Beyond Full Scale IQ: A New WAIS-III Indicator of Mental Retardation

Christina M. Almstrom
Oklahoma State University

Nick M. Wisdom
Oklahoma State University

Jennifer L. Callahan
University of North Texas

Abstract

It is paradoxical that individuals with mental retardation are disproportionately represented in the criminal justice system, suggesting a greater proportion of false positives, while at the same time there is concern regarding a potential motivation for individuals without mental retardation to simulate such a condition for the purpose of avoiding the death penalty, a concern for false negatives. To determine if the likelihood of accurate classification of impairment could be enhanced, we examined the pattern of scaled scores in the WAIS-III standardization dataset. Specifically, we compared the scores obtained by individuals obtaining a Full Scale IQ within the Borderline range with those in the Mild or Moderate range of impairment. Difference scores among the indexes were used to create a composite indicator making it possible to more than double the likelihood of correctly detecting impairment. Given the potential motivation to simulate impairment, a unique feature of the composite indicator described here is the inclusion of a regression equation to estimate missing values.

The American Association on Intellectual and Developmental Disabilities (AAIDD; formerly American Association on Mental Retardation; AAMR) reports that individuals with mental retardation are disproportionately represented in the United States criminal justice system, with estimates ranging from 4% to 10% in comparison to estimates of mental retardation ranging between 1.5% and 2.5% in the general population (AAMR, 2002). Such individuals pose unique considerations within the criminal justice system. For example, an individual with mental retardation may have a heightened desire to appease authority figures, have poor comprehension of causality, be easily led, and/or display an exaggerated willingness to talk (Brodsky & Bennett, 2005). Illustrating these vulnerabilities, a study reported by the American Civil Liberties Union (ACLU, 2004) examining a defendant's ability to comprehend their Miranda rights found that 73% of individuals with intellectual impairment did not truly understand that their statements could be used against them in legal proceedings. It is not surprising then that in a study of 125 cases where individuals were later exonerated, Drizin and Leo (2004) found that 22% of the falsely convicted individuals had mental retardation. Such alarming figures merit researchers' attention to the how the needs of individuals that are intellectually impaired may be better identified and served in the criminal justice system.

The implications of mental retardation are particularly salient in potential death penalty cases. Between 1976 and 1995 there were 18 people with documented mental disabilities, including mental retardation, executed for capital crimes in the United States. Of 30 executions in 1995, 20% of the individuals put to death had mental retardation, despite the U.S. Supreme Court having previously expressed the opinion that mental retardation poses a mitigating circumstance (e.g., *Penry v. Lynaugh*; 492, U.S. 302, 1989) that must be considered. Further, in 2002, the U.S. Supreme Court opined more strongly, in *Atkins v. Virginia* (536, U.S. 304, 2002), that execution of an individual with mental retardation violates the Eighth Amendment. As a result, 26 of the 40 death penalty states have enacted legislation protecting individuals with mental retardation from execution.

The definitions of mental retardation described in state and federal statutes typically draw upon those provided by professional organizations such as the AAIDD and/or the American Psychiatric Association (APA), but important differences between professional guidelines and statutes exist. Generally, legislation describes mental retardation as characterized by persisting deficits in both intellectual and adaptive functioning (e.g., social, conceptual, and practical skills in everyday life) with onset beginning prior adulthood (see, for example, Oklahoma Statutes §10-1408, 2005). Intellectual deficit is typically defined as obtaining a full scale IQ that is two or more standard deviations below the mean on a standardized intelligence test, such as the Wechsler Adult Intelligence Scale-III (WAIS-III), and may be specifically set at 70 or below (again, see for example, Oklahoma Statutes §10-1408, 2005). Unfortunately, such legislative definitions often deviate from AAMR and APA guidelines, which provide a more ambiguous cut score for considering full scale IQ to be consistent with mental retardation.

Correspondence regarding this manuscript may be directed to Jennifer L. Callahan, PhD ABPP, Department of Psychology, PO Box 311280, Denton TX 76203-1280, USA. jennifercallahan@unt.edu

In particular, the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) specifies that an individual's full scale IQ may be as high as 75 and still be indicative of mental retardation. Literature provides no support of legislative statutes that set an absolute cut score of 70 and the authors view this as a truly fatal flaw in legislation for those with intellectual disabilities.

