Journal of the International Neuropsychological Society (1999), **5**, 26–31.

Performance discrepancies on the California Verbal Learning Test–Children's Version in the standardization sample

JACQUES DONDERS
Mary Free Bed Hospital, Grand Rapids, MI

(Received January 14, 1998; Revised April 9, 1998; Accepted April 17, 1998)

Abstract

The standardization data for the California Verbal Learning Test–Children's Version (CVLT–C) were used to evaluate statistically significant discrepancies between key quantitative variables of this instrument, as well as the base rate of specific discrepancies. The results indicated that apparently large discrepancies between the respective standard scores were actually fairly common. However, for 3 of the 4 contrasts, discrepancies that equaled or exceeded 1.5 *z*-score points in the hypothesized direction were sufficiently unusual to be considered clinically significant. For a 4th contrast, discrepancies that equaled or exceeded 1 *z*-score point in the hypothesized direction appeared to meet this criterion. It is suggested that the interpretation of clinically obtained CVLT–C profiles should focus primarily on specific quantitative variables, with inclusion of consideration of the presented base rates of discrepancies between the respective *z* scores. (*JINS*, 1999, *5*, 26–31.)

Keywords: Assessment, Memory, Test score discrepancies

INTRODUCTION

Learning and remembering new information is an essential part of children's development, and these skills are often negatively affected by acquired brain damage (Fletcher et al., 1995; Ylvisaker et al., 1994). Therefore, incorporation of psychometrically sound measures of memory abilities is crucial in pediatric neuropsychological evaluations. The California Verbal Learning Test–Children's Version (CVLT–C; Delis et al., 1994) is a recently developed instrument that offers clinicians the opportunity to evaluate quantitative and qualitative aspects of children's memory abilities through standard scores with reference to age-based norms. The reliability of the CVLT–C appears to be acceptable, with average values of coefficient alpha ranging from .72 to .85 (Delis et al., 1994). The validity of this instrument in terms of sensitivity to acquired brain damage has also been demonstrated in several recent studies (Jaffe et al., 1993; Levin et al., 1993).

Reprint requests to: Jacques Donders, Psychology Service, Mary Free Bed Hospital, 235 Wealthy S.E., Grand Rapids MI 49503. E-mail: jdonders@mfbrc.com

The CVLT–C manual provides general interpretive guidelines with regard to the consideration of contrast variables (e.g., the difference between free recall and recognition), but no supporting actuarial data are included. Yeates and his colleagues have recently provided information about the contrasts between specific CVLT–C variables in children with traumatic brain injury (Yeates et al., 1995a) and children with myelomeningocele and shunted hydrocephalus (Yeates et al., 1995b). Compared to matched controls, children in both groups demonstrated greater proportional decline in free recall between short and long delayed recall trials, and relatively greater facilitation of retrieval under a multiple-choice recognition format. However, the findings were limited somewhat by the fact that raw scores were used (which do not reflect normative ranks), and by the fact that the comparisons involving the recognition format did not take into account incorrect responses.

For all of these reasons, it was decided that, despite the appeal of the CVLT–C, clinical interpretation is currently hindered by the fact that insufficient psychometric information is available to facilitate direct comparisons of the various variables that can be obtained from this instrument. The purpose of this investigation was to determine the base rate of specific discrepancies between selected CVLT–C

https://doi.org/10.1017/S1355617799511041 Published online by Cambridge University Press

variables in the standardization sample, and to determine what levels of discrepancy are required for statistical significance.

METHODS

Research Participants

The 920 children from the CVLT–C standardization sample were used for this study. This sample is fairly representative of the population of the United States in terms of ethnicity, parental education, and geographic region, with approximately equal numbers of boys and girls at each of 12 age levels.

