The Value of the WISC-IV Digit Span Subtest in Detecting Noncredible Performance during Pediatric Neuropsychological Examinations[†]

Michael W. Kirkwood*, David D. Hargrave, John W. Kirk

Department of Physical Medicine & Rehabilitation, The Children's Hospital, Colorado and University of Colorado School of Medicine, Aurora, CO, USA

*Corresponding author at: The Children's Hospital, Concussion Program, B285, 13123 E. 16th Ave., Aurora, CO 80045, USA. Tel.: +720-777-6193; fax: +720-777-7297.
E-mail address: kirkwood.michael@tchden.org (M.W. Kirkwood).

Abstract

In adult populations, research on methodologies to identify negative response bias has grown exponentially in the last two decades. Far less work has focused on methods appropriate for children. Although several recent studies have demonstrated the appropriateness of using stand-alone symptom validity tests with younger populations, a near absence of pediatric work has investigated embedded validity indicators. The present study examined the classification value of several scores derived from the WISC-IV Digit Span subtest. The sample consisted of 274 clinically referred mild traumatic brain injury patients aged 8 through 16 years. Fourteen percent of the participants failed both the Medical Symptom Validity Test and Test of Memory Malingering, which was used as the criterion for noncredible effort. For age-corrected scaled scores, a score of ≤ 5 resulted in the optimal cut-score, yielding sensitivity of 51% and specificity of 96%. For Reliable Digit Span, the optimal cut-score was ≤ 6, with sensitivity of 51% and specificity of 92%. Although only moderately sensitive, Digit Span scores are likely to have good utility in identifying noncredible performance in relatively high-functioning older children and adolescents. Indeed, classification statistics produced in this pediatric sample compare favorably with those produced in many real-world adult patients.

Keywords: Digit span; Reliable digit span; Wechsler intelligence scale for children; Symptom validity testing; Response bias; Postconcussion; Mild traumatic brain injury

Introduction

In comparison to the vast literature focused on noncredible neuropsychological performance in adults, the pediatric literature is relatively sparse. However, a number of single-case reports have clearly documented that children can feign cognitive impairment during neuropsychological examination (Flaro & Boone, 2009; Henry, 2005; Kirkwood, Kirk, Blaha, & Wilson, 2010; Lu & Boone, 2002; McCaffrey & Lynch, 2009). Several recent clinical case series have also consistently found that a small percentage of general pediatric patients perform suboptimally because of effort-related problems (Carone, 2008; Donders, 2005; Kirk et al., 2011; MacAllister, Nakhutina, Bender, Karantzoulis, & Carlson, 2009). Two other recent studies suggest that, under certain conditions, rates of negative response bias in children are likely to be considerably higher. In a mild traumatic brain injury (TBI) case series of ours consisting of 193 children and adolescents referred exclusively for clinical neuropsychological evaluation, 17% of the sample failed the Medical Symptom Validity Test (MSVT), which was the same percentage estimated to have put forth noncredible effort more broadly across the examinations (Kirkwood & Kirk, 2010). Chafetz, Abrahams, and Kohlmaier (2007) found an even higher percentage of children (28–37%) who failed a symptom validity test (SVT) during determination evaluations for Social Security Disability benefits.

The two primary objective methods to evaluate the validity of an individual's neuropsychological performance are stand-alone SVTs and indices derived from conventional tests or "embedded indicators." In adult populations, an extensive

---

[†] A preliminary version of this paper was presented at the annual meeting of the American Academy of Clinical Neuropsychology in Chicago, June 2010.

body of evidence supports both approaches (Boone, 2007; Larrabee, 2007). Relying on objective tools to determine negative response bias in pediatric populations is no less important, because subjective clinical judgment alone is unlikely to be consistently effective (Faust, Hart, & Guilmette, 1988; Faust, Hart, Guilmette, & Arkes, 1988). Over the last decade, a number of studies have demonstrated that certain stand-alone SVTs can be used appropriately with pediatric populations (Kirkwood, in press). For example, pediatric patients down to age 5 or 6 years can pass the Test of Memory Malingering (TOMM; Constantinou & McCaffrey, 2003; Donders, 2005; Kirk et al., 2011; MacAllister et al., 2009). The Word Memory Test (WMT) and Computerized Assessment of Response Bias (CARB) require some facility with reading and numbers, so are inappropriate for early elementary-aged children; however, on the primary effort indices, children who are older than 10 years score above adult cutoffs (Courtney, Dinkins, Allen, & Kuroski, 2003). On the WMT, no age effect has been found in children with at least a third-grade reading level (Green & Flaro, 2003). The MSVT is similar to the WMT but is shorter and easier and has been successfully used in children as young as second or third grade (Blaskewitz, Merten, & Kathmann, 2008; Carone, 2008; Kirkwood & Kirk, 2010).