Since the *Atkins v. Virginia* ruling, it has been argued successfully in court proceedings that a potential motivation exists for individuals without mental retardation to simulate such a condition for the purpose of avoiding the death penalty. Historically, the feigning of mental retardation was considered unlikely because of the associated negative societal connotations (AAMR, 2002). However, if a motivation exists, Hall and Pritchard (1997) suggest that faking bad is made particularly easy on assessments of cognitive ability because the individual is told what is required in order to do well thereby making it obvious what is required in order to do poorly. Although administration of a standardized measure of intelligence is required as part of the diagnostic process for establishing mental retardation (Bonnie, 2004), such instruments contain no specific validity scale(s) for use in detecting simulated mental retardation.

The most commonly administered intelligence scale for adults is the Wechsler Adult Intelligence Scale, which is presently in its third edition (Wolfe-Christensen, & Callahan, in press). Aside from a measure of intelligence, measurement of adaptive functioning must be conducted to document mental retardation. For the purpose of considering whether impairments on these measures might be feigned, an individual examiner may choose to administer broad symptom validity measures (e.g., Structured Interview of Reported Symptoms, SIRS,) or memory malingering assessments (e.g., "TOMM"; Test of Memory Malingering). However, while these instruments serve as psychometrically sound measures of negative impression management within discrete areas, they are not necessarily specific indicators of simulated impairment on measures of mental retardation (Greiffenstein, Gola, & Baker, 1995; Meyers & Volbrecht, 2003). Nevertheless, such instruments are routinely used by examiners conducting evaluations of mental retardation in death penalty cases (e.g., Oklahoma Forensic Center evaluations).

In order to determine whether or not individuals can effectively, and willfully, perform within the impaired range on a measure of intelligence, Johnstone and Cooke (2003) instructed a group of incarcerated youthful offenders and a group of postgraduate students to simulate intellectual impairment falling below a full scale IQ of 69 on the Wechsler Adult Intelligence Scale – Revised (WAIS-R). Although both groups were able to generate full scale IQ scores below the criterion, analysis of their responses revealed easily detected patterns of responding and types of errors. As a result, Johnstone and Cooke suggested that greater attention be given to subtest inconsistencies and floor effects in detecting mental retardation simulations. While the Johnstone and Cooke results have implications for the single case review, they did not provide any exploration of aggregate patterns. Further, they used the WAIS-R while the Daubert standard (1993) necessitates that current assessment of adults use a more recent instrument, such as the WAIS-III.

A large number of studies have reported success in the detection of malingering related to traumatic head injury, alcohol abuse, and polysubstance abuse using the subtests of the WAIS-III. For example, Vocabulary minus Digit Span and the Rarely Missed Index (Miller, Ryan, Caruthers, & Cluff, 2004), the Reliable Digit Span (Mathias, Greve, Bianchini, Houston, & Crouch, 2002), the Discriminate Function Score (Greve, Bianchini, Mathias, Houston, & Crouch, 2003), unusual digit span patterns (Iverson & Tulsky, 2003; Strauss, Slick, Levy-Bencheton, Hunter, MacDonald, & Hultsch, 2002), and profile analysis (Mittenberg, Aguila-Puentes, Patton, Canyock, & Heilbronner, 2002) have all demonstrated some benefit in accurately identifying simulated impairment.

This research is largely reflective of detecting malingering of traumatic head injury and substance induced brain dysfunction, but it suggests that similar approaches might be useful in the identification of feigned mental retardation. In an effort to better assess the clinical usefulness of the WAIS-III, normative data was obtained from The Psychological Corporation (the test developer and publisher) for all individuals within the standardization dataset with an obtained full scale IQ below 80.

Similar to the approach of Keith Hawkins (1997) in examining standardization data on the WAIS-III to aid clinicians in identifying individuals with various forms of acquired brain dysfunction, this article presents graphical representations of inter-subscale patterns at different levels of impairment as measured by full scale IQ. This approach was selected because it is uniquely consistent with the recently announced planned renaming, effective 2009, of *mental retardation* to *intellectual disability* (Schalock, Luckasson, & Shogren, 2007). More specifically, subscale patterns for individuals known to have mental retardation in the mild to moderate range (IQ of 45 to 69) were compared to individuals in the qualitatively borderline range (IQ of 70 to 80).