Materials

The CVLT–C is an individually administered measure of a child's ability to learn and remember verbally presented information. It contains two word lists that each contain 15 shopping items, including five words from each of three semantic categories. There are five trials of full presentation and immediate reproduction of the first list (A), followed by one-time presentation and immediate reproduction of the second list (B). Variables of free recall and semantically cued recall of list A are obtained immediately after the trial with List B, and again after a 20-min delay. Finally, a recognition trial is presented in which the child is asked to identify the 15 items from List A from a larger list containing distractor items. The task is the same for all ages, ranging from 5 to 16 years. Performance on the CVLT–C is characterized in terms of a summary *T* score ($M = 50$, $SD = 10$), reflecting a global index of immediate free recall over five successive trials of the first list, as well as in *z* scores ($M = 0$, $SD = 1$) for a variety of other quantitative (e.g., number of words recalled) and qualitative (e.g., learning strategy) variables.

Four specific contrasts are offered by the CVLT–C format. One concerns the difference between total numbers of correct words recalled on, respectively, the second list (B) and the first trial of List A (A1). This contrast (PI) is thought to reflect susceptibility to proactive interference (Delis et al., 1994): Poor performance on B, as compared to A1, may be the result of a decremental effect of prior learning on subsequent learning.

A second CVLT–C contrast concerns the savings on the short delay free recall trial (SD) of correct words that the child recalled on the fifth trial of List A (A5). This contrast (RI) is purported to reflect susceptibility to retroactive interference (Delis et al., 1994): Poor performance on SD, as compared to A5, may be the result of a decremental effect of new learning (involving List B) on previous learning.

The savings on the long delay free recall trial (LD) of correct words that the child recalled on SD is a third contrast that is offered by the CVLT–C. This contrast (RF) is proposed as a reflection of rapid forgetting (Delis et al., 1994): Poor performance on LD, as compared to SD, may be the result of an increased rate of forgetting during the delay interval.

The fourth and final CVLT–C contrast concerns a comparison between discriminability (DI, a variable of overall performance on the recognition trial that takes into account both correct and incorrect responses) and LD. This contrast (RP) is supposed to reflect retrieval problems (Delis et al., 1994): A significantly better score on DI, as compared to LD, may indicate that the child had difficulty with the independent recollection of information but that the material was not truly lost.

Procedure

The following primary CVLT–C variables were included in this investigation: total words correctly recalled at the first trial of the first list (A1); total words correctly recalled at the fifth trial of the first list (A5); total words correctly recalled at the second list trial (B); total words correctly recalled on the short delay free recall trial of list A (SD); total words correctly recalled on the long delay free recall trial of list A (LD); and discriminability (DI), determined by the following formula: $DI = 100 \times (1 - X/45)$, where X is the sum of the numbers of, respectively, false positives and misses on the recognition trial. The *z* scores for each of these six variables were used, instead of raw scores, in order to allow direct comparability of the scores. Contrasts were then calculated by subtracting the respective *z* scores from each other, according to the following formulas: Proactive Interference (PI) = B − A1, Retroactive Interference (RI) = SD − A5, Rapid Forgetting (RF) = LD − SD, and Retrieval Problems (RP) = DI − LD.

It is important to realize that the contrast variables themselves are not *z* scores but merely reflect the numeric difference between the component *z* scores. In addition, the direction of the difference between the components of each contrast variable is of great importance in the interpretation of CVLT–C results. True Proactive Interference would be reflected in a negative value of PI (i.e., B < A1), true Retroactive Interference in a negative value of RI (i.e., SD < A5), true Rapid Forgetting in a negative value of RF (i.e., LD < SD), and true Retrieval Problems in a positive value of RP (i.e., DI > LD).

The minimum contrast values that are required for statistical significance at the .10, .05, and .01 levels were calculated on the basis of the standard error of estimate of the difference between the respective scores. Multiplying this statistic by an appropriate factor yields the amount of the difference that is statistically significant at any given level of confidence. Because of the absence of information about the reliability of some of the scores, the simple difference method was used for this purpose (Mittenberg et al., 1991; Reynolds, 1984). The associated formula is as follows: $D = Z \times SD \times \sqrt{(2 - 2r_{xy})}$, where D is the difference score, Z is the normal curve value associated with the desired level of statistical significance (i.e., 1.28 for an alpha of .10, 1.65

https://doi.org/10.1017/S1355617799511041 Published online by Cambridge University Press

for an alpha of .05, and 2.33 for an alpha of .01), *SD* is the standard deviation of the scales being compared (i.e., 1 in these cases), and r_{xy} is the correlation between the scales.