In contrast to this burgeoning pediatric research focused on stand-alone SVTs, a near absence of published work has focused on the utility of embedded measures in youth populations. In adults, the value of embedded indicators is well established, as they are time efficient, resistant to coaching, and allow for more continuous monitoring of effort than stand-alone SVTs. The Digit Span subtest from the Wechsler instruments is one of the most thoroughly investigated of all embedded measures. Babikian and Boone (2007) and Suhr and Barrash (2007) have recently reviewed dozens of adult studies focused on using Digit Span as an SVT. Classification statistics vary considerably across studies. Generally speaking, sensitivity is higher in simulators than real-world patients and specificity is higher in mildly affected samples like mild TBI than those with more significant neurological problems like stroke or moderate/severe TBI. Across studies, a Digit Span age-corrected scaled score (ACSS) of ≤ 5 has typically been associated with $>90\%$ specificity, with sensitivity ranging from about 25% to 50% at this cutoff (Axelrod, Fichtenberg, Millis, & Wertheimer, 2006; Babikian, Boone, Lu, & Arnold, 2006).

In addition to ACSS, other indices of poor effort derived from Digit Span performance have also been examined. The most well known of these is Reliable Digit Span (RDS), introduced by Greiffenstein, Baker, and Gola (1994). Reliable Digit Span is calculated by summing the longest string of digits repeated without error over two trials under both forward and backward conditions. Using an RDS, cutoff of ≤ 8 or ≤ 7 has produced sensitivity values above 50% in nearly all adult studies, though specificity at this level has been found to be unacceptably low in more severely affected clinical populations (Etherton, Bianchini, Ciota, & Greve, 2005; Greiffenstein, Gola, & Baker, 1995; Larrabee, 2003; Ylioja, Baird, & Podell, 2009). When RDS is set to ≤ 6, sensitivity is lowered to 40% to 60%, but specificity improves more consistently to at least 90% (Babikian et al., 2006; Duncan & Ausborn, 2002).

Only one identified pediatric study has examined Digit Span performance as an indicator of noncredible neuropsychological performance. In a simulation study with 38 German children aged 6–12 years, Blaskewitz et al. (2008) administered a neuropsychological battery that included the Wechsler Intelligence Scale for Children-Third Edition Digit Span subtest. The mean RDS score for the experimental malingerers was 2.7 ($SD = 2.7$), with 90% failing at a cutoff score of 7/8. However, the majority of the matched controls (59%) also failed using this cutoff score, supporting the sensible idea that RDS cutoffs used for adults are unlikely to be appropriate for young children. Unfortunately, the authors did not publish the classification statistics for lower RDS cut scores or for other Digit Span scores.

In the present study, we set out to examine the classification statistics of the WISC-IV Digit Span scaled scores, RDS, and raw scores from Forward and Backward conditions in detecting noncredible performance in real-world pediatric patients. The sample consisted of 274 school-aged children and adolescents referred consecutively for outpatient clinical neuropsychological consultation following a mild TBI. We anticipated significant differences for each Digit Span variable between the group of children classified as providing credible effort versus those classified as providing noncredible effort. In this relatively mildly affected pediatric sample, we also anticipated to find classification statistics comparable with those apparent in studies with real-world adult patients (i.e., sensitivity near 50% when specificity $>90\%$).