METHOD

Participants

The data utilized for this investigation were acquired from the publisher of the WAIS-III and included the normative data generated during the standardization process. Data were obtained for all individuals within the standardization dataset that obtained a full scale IQ below 80 and consisted of the scaled scores for each subtest as well as the standard scores for each of the four indexes (Verbal Comprehension, VCI; Working Memory, WMI; Perceptual Organization, POI; and Processing Speed, PSI) and the three IQ scores (Full Scale IQ, FSIQ; Verbal IQ, VIQ; and Performance IQ, PIQ).

There were 183 individuals with a FSIQ falling in the Borderline range (FSIQ 70-80) and 64 individuals with a FSIQ falling in the Mild to Moderate impairment range (FSIQ 45-69). The average age of individuals with a FSIQ in the Borderline range was 52.10 with a median stratified education level of 9-11 years. The average age of individuals with a FSIQ in the Mild/Moderate range was 47.03 with a median stratified education level of eight years or less. Males comprised 63% of the participants in the Borderline range and 55% of the participants in the Mild/Moderate range with 37% and 45% females, respectively.

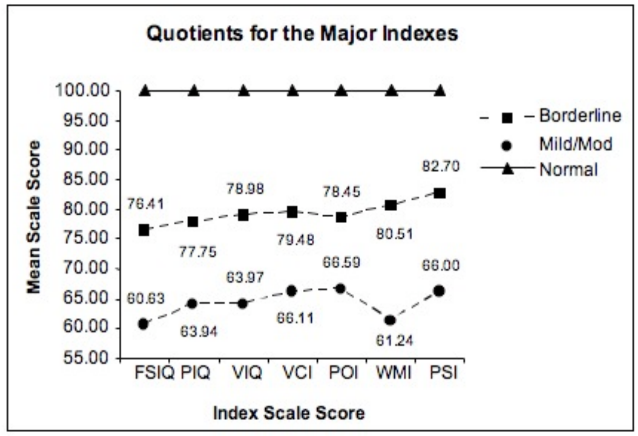
Procedure

To compare patterns of performance, scaled scores for the three WAIS-III intelligence quotients and the four major indexes of 64 subjects were randomly selected from the 183 subjects in the Borderline range to achieve an equal sample size to the Mild/Moderate impairment group. The Working Memory Index contained missing observations in each of these groups due to many participants being unable to complete the Letter-Number Sequencing subtest (45.36% of those in the Borderline range and 68.75% in the Mild/Moderate Mental Retardation range). Instead of simply substituting group means for these data, a regression equation was computed using the FSIQ and WMI scores where there were no missing observations. This equation, WMI = 1.165(FSIQ) – 9.552, was then used to estimate the missing values in both groups.

RESULTS AND DISCUSSION

The means for each of the WAIS-III quotients and indexes are depicted in Figure 1. Separate within-subjects ANOVAs revealed significant differences for each of the comparison groups: Borderline: $F_{(6,378)}$ = 8.197, $p < .001$, $\eta_p^2 = .115$, and Mild/Moderate: $F_{(6,378)}$ = 22.93, $p < .001$, $\eta_p^2 = .259$. The results of post hoc comparisons within the samples revealed several significant differences of various effect sizes. The magnitude of the difference between the effect sizes within the samples was used as an indication of how well the comparison would be expected to discriminate Borderline from Mild/Moderate subjects. The comparisons that yielded the largest effect size differences were VCI-VIQ, POI-PIQ, POI-WMI, and PSI-WMI.

***Figure 1.* Individual mean scale scores for WAIS-III Quotients and Major Indexes.**



As can be seen from Figure 2, η_p^2 is relatively large in all four comparisons for the Mild/Moderate group and near zero (what would be expected in a sample of subjects in the normal range) in three of the four comparisons for the Borderline group. With this information, difference scores on each of these predictor comparisons were computed and analyzed in a 2 x 4 between/within ANOVA. The main effect for Group was significant, $F_{(1,126)} = 20.81$, $p < .001$, $\eta_p^2 = .142$, as was the main effect for Predictor Comparison $F_{(3,\ 378)} = 3.43$, $p = .017$, $\eta_p^2 = .026$. The interaction of Group x Predictor Comparison was also significant, $F_{(3,378)} = 6.57$, $p < .001$, $\eta_p^2 = .05$. These interaction means are shown in Figure 3. A significant interaction is indicative of different performance patterns for the Borderline and Mild/Moderate subjects which, potentially, can be used to distinguish between them.