A variety of other CVLT–C variables were also included in this investigation in order to evaluate possible reasons why some children might have unusually large contrast variables. These variables included the following: proportion of recall of words in groups from the same semantic category across consecutive presentations of List A (Semantic Clustering); percentage of correct words recalled from the end region across consecutive presentations of List A (Recency Effect); percentage of correct words consistently recalled across consecutive presentations of List A (Recall Consistency); total number of repetitions of any response on the same trial across all free and cued recall trials (Perseverations); and total number of incorrectly recalled words across all free and cued recall trials (Intrusions). Again, the *z* scores for each of these five secondary variables were used, instead of raw scores, in order to allow direct comparability of the scores.

RESULTS

Table 1 presents by age group the minimum differences between a child's specific CVLT–C *z* scores that are required for statistical significance at three levels of confidence: .10, .05, and .01. For example, a 10-year-old child would need to have a difference of at least 1 *z*-score point between its A5 and SD performances (reflected in the RI contrast) in order for this discrepancy to be considered statistically significant with 90% confidence.

Because of the well-known fact that statistical significance does not necessarily imply clinical significance (Silverstein, 1981), the base rates of specific sizes of CVLT–C contrast discrepancies were also evaluated. These data are presented in Table 2. Inspection of Table 2 reveals that seemingly large magnitudes of discrepancies between specific *z* scores are not uncommon in the CVLT–C standardization sample. For example, more than a quarter of the standardization sample had a RI discrepancy (considered regardless of the direction of the difference) that equaled or exceeded 1 standard deviation. However, as was stated above, the direction of the difference is of great significance in the interpretation of CVLT–C results. Inspection of Table 2 suggests that for PI, RI, and RF, fairly large negative values (i.e., < -1) are about as common as fairly large positive values (i.e., > 1). For RP, however, positive values of this magnitude (occurring in approximately 12% of the standardization sample) are twice as common as negative values of the same size (occurring in approximately 6% of the standardization sample).

When taking into account the direction of performance patterns, it is common practice to consider clinical phenomena that occur in less than approximately 10% of the general pediatric population as extraordinary or unusual (Achenbach, 1991; Donders, 1997). Considering the data in Table 2, this could be defined in the current context as PI values ≤ -1.5, RI values ≤ -1.5, RF values ≤ -1, and RP values ≥ 1.5. It was considered to be of interest to determine if children who displayed any such unusual contrasts would demonstrate characteristics on other CVLT–C variables that might add insights into the reasons for their unusual performance. For these purposes, the performance of all children who did or did not meet at least one of the above criteria for unusually large contrasts was considered on sev-

Table 1. Magnitudes (in *z*-score units) of CVLT–C contrast variables required for statistical significance at three levels, by age

Age (years)	Level	PI	RI	RF	RP
5	.10	1.36	1.56	1.25	1.46
	.05	1.75	2.01	1.62	1.88
	.01	2.47	2.84	2.28	2.66
6	.10	1.55	1.31	1.18	1.37
	.05	2.00	1.68	1.52	1.77
	.01	2.82	2.38	2.14	2.49
7	.10	1.43	1.17	1.17	1.14
	.05	1.85	1.50	1.50	1.47
	.01	2.61	2.12	2.12	2.07
8	.10	1.51	1.28	.88	1.34
	.05	1.95	1.65	1.14	1.73
	.01	2.75	2.33	1.61	2.45
9	.10	1.34	1.06	.91	1.19
	.05	1.73	1.37	1.17	1.54
	.01	2.45	1.93	1.65	2.17
10	.10	1.41	1.00	.85	1.19
	.05	1.82	1.29	1.09	1.54
	.01	2.56	1.82	1.54	2.17
11	.10	1.41	1.09	1.08	1.02
	.05	1.82	1.40	1.39	1.32
	.01	2.56	1.98	1.96	1.86
12	.10	1.34	.95	.85	.96
	.05	1.73	1.22	1.09	1.24
	.01	2.45	1.72	1.54	1.75
13	.10	1.54	1.25	.91	1.29
	.05	1.98	1.62	1.17	1.67
	.01	2.80	2.28	1.65	2.35
14	.10	1.45	.91	.79	1.14
	.05	1.87	1.17	1.02	1.47
	.01	2.63	1.65	1.45	2.07
15	.10	1.42	1.00	1.11	.95
	.05	1.83	1.29	1.44	1.22
	.01	2.59	1.82	2.03	1.72
16	.10	1.43	1.28	.83	.95
	.05	1.85	1.65	1.07	1.22
	.01	2.61	2.33	1.52	1.72
All ages	.10	1.43	1.18	1.00	1.23
	.05	1.85	1.52	1.29	1.58
	.01	2.61	2.14	1.82	2.24

