = Coding; SM = Similarities; PA = Picture Arrangement; AR = Arithmetic; BD = Block Design; VO = Vocabulary; OA = Object Assembly; CM = Comprehension.

Table 8
Mean Wechsler Intelligence Scale for Children–Third Edition (WISC-III) IQ and Subtest Scores at Time 2 Across Profile Types

	Profile Type								
Score	1	2	3	4	5	6	7	8	Unusual
FSIQ	126	112	106	104	98	87	85	72	77
VIQ	124	110	109	97	100	84	90	74	77
PIQ	124	113	103	110	97	94	82	74	82
PC	13	13	9	11	11	10	8	6	8
IN	14	12	12	10	10	7	8	5	6
CD	12	9	13	12	7	8	8	6	6
SM	14	12	12	10	11	8	9	6	7
PA	14	13	10	13	9	9	8	6	7
AR	13	10	10	9	9	7	8	5	5
BD	16	13	11	11	10	9	6	5	7
VO	14	12	11	9	9	6	8	5	6
OA	13	13	8	11	10	10	6	6	7
CM	14	12	13	10	10	8	9	6	6

Note: FSIQ = Full Scale IQ; VIQ = Verbal IQ; PIQ = Performance IQ; PC = Picture Completion; IN = Information; CD = Coding; SM = Similarities; PA = Picture Arrangement; AR = Arithmetic; BD = Block Design; VO = Vocabulary; OA = Object Assembly; CM = Comprehension.

Profiles 6 and 8: Characteristics and Directions for Future Research

Compared to the overall sample, members of Profile 6 had a higher frequency of students diagnosed with learning disabilities and fewer diagnosed with mental retardation and other disabilities. Having a smaller number of participants with mental retardation is not unexpected given that the average FSIQ for this profile ranged from 87 (Time 2) to 88 (Time 1), whereas the generally accepted cutoff for diagnosis of mental retardation is 70 (Spruill, 1998).

Compared to the overall sample, Profile 8 contained more students diagnosed with mental retardation, speech and language disabilities, and other disabilities and fewer students with learning disabilities. Having a higher proportion of students with mental retardation is not unexpected given that the average FSIQ of this profile was near 70. In addition, Profile 8 included more females and not as many males compared to the overall sample. More Black students and fewer White students were members of Profile 8, a finding that is not unanticipated given that the same trend was seen among children aged 6 to 21 who received special education services during 2000-2001 (U.S. Department of Education [USDOE], 2001).

Average WISC-III IQ and subtest scores of students in Profiles 6 and 8 resembled those described by Konold et al. (1999). Although certain subtest score trends can be established for Profile 6, future research will be better directed to exploration of the meaning of membership in Profile 6 or 8, rather than the meaning of isolated components of those profiles: The premise of this article's investigation of an empirical approach to profile analysis was due to the many limitations of clinically based methods. Membership in Profiles 6 and 8 can be investigated for correspondence with certain outcome variables, such as an aspect of classroom behavior or response to a specific mode of instruction.

When conducting future research, it is critical to remember that classification to Profile 6 or 8 is only meaningful if this information is useful in making predictions or if it is used to generate effective interventions (Glutting, McDermott, Prifitera, & McGrath, 1994). Also, it must be kept in mind that to be meaningful, profile membership information must combine with overall intelligence scores to produce results that exceed what is currently predictable through knowledge of global IQ scores alone. That is, knowledge of profile membership must add incremental validity relative to global intelligence scores (Lubinski, 2004).

Limitations and Additional Directions for Future Research

Interpretation of the results of this study must be made within the context of its limitations. Participants represented a subset of the sample obtained by Canivez and Watkins (1998). Their low survey response rate removed the randomness with which participants were originally selected. That is, participation was based on the voluntary decision of school psychologists to respond. Thus, participants of the present study were also not randomly selected. Nonrandomness of a sample reduces generalizability of results. On the other hand, there was no reason to suspect a selection bias among more than 100 school psychologists from 33 states.

Several other factors also limit generalizability of the current findings. Only scores of students who were reevaluated were included. As such, results should not be extended to students who were only evaluated once, such as those no longer requiring special education services. In addition, results are most representative of students in Grades 1 through 5. Although participants were representative of the population of students receiving special education (USDOE, 2001), it is difficult to generalize results to children who are not White as well as to students without disabilities and to those having a disability other than a learning disability.

Further research is needed before results can be generalized to students' profiles on measures of cognitive functioning other than the WISC-III. For example, the WISC-III has been replaced by the WISC-IV, and it is important that the current study be replicated using this updated measure and that suggestions for future research based on the results of this study be modified for similar studies involving the WISC-IV.