Method

Participants

The study was reviewed and approved by the university-affiliated institutional review board. Participants were drawn from a 4-year series of consecutive clinical cases referred to an outpatient concussion program at a children's hospital in the Rocky Mountain region of the United States. Patients were considered eligible for participation if they were administered the WISC-IV Digit Span subtest, were aged 8 through 16 years at the time of evaluation, were within 1 year of sustaining a blunt head trauma, and were referred because of concerns or questions about the effects of underlying brain injury. Other

than a few select cases in which the head injury was unwitnessed and reliable acute injury data were unavailable, patients displayed evidence of concussion such as alteration in mental status, loss of consciousness, post-traumatic amnesia, or transient neurologic disturbance. The most common causes of injury were recreation or sports (65%), falls (18%), motor vehicle-related trauma (11%), and assaults (3%). Children who had intracranial pathology on neuroimaging were included if their Glasgow Coma Scale (GCS) score was never <13, consistent with most common definitions of mild TBI (e.g., Carroll, Cassidy, Holm, Kraus, & Coronado, 2004). Exclusionary criteria were: forensic referral, neurosurgical intervention, injury resulting from abuse, and non-TBI such as hypoxia, stroke, or infectious illness. If a patient was evaluated more than once clinically, only the first encounter data were utilized. The final sample included 274 participants. Background and injury characteristics of the sample are provided in Table 1.

Measures

The Digit Span subtest from the WISC-IV (Wechsler, 2003) consists of two parts: Digit Span Forward and Digit Span Backward. For Digit Span Forward, the child repeats increasingly longer strings of numbers in the same order as presented aloud by the examiner. For Digit Span Backward, the child repeats increasingly longer strings of numbers in reverse order of that presented aloud by the examiner. For both Digit Forward and Backward, each number is read at a rate of one per second. Administration is discontinued when both items from a given pair are failed. Age-corrected scaled scores (ACSS) and raw scores from Digit Span Forward and Digit Span Backward were recorded. Reliable Digit Span was calculated according to the guidelines recommended by Greiffenstein and colleagues (1994; sum of the longest string of digits repeated without error over two trials under both forward and backward conditions). A score of 2 was assigned to individuals who failed at least one trial each of the floor items (two forward and two backward).

The MSVT is a computerized forced-choice verbal memory test designed to evaluate effort and memory. The primary effort indices are the Immediate Recognition (IR), Delayed Recognition (DR), and Consistency (CNS) scores. The test requires about 5 min of direct administration time (i.e., not including the delay time between IR and DR). Examinees are presented with 10 semantically related word pairs twice on a computer screen. They are then asked to choose the correct word from pairs consisting of the target and a foil, during IR and DR conditions. After each response, examinees receive auditory and visual feedback. Examinees are then asked to recall the words during paired associate (PA) and free recall (FR) conditions. Participants in the current project were administered the MSVT in a standardized fashion, except that the examiner stayed in the room during the entire administration. The actuarial criteria proposed by Green (2004) were considered indicative of suboptimal effort.

The TOMM is a 50-item forced-choice recognition test designed to detect individuals who exaggerate or fake memory impairment. It includes two learning trials, two recognition trials, and an optional retention trial. On each learning trial, the examinee is presented 50 line drawings one at a time. Examinees are then asked to choose the correct drawing from a pair

Table 1. Background and injury characteristics of participants

Total participants	$N = 274$
Age	$M = 14.2, SD = 2.2$
Grade	$M = 8.3, SD = 2.2$
Male	$n = 172 (63\%)$
Caucasian	$n = 230 (84\%)$
Estimated Full Scale IQ[a]	$M = 103.5, SD = 12.6$
Maternal years of education	$M = 15.1, SD = 2.2$
Paternal years of education	$M = 15.2, SD = 2.6$
Premorbid history of attention-deficit/hyperactivity disorder	$n = 45 (16\%)$
Premorbid history of diagnosed learning disability	$n = 29 (11\%)$
Premorbid history of special education services	$n = 35 (13\%)$
Weeks since injury (range 1–52 weeks)	$M = 9.7, SD = 9.1$
	$Mdn = 6.0$
Loss of consciousness	$n = 49 (18\%)$
Neuroimaging conducted	$n = 199 (73\%)$
Intracranial findings identified by CT or MRI	$n = 27 (10\%)$
Families in or planning litigation	$n = 22 (8\%)$
Families seeking disability compensation	$n = 0 (0\%)$
Participants charged with a crime	$n = 0 (0\%)$

Note: [a]Based on performance of the 261 participants administered the Wechsler Abbreviated Scale of Intelligence.

consisting of the target and a foil, during two recognition trials, with the examiner providing oral feedback after each response. The actuarial criterion on Trial 2 proposed by Tombaugh (1996) was considered indicative of suboptimal effort.