Note. PI = Proactive Interference; RI = Retroactive Interference; RF = Rapid Forgetting; RP = Retrieval Problems. Standardization data of the California Verbal Learning Test–Children's Version (CVLT–C).

https://doi.org/10.1017/S1355617799511041 Published online by Cambridge University Press

Table 2. Cumulative percentages of the CVLT–C standardization sample obtaining various values of contrast variables

Value	PI	RI	RF	RP
≤ −3	1	1	1	1
−2.5	3	1	1	2
−2	5	3	1	3
−1.5	11	7	3	6
−1	24	17	10	15
−0.5	41	36	32	32
0	61	63	65	55
0.5	77	85	86	75
1	87	95	96	88
1.5	94	98	99	97
2	98	99	99	99
2.5	99	100	100	100
≥3	100	100	100	100

Note. PI = Proactive Interference; RI = Retroactive Interference; RF = Rapid Forgetting; RP = Retrieval Problems. Standardization data of the California Verbal Learning Test–Children's Version (CVLT–C).

eral quantitative and qualitative CVLT–C variables pertaining to learning characteristics and recall errors. Only variables that were not dependent on the values of the variables contributing directly to the contrasts, and that were not interdependent with each other, were selected in this regard. These data are presented in Table 3.

Inspection of Table 3 reveals that neither the variables pertaining to learning characteristics (Semantic Clustering, SEM; Recency Effect, REC; and Recall Consistency, CON) nor the variables pertaining to recall errors (Perseverations, PSV; and Intrusions, INT) showed distinct patterns for any of the groups of children who met one or more of the criteria for unusual contrasts. The average performance of these children was within a .5 standard deviation from the mean on all of these variables. Statistical analyses between these four groups with significantly large contrasts were not performed because some children had two ($N = 61$) or three ($N = 3$) unusual contrasts, and therefore the groups were not independent of each other. When children had more than one unusual contrast, a RP value ≥ 1.5 was clearly most often one of them.

Even though no distinct pattern was clearly evident for any of the four groups with unusual contrasts, it would still be possible that there could be differences in learning style between children with and without unusual contrasts. For this purpose, separate multivariate analyses of variance (MANOVAs) were performed to compare children who did or did not display specific individual contrasts of sufficiently large magnitude (see Table 3). In each of these analyses, there were five dependent variables (SEM, REC, CON, PSV, and INT). *Post-hoc* comparisons were performed with the Bonferroni correction of alpha to .01 (.05/5) in each case.

The MANOVA for PI yielded a statistically significant main effect of groups [$F(5,914) = 3.22$, $p < .01$]. *Post-hoc* comparisons revealed that children with PI effects ≤ −1.5 had significantly higher SEM *z* scores than children who did not have such an unusual contrast [$t(918) = 3.45$, $p < .001$]. The other group differences on this variable did not meet the *a priori* established level of statistical significance.

The MANOVA for RI yielded a statistically significant main effect of groups [$F(5,914) = 6.31$, $p < .0001$]. *Post-hoc* comparisons revealed that children with RI effects ≤ −1.5 had significantly higher INT *z* scores than children who did not have such an unusual contrast [$t(918) = 4.80$, $p < .0001$]. The other group differences on this variable did not meet the *a priori* established level of statistical significance.