Another difficulty of this study was related to examiner effects. The accuracy of WISC-III administration to students at both Time 1 and at Time 2 could only be assumed. Also, because the assessors of a given student at Time 1 and Time 2 may have varied, different degrees of standardized administration on the part of the examiners might have influenced WISC-III results and distorted classification stability findings.

A final consideration is the small membership of some profile types. The lack of statistical significance for Profiles 1 and 2 might have been due to inadequate power. Future research should enlist a larger number of participants across profile types to address this limitation.

Conclusion

Although it appears that profile-type membership possesses some degree of stability in the short term across a number of cognitive measures (Glutting et al., 1992; Glutting & McDermott, 1990b; McDermott, Glutting, Jones, Watkins, et al., 1989), results of this study revealed that profile-type membership of empirically derived WISC-III subtest profiles did not remain stable in the long term (i.e., 3 years). As such, empirically based WISC-III subtest profile-type membership cannot be relied upon to make educational decisions for students. Even though a nonlinear multivariate approach to profile analysis has advantages over clinically based techniques, to date *neither* approach has been empirically supported in its contribution to diagnosis or educational decision making. Two possible exceptions were Profiles 6 and 8, which had fair agreement coefficients. Thus, future validity research as well as replication research taking stated limitations into account will be important for ultimate interpretation of the current findings.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Canivez, G. L., & Watkins, M. W. (1998). Long-term stability of the Wechsler Intelligence Scale for Children–Third Edition. *Psychological Assessment, 10*, 285-291.

Canivez, G. L., & Watkins, M. W. (1999). Long term stability of the Wechsler Intelligence Scale for Children–Third Edition among demographic subgroups: Gender, race, and age. *Journal of Psychoeducational Assessment, 17*, 300-313.

Canivez, G. L., & Watkins, M. W. (2001). Long term stability of the Wechsler Intelligence Scale for Children–Third Edition among students with disabilities. *School Psychology Review, 30*, 361-376.

Cattell, R. B. (1949). r_p and other coefficients of pattern similarity. *Psychometrika, 14*, 279-298.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284-290.

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

Davidow, J., & Levinson, E. M. (1993). Heuristic principles and cognitive bias in decision making: Implications for assessment in school psychology. *Psychology in the Schools, 30*, 351-361.

Donders, J. (1996). Cluster subtypes in the WISC-III standardization sample: Analysis of factor index scores. *Psychological Assessment, 8*, 312-318.

Elliott, C. D. (1990). *Differential Ability Scales: Introductory and technical handbook*. San Antonio, TX: Psychological Corporation.

Faust, D. (1986). Research on human judgment and its application to clinical practice. *Professional Psychology: Research and Practice, 17*, 420-430.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*, 378-382.

Glutting, J. J., & McDermott, P. A. (1990a). Patterns and prevalence of core profile types in the WPPSI standardization sample. *School Psychology Review, 19*, 471-491.

Glutting, J. J., & McDermott, P. A. (1990b). Score structures and applications of core profile types in the McCarthy Scales standardization sample. *Journal of Special Education, 24*, 212-233.

Glutting, J. J., McDermott, P. A., & Konold, T. R. (1997). Ontology, structure, and diagnostic benefits of a normative subtest taxonomy from the WISC-III standardization sample. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 349-372). New York: Guilford.

Glutting, J. J., McDermott, P. A., Konold, T. R., Snelbaker, A. J., & Watkins, M. W. (1998). More ups and downs of subtest analysis: Criterion validity of the DAS with an unselected cohort. *School Psychology Review, 27*, 599-612.

Glutting, J. J., McDermott, P. A., Prifitera, A., & McGrath, E. A. (1994). Core profile types for the WISC-III and WIAT: Their development and application in identifying multivariate IQ-achievement discrepancies. *School Psychology Review, 23*, 619-639.

Glutting, J. J., McDermott, P. A., Watkins, M. W., 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.

Glutting, J. J., McGrath, E. A., Kamphaus, R. W., & McDermott, P. A. (1992). Taxonomy and validity of subtest profiles on the Kaufman Assessment Battery for Children. *Journal of Special Education, 26*, 85-115.

Holland, A. M., & McDermott, P. A. (1996). Discovering core profile types in the school-age standardization sample of the Differential Ability Scales. *Journal of Psychoeducational Assessment, 14*, 131-146.

Kamphaus, R. W. (1998). Intelligence test interpretation: Acting in the absence of evidence. In A. Prifitera & D. Saklofske (Eds.), *WISC-III clinical use and interpretation: Scientist-practitioner perspectives* (pp. 39-57). San Diego, CA: Academic Press.

Kamphaus, R. W., Petoskey, M. D., & Rowe, E. W. (2000). Current trends in psychological testing of children. *Professional Psychology: Research and Practice, 31*, 155-164.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York: John Wiley.