Procedure

Patients underwent testing at the earliest 1 week post-injury and at the latest 52 weeks post-injury. The median testing time was 6 weeks post-injury. Most children underwent an abbreviated battery of neuropsychological tests rather than a more comprehensive evaluation (as discussed in Kirkwood et al., 2008), though the actual tests administered varied depending on clinical need. The WISC-IV Digit Span subtest and MSVT were administered to all patients. The battery included the Wechsler Abbreviated Scale of Intelligence (WASI) for 95% of the sample and the Woodcock-Johnson III Tests of Achievement Letter-Word Identification subtest for 81% of the sample. If a patient failed any of the three primary effort indices of the MSVT, Trials 1 and 2 of the TOMM were also administered (the TOMM retention trial was not administered consistently). Thus, the only patients who were administered the TOMM were those who first demonstrated evidence of suspect effort on the MSVT.

Of the 274 participants, 50 (18.2%) failed at least one of the three primary effort indices of the MSVT, with 37 (13.5%) failing both the MSVT and TOMM. Based on MSVT and TOMM performance, two distinct groups were formed: (i) a credible effort group consisting of all patients who passed the MSVT ($n = 224$) and (ii) a noncredible effort group consisting of those patients who failed both the MSVT and TOMM ($n = 37$). There was a small group of children ($n = 13$) who failed the MSVT but passed the TOMM. These children with more equivocal SVT performance were excluded in the classification analyses to reduce the chances of false-positive errors.

Results

Performance profiles in those participants who failed both the MSVT and TOMM (noncredible group) and those who did not (credible group) are provided in Table 2. The two groups did not differ in age, grade, or single word reading grade level, nor did they differ by gender, ethnic/racial status (classified as Caucasian or other), maternal education, history of premorbid LD, ADHD, or need for special education services, time since injury, or whether the injury was associated with loss of consciousness or neuroimaging pathology. Litigation status of the groups did not differ either. A total of 22 families reported that they were engaged in or planning litigation. Only two of these patients failed both the MSVT and TOMM. Thus, assuming honest reporting, a maximum of 2 out of the 37 cases failing both measures were potentially driven by compensation seeking.

Independent sample *t*-test results and effect sizes for the various Digit Span scores are presented in Table 2. To account for multiple comparisons, a *p* value of .01 was chosen a priori to identify significance. The noncredible effort group performed significantly worse across all Digit Span scores: ACSS ($t = 8.20$, $p < .001$), RDS ($t = 7.06$, $p < .001$), Forward raw scores ($t = 7.57$, $p < .001$), and Backward raw scores ($t = 6.09$, $p < .001$). Effect sizes were uniformly large (Cohen, 1988).

Sensitivity and specificity values and their respective cutoff scores for the various Digit Span scores are presented in Table 3. Calculation of cutoff scores was confirmed through receiver operating characteristic (ROC) area under the curve (AUC) analyses. The AUCs for ACSS (.85), RDS (.81), and Digits Forward raw score (.82) suggested excellent diagnostic accuracy, with the AUC for Digits Backward raw score (.79) just slightly lower. Cut scores were considered optimal when specificity was 90% or greater while maintaining maximum sensitivity. For ACSS, the optimal cutoff score was ≤5, resulting

Table 2. WISC-IV Digit Span performance in the credible and noncredible effort groups