The MANOVA for RF yielded a statistically significant main effect of groups [$F(5,914) = 2.77$, $p < .05$]. *Post-hoc* comparisons revealed that children with RF effects ≤ −1 had significantly lower CON *z* scores than children who did not have such an unusual contrast [$t(918) = -3.03$, $p <$

Table 3. CVLT–C performance of children with and without unusual contrast values

		SEM		REC		CON		PSV		INT	
Contrast	*N*	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
PI ≤ −1.5	101	.31	(1.06)	−.04	(.76)	−.01	(.87)	.13	(1.10)	−.18	(1.02)
PI > −1.5	819	−.08	(1.05)	.06	(1.03)	.01	(.95)	−.04	(1.09)	−.12	(1.02)
RI ≤ −1.5	68	−.09	(.99)	−.06	(1.00)	.20	(.75)	.13	(1.25)	.43	(1.51)
RI > −1.5	852	−.03	(1.06)	.06	(1.00)	−.01	(.95)	−.03	(1.08)	−.17	(.95)
RF ≤ −1	92	−.23	(.90)	.03	(1.08)	−.28	(1.06)	−.05	(1.26)	.06	(1.14)
RF > −1	828	−.01	(1.07)	.06	(1.00)	.04	(.92)	−.02	(1.07)	−.15	(1.00)
RP ≥ 1.5	114	−.14	(1.05)	.20	(1.10)	−.39	(1.16)	−.38	(.73)	−.15	(1.06)
RP < 1.5	806	−.01	(1.06)	.03	(.99)	.06	(.89)	.03	(1.13)	−.13	(1.01)

Note. Population standard scores $M = 0$ and $SD = 1$ for all variables. PI = Proactive Interference; RI = Retroactive Interference; RF = Rapid Forgetting; RP = Retrieval Problems; SEM = Semantic Clustering; REC = Recency Effect; CON = Recall Consistency; PSV = Perseverations; INT = Intrusions. Standardization data of the California Verbal Learning Test–Children's Version (CVLT–C).

https://doi.org/10.1017/S1355617799511041 Published online by Cambridge University Press

.01]. The other group differences on this variable did not meet the *a priori* established level of statistical significance.

The MANOVA for RP yielded a statistically significant main effect of groups [$F(5,914) = 7.25$, $p < .0001$]. *Post-hoc* comparisons revealed that children with RP effects ≥ 1.5 had significantly lower CON *z* scores [$t(918) = -4.84$, $p < .0001$], and significantly lower PSV *z* scores [$t(918) = -3.78$, $p < .001$], than children who did not have such an unusual contrast. The other group differences on this variable did not meet the *a priori* established level of statistical significance.

DISCUSSION

The purpose of this investigation was to determine the base rates of specific discrepancies between key CVLT–C variables in the standardization sample, and to determine what level of contrast variables would be required for statistical significance. The results indicate that for several of the contrast variables at various age levels, values as small as 1 standard deviation can sometimes be considered to be statistically significant at a level (90% confidence) that has been suggested by some authors (e.g., Kaufman, 1994) as being sufficient for most clinical situations. However, inspection of the base rates suggests that such discrepancies are actually not all that unusual. Therefore, statistically significant differences between specific CVLT–C variables are not necessarily clinically significant. These findings are consistent with those reported in investigations of other memory instruments used with adults, such as the Wechsler Memory Scale–Third Edition (Wechsler, 1997).

The decision about which level of statistical confidence to consider will also depend on how many comparisons one makes in an individual case. If one compares all four contrasts, the chance of Type 1 error is greatly increased. For example, the chance would not be .1 but .4 that at least one of the particular contrasts would fall outside of the 90% confidence range. Therefore, if multiple comparisons are made, a more conservative level of statistical significance may be indicated. On the other hand, guarding too much against Type 1 error (rejecting a true null hypothesis) may spuriously inflate the chance of Type 2 error (accepting a false null hypothesis). Ideally, one would like to address this matter by the calculation of confidence intervals for each of the CVLT–C *z* scores. However, there is simply no information about the reliability of all those variables. For all of these reasons, a feasible rule of thumb for clinical practice may be the following: (1) a 90% confidence interval may suffice if only one or two contrasts needs to be considered, (2) a 95% confidence interval may be indicated when making three or four contrast comparisons, and (3) a 99% confidence interval may be needed only in unusual situations (e.g., to demonstrate convincingly the need for specific costly cognitive rehabilitation).