***Figure 2.* Comparison of the relative effect sizes (η_p^2) within the Borderline and the Mild/Moderate Impairment samples.**

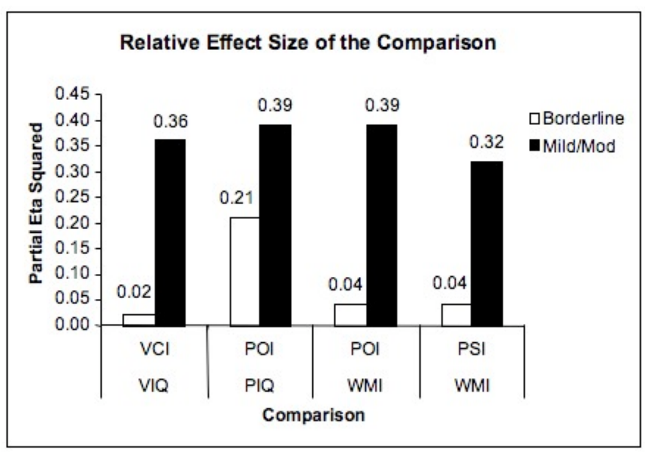
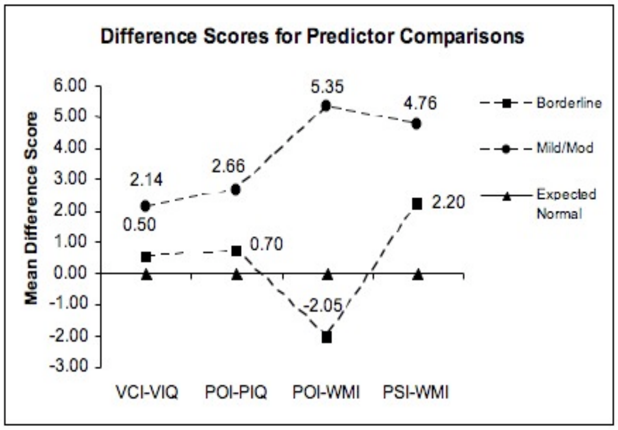
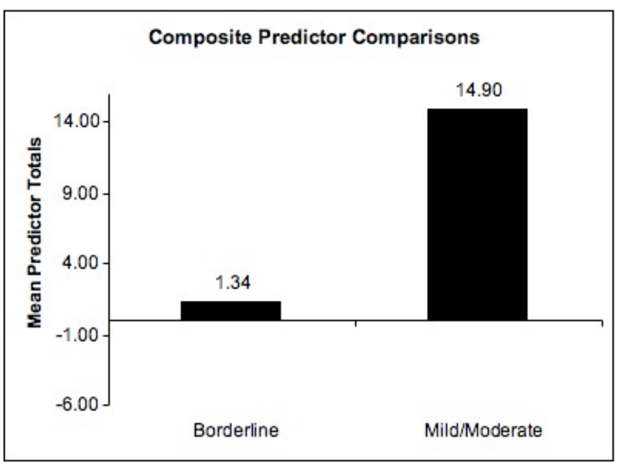


Figure 3. **Predictor Comparisons from Major Index Difference** Scores.



A simple effects analysis revealed significant comparisons between Groups for the VCI-VIQ comparison, $F_{(1,126)} = 7.56$, $p = .007$, $\eta_p^2 = .056$; the POI-PIQ comparison, $F_{(1,126)} = 9.41$, $p = .003$, $\eta_p^2 = .070$; and the POI-WMI comparison $F_{(1,126)} = 25.89$, $p < .001$, $\eta_p^2 = .166$; however, the PSI-WMI comparison was not significant, $F_{(1,126)} = 2.46$, $p = .120$, $\eta_p^2 = .017$. To provide even better discrimination between performance patterns between the Borderline and Mild/Moderate groups, composite scores for these predictor comparisons were computed. These composite scores are shown in Figure 4.

Figure 4. **Mean composite scores from the four predictor comparisons VCI-VIQ, POI-PIQ, POI-WMI, and PSI-WMI.**



The composite mean for the Borderline subjects, $M = 1.34$, was close to what would be expected for a group of normal subjects, $E(M) = 0$, while the mean composite for the Mild/Moderate subjects was substantially larger, $M = 14.90$. The difference between these groups was significant, $F_{(1,126)} = 20.89$, $p < .001$, $\eta_p^2 = .142$.