	Credible effort group ($n = 224$; 62% male)				Noncredible effort group ($n = 37$; 68% male)				*p*	Cohen's *d*
	M	*SD*	*Mdn*	*Range*	*M*	*SD*	*Mdn*	*Range*		
Age	14.2	2.2	14.9	8–16.9	14.1	2.0	14.7	9–16.6	.775	
Grade	8.3	2.2	9.0	2–12	8.4	2.0	9.0	3–11	.911	
Single Word Reading Grade Level[a]	9.1	3.2	8.9	2.4–18.0	8.8	3.3	9.1	0.5–17.3	.622	
Digit Span Scaled Score	9.9	2.9	10.0	3–19	5.6	3.1	5.0	1–14	<.001	1.5
Reliable Digit Span	8.6	1.8	8.0	5–16	6.4	1.8	6.0	2–10	<.001	1.2
Digit Span Forward Raw Score	9.6	2.1	10.0	4–16	6.7	2.3	7.0	2–11	<.001	1.4
Digit Span Backward Raw Score	7.7	2.1	7.0	3–15	5.5	1.9	5.0	0–9	<.001	1.1

Note: [a]Based on performance of the 222 participants ($n = 190$ credible group; $n = 32$ noncredible group) administered the Woodcock Johnson Tests of Achievement Letter-Word Identification subtest.

Table 3. Sensitivity and specificity values for various WISC-IV Digit Span score cutoffs

	Sensitivity %	Specificity %
Digit Span Scaled Score		
≤3	22	100
≤4	38	99
≤5	51	96
≤6	68	89
≤7	78	76
≤8	81	67
≤9	84	53
≤10	95	39
Reliable Digit Span		
≤4	14	100
≤5	27	100
≤6	51	92
≤7	76	69
≤8	89	49
≤9	97	24
≤10	100	15
Digit Span Forward Raw Score		
≤4	19	100
≤5	24	99
≤6	41	94
≤7	62	84
≤8	81	68
≤9	87	54
≤10	97	31
Digit Span Backward Raw Score		
≤3	8	100
≤4	24	97
≤5	51	91
≤6	76	70
≤7	84	48
≤8	95	28
≤9	100	18
≤10	100	10

in sensitivity of 51% and specificity of 95%. For RDS, the optimal cutoff score was ≤6, yielding sensitivity of 51% and speci-ficity of 92%. For Forward raw scores, the optimal cutoff was ≤6, which resulted in sensitivity of 41% and specificity 93%. For Backward raw scores, the optimal cutoff was ≤5, yielding sensitivity of 51% and specificity of 91%.

Positive predictive value (PPV) is the proportion of participants below the cutoff in the noncredible effort group, whereas negative predictive value (NPV) is the proportion of participants above the cutoff in the credible group. Positive predictive value and NPV for Digit Span indices at noncredible effort base rates of 5%, 15%, 25%, and 40% are displayed in Table 4. A base rate of 15% (which approximates the rate in our sample) for Digit Span ACSS cutoff of ≤5 yielded a PPV of 69%, and an NPV of 92%. An RDS cutoff of ≤6 yielded PPV of 53% and NPV of 91% at a base rate of 15%. Utilizing a base rate of 25% with an ACSS cutoff of ≤5 resulted in PPV of 81% and NPV of 85%. At the same base rate, an RDS cutoff of ≤6 resulted in PPV of 68% and NPV of 85%.

Discussion

Though given scant attention historically, the current study supports the idea that adolescents and children down to at least age 8 years are capable of noncredible performance during neuropsychological evaluation, even in clinical contexts when sec-ondary gain may not be readily apparent. In this relatively large homogenous case series, 14% of the children performed below the actuarial cutoffs on *both* the MSVT and TOMM. The number of those failing the MSVT alone was 18%, comparable with the number of patients who were judged to have provided noncredible effort in an earlier version of the same case series once possible false-positives and false-negatives were taken into account (Kirkwood & Kirk, 2010). Although children are clearly capable of feigning cognitive symptoms in pursuit of financial gain, compensation-seeking behavior did not drive the majority

Table 4. PPVs and NPVs for select WISC-IV Digit Span scores at base rates of noncredible effort of 5%, 15%, 25%, and 40%