The current findings also suggest that qualitative CVLT–C variables pertaining to learning characteristics, and quantitative variables pertaining to recall errors, do not necessarily provide clear information with regard to why a particular child might obtain a large discrepancy between specific quantitative CVLT–C variables of accurate recall. For example, contrary to clinical lore that Semantic Clustering is associated with better learning, it was found that children with unusually large Proactive Interference effects actually engaged relatively more in this strategy than children who did not have such an unusual contrast. Similarly, children with unusually large degrees of Retrieval Problems were found to provide fewer Perserverations than children who did not have such an unusual contrast. There were no differences between any of the groups with regard to the variable Recency Effect. On the other hand, the variables Recall Consistency and Intrusions did demonstrate group differences that were consistent with expectation. Specifically, a higher number of Intrusions was associated with the presence of an unusually large Retroactive Interference effect, and children who had unusually large contrasts on either Rapid Forgetting or Retrieval Problems tended to have lower degrees of Recall Consistency than children who did not have such significant contrasts.

The inconsistent findings in the area of learning characteristics and recall errors may be related to the fact that some of these quantitative variables (e.g., Perseverations) had poor sampling characteristics, and some of these qualitative variables (e.g., Semantic Clustering) were only weakly associated with reliable theoretical constructs, in a recent investigation of the factor structure of the CVLT–C (Donders, in press). In contrast, high factor loadings were consistently found in that study for the variables contributing directly to the contrasts that were the focus of the present investigation. Combined, these findings suggest that the most reliable basis for clinical interpretation of CVLT–C protocols may be the *z* scores for key quantitative variables, and the associated contrasts.

Inspection of the base rates for contrast values, with consideration of (1) the direction of the difference between the respective variables, and (2) the above-mentioned risk for increased Type 1 error when making multiple comparisons, suggests that *z*-score discrepancies can be considered in a valid manner when this is done cautiously. Specifically, the following guidelines for the interpretations of contrasts on the CVLT–C are proposed at this time.

An unusual degree of *Proactive Interference* can be considered when the *z* score for recall on the first trial of the first list (A1) exceeds the *z* score for recall of the second list (B) by at least 1.5 points. An unusual degree of *Retroactive Interference* can be considered when the *z* score for recall on the fifth trial of the first list (A5) exceeds the *z* score for the short delay free recall trial of List A (SD) by at least 1.5 points. An unusual degree of *Rapid Forgetting* can be considered when the *z* score for SD exceeds the *z* score for the long delay free recall trial of List A (LD) by at least 1 point. A *Retrieval Problem* can be considered when the *z* score for the discriminability index (DI) exceeds the *z* score for LD by at least 1.5 points.

It must also be realized that some children in the CVLT–C standardization sample demonstrated large contrasts in a direction that was opposite of what might be expected on the basis of the theoretical assumptions underlying those contrasts. In fact, on three of the four contrasts, fairly large positive values appeared to occur about just as frequently as fairly large negative values. Possible explanations (which must be considered to be only speculative at this time) may be the following.

A *z* score on B that is more than 1 standard deviation better than the *z* score on A1 may raise the question whether the child was initially anxious on the test and then settled down after repeated practice with list A. A *z* score on SD that is more than 1 standard deviation better than the *z* score on A5 may be related to inconsistencies in retrieval (in which case facilitation of recall on the recognition trial would be expected). A *z* score on LD that is at least 1 standard deviation better than the *z* score on SD may suggest that the child learned to remember some more words because of the semantic category prompts that were provided at the time of the short delay cued recall trial. Deterioration of performance on the recognition trial, as evidenced by a *z* score on DI that is significantly worse than the *z* score on LD, is much less common in the CVLT–C standardization sample than facilitation of retrieval on the recognition trial. When examiners encounter the former pattern, an unusual response set (possibly related to motivation issues) may need to be considered.