Kaufman, A. S., & Kaufman, N. L. (1983a). *K-ABC administration and scoring manual*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (1983b). *K-ABC interpretive manual*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Lichtenberger, E. O. (2000). *Essentials of WISC-III and WPPSI-R assessment*. New York: John Wiley.

Kavale, K. A., & Forness, S. R. (1984). A meta-analysis of the validity of Wechsler scale profiles and recategorizations: Patterns or parodies? *Learning Disability Quarterly, 7*, 136-156.

Konold, T. R., Glutting, J. J., McDermott, P. A., Kush, J. C., & Watkins, M. W. (1999). Structure and diagnostic benefits of a normative subtest taxonomy developed from the WISC-III standardization sample. *Journal of School Psychology, 37*, 29-48.

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.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General intelligence,' objectively determined and measured." *Journal of Personality and Social Psychology, 86*, 96-111.

McCarthy, D. (1972). *Manual for the McCarthy Scales of Children's Abilities*. New York: Psychological Corporation.

McDermott, P. A. (1998). MEG: Megacluster analytic strategy for multistage hierarchical grouping with relocations and replications. *Educational and Psychological Measurement, 58*, 677-686.

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.

McDermott, P. A., Fantuzzo, J. W., Glutting, J. J., Watkins, M. W., & Baggaley, A. R.(1992). Illusions of meaning in the ipsative assessment of children's ability. *Journal of Special Education, 25*, 504-526.

McDermott, P. A., Glutting, J. J., Jones, J. N., & Noonan, J. V. (1989). Typology and prevailing composition of core profiles in the WAIS-R standardization sample. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*, 118-125.

McDermott, P. A., Glutting, J. J., Jones, J. N., Watkins, M. W., & Kush, J. (1989). Core profile types in the WISC-R national sample: Structure, membership, and applications. *Psychological Assessment, 1*, 292-299.

Milligan, G. W., & Hirtle, S. C. (2003). Clustering and classification methods. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* (pp. 165-186). Hoboken, NJ: Wiley.

Osgood, C. E., & Suci, G. J. (1952). A measure of relation determined by both mean differences and profile information. *Psychological Bulletin, 49*, 251-262.

Pfeiffer, S. I., Reddy, L. A., Kletzel, J. E., Schmelzer, E. R., & Boyer, L. M. (2000). The practitioner's view of IQ testing and profile analysis. *School Psychology Quarterly, 15*, 376-385.

Reschly, D. J. (1997). Diagnostic and treatment utility of intelligence tests. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 437-456). New York: Guilford.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Jerome M. Sattler.

Spengler, P. M., Strohmer, D. C., Dixon, D. N., & Shivy, V. A. (1995). A scientist-practitioner model of psychological assessment: Implications for training, practice, and research. *The Counseling Psychologist, 23*, 506-534.

Spruill, J. (1998). Assessment of mental retardation with the WISC-III. In A. Prifitera & D. Saklofske (Eds.), *WISC-III clinical use and interpretation: Scientist-practitioner perspectives* (pp. 73-90). San Diego, CA: Academic Press.

Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. *Journal of Psychoeducational Assessment, 12*, 331-350.

Tatsuoka, M. M., & Lohnes, P. R. (1988). *Multivariate analysis: Techniques for educational and psychological research* (2nd ed.). New York: Macmillan.

U.S. Department of Education. (2001). *To assure the free appropriate public education of all children with disabilities: Twenty-fourth annual report to Congress on the implementation of the Individuals With Disabilities Education Act* (USDOE Contract No. HS97020001 with Westat). Washington, DC: Author.

Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *American Statistical Association Journal, 58*, 236-244.

Watkins, M. W. (1998). MacKappa [Computer software]. University Park, PA: Author.

Watkins, M. W. (2003). IQ subtest analysis: Clinical acumen or clinical illusion? *Scientific Review of Mental Health Practice, 2*, 118-141.

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 WISC-III profile elevation, scatter, and shape information for predicting reading and math achievement. *Psychological Assessment, 12*, 402-408.

Wechsler, D. (1967). *Manual for the Wechsler Preschool and Primary Scale of Intelligence*. New York: Psychological Corporation.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children–Revised*. New York: Psychological Corporation.

Wechsler, D. (1981). *Manual for Wechsler Adult Intelligence Scale–Revised*. New York: Psychological Corporation.

Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children–Third Edition*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003a). *Wechsler Intelligence Scale for Children–Fourth Edition administration and scoring manual*. San Antonio, TX: Psychological Corporation.
Wechsler, D. (2003b). *Wechsler Intelligence Scale for Children–Fourth Edition technical and interpretive manual*. San Antonio, TX: Psychological Corporation.