	5% Base rate		15% Base rate		25% Base rate		40% Base rate	
	PPV (%)	NPV (%)	PPV (%)	NPV (%)	PPV (%)	NPV (%)	PPV (%)	NPV (%)
Digit Span Scaled Score								
≤3	100	96	100	88	100	79	100	66
≤4	67	97	87	90	93	83	96	71
≤5	40	97	69	92	81	85	89	75
≤6	25	98	52	94	67	89	80	81
≤7	15	98	36	95	52	91	68	84
Reliable Digit Span								
≤4	100	96	100	87	100	78	100	64
≤5	100	96	100	89	100	80	100	67
≤6	25	97	53	91	68	85	81	74
≤7	11	98	30	94	45	90	62	81
Digit Span Forward Raw Score								
≤4	100	96	100	87	100	79	100	65
≤5	56	96	81	88	89	80	94	66
≤6	26	97	55	90	69	83	82	71
≤7	17	98	41	93	56	87	72	77
Digit Span Backward Raw Score								
≤3	100	95	100	86	100	77	100	62
≤4	30	96	59	88	73	79	84	66
≤5	23	97	50	91	65	85	79	74
≤6	12	98	31	94	46	90	63	81
≤7	8	98	22	94	35	90	52	82

of noncredible cases in this clinical sample. At the time of the neuropsychological contact, no cases were seeking disability compensation and only 2 out of the 22 cases engaged in or planning litigation were in the noncredible effort group. Of course, noncredible performance during a pediatric neuropsychological examination can occur for many reasons other than compensation-seeking (Kirkwood et al., 2010).

Multiple studies with adults have indicated that Digit Span performance has utility as an SVT. The primary purpose of the present study was to examine the value of Digit Span scores as an embedded indicator in an exclusively pediatric population consisting of real-world patients. When comparing the group who failed both the MSVT and TOMM with the group who did not, significant differences were found for each examined Digit Span variable (ACSS, RDS, Forward raw score, Backward raw score). Effect sizes for all differences were impressively large ($d > 1.0$), with the biggest effect seen for ACSS. Examination of the ROC area under the curve and inspection of the sensitivity and specificity values for each variable also suggested reasonable classification statistics, with ACSS serving as the best discriminator between groups.

Using an ACSS cutoff of ≤5 resulted in 51% sensitivity and 96% specificity. Classification statistics were similar when using a cutoff for RDS ≤6, whereby sensitivity was 51% and specificity was 92%. Setting RDS at ≤7, a value commonly employed in adult studies, resulted in an unacceptably high false-positive rate (31%) in this pediatric sample, supporting the intuitively sensible notion that RDS cut scores should be more conservative in children. Using ACSS ≤4 and RDS ≤5 reduced sensitivity to 38% and 27%, though specificity then increased to 99% and 100%, respectively.

Although the sensitivity and specificity values appear strong at the group level, clinicians are more interested in determining whether a specific SVT score suggests the presence of noncredible effort for an individual patient. Positive and negative predictive power have more relevance here, as they allow clinicians to state the probability that noncredible effort is or is not present given a specific test score. In contrast to sensitivity and specificity, predictive power estimates depend on the prevalence of the condition in the population of interest. Unfortunately, to date, very little attention has focused on the base rates of noncredible effort during neuropsychological examination in pediatric populations. As discussed in the introduction, the limited extant studies suggest that the base rate is anywhere between roughly 5% and 40%, depending on presenting condition and evaluative context. We provide positive and negative predictive power values across Digit Span variables to approximate the various base rates reported in available pediatric studies (5%, 15%, 25%, 40%).

The results of the present study need to be interpreted in the context of several limitations. The participants were drawn from a sample of convenience comprised of children and adolescents for whom persistent questions or concerns were apparent following a mild TBI. Because most youth can be expected to recover relatively quickly after such injury (Belanger & Vanderploeg, 2005; Carroll, Cassidy, Peloso et al., 2004; Maillard-Wermelinger et al., 2009), the participants are unlikely

to be representative of many patients with mild TBI. At the same time, all patients had suffered at worst a complicated mild TBI. Classification statistics for Digit Span scores will almost certainly be worse in more severely affected populations. The current sample was also skewed toward adolescent Caucasian patients who were from well-educated families. Further research will be required to examine if the results generalize to youth from more varied backgrounds. Another limitation is that the MSVT was the only stand-alone SVT administered to all patients. Because the MSVT has been found to have a higher false-negative rate than other lengthier SVTs in adult populations (Green, 2007) and some individuals feign deficits other than verbal memory impairment, noncredible effort cases in the present sample might have been underestimated.