In summary, the current findings support the previously suggested approach (Donders, in press) of focusing first and foremost on quantitative variables when interpreting CVLT–C protocols of individual children. The global *T* score is a reliable summary index of overall efficacy of the child's learning of list A. The *z* scores on variables such as A1, A5, B, SD, LD, and DI, and the presence of unusually large contrasts between these variables, may provide important additional information about the child's memory abilities. However, CVLT–C results should never be interpreted in isolation. None of the interpretive considerations offered here should be construed as definite diagnoses unless they can be corroborated by evidence from other independent instruments during a broader neuropsychological evaluation. A goal for future research is to evaluate the frequency of occurrence of unusually large CVLT–C *z* score contrasts in specific clinical samples.

ACKNOWLEDGMENTS

This research was supported by a grant from the Mary Free Bed Fund (Project #74). I thank the Psychological Corporation for allowing access to the California Verbal Learning Test–Children's Version (CVLT–C) national sample. This research was based on standardization data of the CVLT–C.

REFERENCES

Achenbach, T.M. (1991). *Integrative guide for the 1991 CBCL/4–18, YSR, and TRF profiles*. Burlington, VT: University of Vermont.

Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, B.A. (1994). *California Verbal Learning Test–Children's Version*. San Antonio, TX: The Psychological Corporation.

Donders, J. (1997). A short form of the WISC–III for clinical use. *Psychological Assessment*, *9*, 15–20.

Donders, J. (in press). A structural equation analysis of the California Verbal Learning Test–Children's Version in the standardization sample. *Developmental Neuropsychology*.

Fletcher, J.M., Levin, H.S., & Butler, I.J. (1995). Neurobehavioral effects of brain injury on children: Hydrocephalus, traumatic brain injury, and cerebral palsy. In M.C. Roberts (Ed.), *Handbook of pediatric psychology* (2nd ed., pp. 362–383). New York: Guilford.

Jaffe, K.M., Fay, G.C., Polissar, N.L., Martin, K.M., Shurtleff, H.A., Rivara, J.B., & Winn, H.R. (1993). Severity of pediatric traumatic brain injury and neurobehavioral recovery at one year—A cohort study. *Archives of Physical Medicine and Rehabilitation*, *74*, 587–595.

Kaufman, A.S. (1994). *Intelligent testing with the WISC–III*. New York: Wiley & Sons.

Levin, H.S., Culhane, K.A., Mendelsohn, D., Lilly, M.A., Bruce, D., Fletcher, J.M., Chapman, S.B, Harward, H., & Eisenberg, H.M. (1993). Cognition in relation to magnetic resonance imaging in head-injured children and adolescents. *Archives of Neurology*, *50*, 897–905.

Mittenberg, W., Thompson, G.B., & Schwartz, J.A. (1991). Abnormal and reliable differences among Wechsler Memory Scale–Revised subtests. *Psychological Assessment*, *3*, 492–495.

Reynolds, C.R. (1984). Critical measurement issues in learning disabilities. *Journal of Special Education*, *18*, 451–475.

Silverstein, A.B. (1981). Reliability and abnormality of test score differences. *Journal of Clinical Psychology*, *37*, 392–394.

Wechsler, D. (1997). *Wechsler Memory Scale–Third Edition*. San Antonio, TX: The Psychological Corporation.

Yeates, K.O., Blumenstein, E., Patterson, C.M., & Delis, D.C. (1995a). Verbal learning and memory following pediatric closed-head injury. *Journal of the International Neuropsychological Society*, *1*, 78–87.

Yeates, K.O., Enrile, B.G., Loss, N., Blumenstein, E., & Delis, D.C. (1995b). Verbal learning and memory in children with myelomeningocele. *Journal of Pediatric Psychology*, *20*, 801–815.

Ylvisaker, M., Hartwick, P., Ross, B., & Nussbaum, N. (1994). Cognitive assessment. In R.C. Savage & G.F. Wolcott (Eds.), *Educational dimensions of acquired brain injury* (pp. 69–119). Austin, TX: PRO-ED.