Despite these limitations, this is the first identified study that has examined an embedded SVT in a real-world pediatric sample. The current findings suggest that although scores from Digit Span may be only moderately sensitive, they are likely to have good utility in identifying noncredible performance in older children and adolescents, at least in those who are relatively high-functioning. Indeed, classification statistics produced in this pediatric sample compare favorably with those produced in many real-world adult patients (approximately 50% sensitivity when specificity is set $\geq$90%). Of course, individuals can produce low Digit Span scores for reasons other than noncredible effort including true neurological impairment and developmental learning and attentional difficulties (Harrison, Rosenblum, & Currie, 2010; Heinly, Greve, Bianchini, Love, & Brennan, 2005). Thus, Digit Span scores need to be integrated with other SVT and ability-based test data, as well as historical and contextual information, before judgments can be made about whether or not an individual youth actually displayed noncredible effort during an examination.

Conflict of Interest

None declared.

References

Axelrod, B., Fichtenberg, N., Millis, S., & Wertheimer, J. (2006). Detecting incomplete effort with digit span from the Wechsler Adult Intelligence Scale-third edition. *The Clinical Neuropsychologist*, *20*, 513–523.

Babikian, T., & Boone, K. (2007). Intelligence tests as measures of effort. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 103–127). New York: Guilford Press.

Babikian, T., Boone, K., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist*, *20*, 145–159.

Belanger, H. G., & Vanderploeg, R. D. (2005). The neuropsychological impact of sports-related concussion: A meta-analysis. *Journal of the International Neuropsychological Society*, *11*, 345–357.

Blaskewitz, N., Merten, T., & Kathmann, N. (2008). Performance of children on symptom validity tests: TOMM, MSVT, and FIT. *Archives of Clinical Neuropsychology*, *23*, 379–391.

K. B. Boone (Ed.). (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: The Guilford Press.

Carone, D. A. (2008). Children with moderate/severe brain damage/dysfunction outperform adults with mild-to-no brain damage on the Medical Symptom Validity Test. *Brain Injury*, *22*, 960–971.

Carroll, L. J., Cassidy, J. D., Holm, L., Kraus, J., & Coronado, V. G. (2004). Methodological issues and research recommendations for mild traumatic brain injury: the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *Journal of Rehabilitation Medicine*, *(Suppl. 43)*, 113–125.

Carroll, L. J., Cassidy, J. D., Peloso, P. M., Borg, J., von Holst, H., Holm, L., et al. (2004). Prognosis for mild traumatic brain injury: results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *Journal of Rehabilitation Medicine*, *(Suppl. 43)*, 84–105.

Chafetz, M. D., Abrahams, J. P., & Kohlmaier, J. (2007). Malingering on the Social Security disability consultative exam: a new rating scale. *Archives of Clinical Neuropsychology*, *22*, 1–14.

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

Constantinou, M., & McCaffrey, R. J. (2003). Using the TOMM for evaluating children's effort to perform optimally on neuropsychological measures. *Child Neuropsychology*, *9*, 81–90.

Courtney, J. C., Dinkins, J. P., Allen, L. M., 3rd, & Kuroski, K. (2003). Age related effects in children taking the Computerized Assessment of Response Bias and Word Memory Test. *Child Neuropsychology*, *9*, 109–116.

Donders, J. (2005). Performance on the Test of Memory Malingering in a mixed pediatric sample. *Child Neuropsychology*, *11*, 221–227.

Duncan, S., & Ausborn, D. (2002). The use of reliable digits to detect malingering in a criminal forensic pretrial population. *Assessment*, *9*, 56–61.

Etherton, J., Bianchini, K., Ciota, M., & Greve, K. (2005). Reliable digit span is unaffected by laboratory-induced pain implications for clinical use. *Assessment*, *12*, 101–106.

Faust, D., Hart, K., & Guilmette, T. J. (1988). Pediatric malingering: the capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *56*, 578–582.

Faust, D., Hart, K. J., Guilmette, T. J., & Arkes, H. R. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice*, *19*, 508–515.

Flaro, L., & Boone, K. (2009). Using objective effort measures to detect noncredible cognitive test performance in children and adolescents. In J. E. Morgan, & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 369–376). New York: Psychology Press.

Green, P. (2004). *Manual for the Medical Symptom Validity Test*. Edmonton: Green's Publishing Inc.

Green, P. (2007). Spoiled for choice: Making comparisons between forced-choice effort tests. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment*. New York: The Guilford Press.

Green, P., & Flaro, L. (2003). Word Memory Test performance in children. *Child Neuropsychology*, *9*, 189–207.

Greiffenstein, M., Baker, W., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, *6*, 218–224.

Greiffenstein, M., Gola, T., & Baker, W. (1995). MMPI-2 Validity scales versus domain specific measures in detection of factitious traumatic brain injury. *The Clinical Neuropsychologist*, *9*, 230–240.

Harrison, A. G., Rosenblum, Y., & Currie, S. (2010). Examining unusual digit span performance in a population of postsecondary students assessed for academic difficulties. *Assessment*, *17*, 283–93.

Heinly, M. T., Greve, K. W., Bianchini, K. J., Love, J. M., & Brennan, A. (2005). WAIS digit span-based indicators of malingered neurocognitive dysfunction: classification accuracy in traumatic brain injury. *Assessment*, *12*, 429–44.

Henry, G. (2005). Childhood malingering: faking neuropsychological impairment in an 8-year-old. In R. L. Heilbronner (Ed.), *Forensic neuropsychology casebook* (pp. 205–217). New York: Guilford Press.

Kirk, J. W., Harris, B., Hutaff-Lee, C. F., Koelemay, S. W., Dinkins, J. P., & Kirkwood, M. W. (2011). Performance on the Test of Memory Malingering (TOMM) among a large clinic-referred pediatric sample. *Child Neuropsychology*. doi: 10.1080/09297049.2010.533166.

Kirkwood, M. W. (in press). Overview of tests and techniques to detect negative response bias in children. In E. Sherman, & B. Brooks (Eds.), *Pediatric forensic neuropsychology*. New York: Oxford University Press.

Kirkwood, M. W., & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the Medical Symptom Validity Test. *The Clinical Neuropsychologist*, *24*, 860–872.

Kirkwood, M. W., Kirk, J. W., Blaha, R. Z., & Wilson, P. E. (2010). Noncredible effort during pediatric neuropsychological exam: A case series and literature review. *Child Neuropsychology*, *16*, 604–618.

Kirkwood, M. W., Yeates, K. O., Taylor, H. G., Randolph, C., McCrea, M., & Anderson, V. A. (2008). Management of pediatric mild traumatic brain injury: A neuropsychological review from injury through recovery. *The Clinical Neuropsychologist*, *22*, 769–800.

Larrabee, G. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, *17*, 410–425.

G. J. Larrabee (Ed.). (2007). *Assessment of malingered neuropsychological deficits*. New York: Oxford University Press.

Lu, P. H., & Boone, K. B. (2002). Suspect cognitive symptoms in a 9-year-old child: malingering by proxy? *Clinical Neuropsychology*, *16*, 90–96.

MacAllister, W. S., Nakhutina, L., Bender, H. A., Karantzoulis, S., & Carlson, C. (2009). Assessing effort during neuropsychological evaluation with the TOMM in children and adolescents with epilepsy. *Child Neuropsychology*, *15*, 521–531.

Maillard-Wermelinger, A., Yeates, K. O., Taylor, H. G., Rusin, J., Bangert, B., Dietrich, A., et al. (2009). Mild traumatic brain injury and executive functions in school-aged children. *Developmental Neurorehabilitation*, *12*, 330–341.

McCaffrey, R., & Lynch, J. (2009). Malingering following documented brain injury: Neuropsychological evaluation of children in a forensic setting. In J. E. Morgan J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 377–385). New York: Psychology Press.

Suhr, J., & Barrash, J. (2007). Performance on standard attention, memory, and psychomotor speed tasks as indicators of malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 131–170). New York: Oxford University Press.

Tombaugh, T. N. (1996). *Test of memory malingering.* Toronto, Ontario, Canada: Multi-Health Systems.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th ed.). San Antonio, TX: The Psychological Corporation.

Ylioja, S. G., Baird, A. D., & Podell, K. (2009). Developing a spatial analogue of the reliable digit span. *Archives of Clinical Neuropsychology*, *24*, 729–739.