Table 1 gives the results of a Bayesian analysis for each of the composite indicators ranging from above 2.0 to above 12.0. The sensitivity ratios varied from .68 to .85. The specificity ratios varied from .55 to .75. The median sensitivity ratio was .77 and the median specificity ratio was .65. In terms of a balance between optimal sensitivity and specificity, a cut-off of 7.0 is appropriate. Using the 7.0 composite indicator cut score for a decision regarding the profile's consistency with that of mental retardation produced good sensitivity (.78) and specificity (.64) with 78.13% True Positives, 64.06% True Negatives, 21.88% False Negatives, and 35.94% False Positives. Because these values are population-based and therefore not as useful at the level of individual assessment, positive and negative likelihood ratios were also computed. The positive likelihood ratio with a composite indicator of 7.0 is 2.17, and the negative likelihood ratio is 0.34. In other words, given a total composite indicator above 7.0, the likelihood that the individual's cognitive abilities are truly consistent with mental retardation is increased more than double that of other individuals who obtain a composite indicator below the 7.0 cut score.

CONCLUSIONS

Since the *Atkins v. Virginia* ruling, Oklahoma courts have heard repeated arguments that a potential motivation exists for individuals without mental retardation to simulate such a condition for the purpose of avoiding the death penalty. At this time there exist no validity scale(s) on standard measures of intelligence to detect malingering, which is quite unfortunate in light of recent research documenting that simulated impairment can be identified (e.g., Johnstone & Cooke, 2003). The authors' observation of legal proceedings in Oklahoma is that the tautology argued by prosecutors before the court is that a defendant *is* malingering if they score within the impaired range on intellectual tests. Indeed, such arguments are made even if the defendant scores within normal limits on administered malingering tests (e.g., *Miller v. State*; O-2005-1072, Okl. Cr., April 5, 2006). We strongly believe a shift in thinking needs to occur with an emphasis on posteriori knowledge. The importance of this issue is underscored when one considers that Oklahoma is ranked third in total number of inmates executed, an alarming achievement given the relatively small population (Criminal Justice Project, 2005). For example, California, another death penalty state with a vastly greater population, ranks 17th in total number of executions. Future research is strongly encouraged to assist in creating a shift to emphasize posteriori knowledge in this high-stakes area.

The determination of whether a protocol is malingered vs. accurate is made especially difficult when the known history and observed adaptive functioning suggest that the individual's intellectual abilities could fall somewhere within either the borderline or the impaired range. Perhaps an individual with borderline

Table 1. Bayesian Analysis Model Applied to Composite Scores

Composite Indicator Above	Sensitivity	Specificity	True Positive %	False Negative %	True Negative %	False Positive %	Likelihood + Ratio	Likelihood - Ratio	Predictive + Value	Predictive - Value
3	0.83	0.55	83	17	55	45	1.82	0.31	0.65	0.76
4	0.81	0.55	81	19	55	45	1.79	0.34	0.64	0.75
5	0.80	0.59	80	20	59	41	1.96	0.34	0.66	0.75
6	0.78	0.61	78	22	61	39	2.00	0.36	0.67	0.74
7	0.78	0.64	78	22	64	36	2.17	0.34	0.68	0.75
8	0.75	0.66	75	25	66	34	2.18	0.38	0.69	0.72
9	0.73	0.67	73	27	67	33	2.24	0.40	0.69	0.72
10	0.72	0.69	72	28	69	31	2.30	0.41	0.70	0.71
11	0.70	0.72	70	30	72	28	2.50	0.41	0.71	0.71
12	0.68	0.75	69	31	75	25	2.75	0.42	0.73	0.71

intellectual abilities that is facing the possibility of being sentenced to death has much to gain from under-performing on intelligence testing resulting in a FSIQ falling in the impaired range. Recognizing this, legal counsel may attempt to argue that a truly impaired individual actually possesses borderline (or better) intellectual abilities and is only feigning more significant deficits. When known history and demonstrated adaptive functioning make such a distinction ambiguous, the potential for a grave error exists in which a truly impaired individual may be mistakenly labeled as malingering.

This investigation sought to determine whether data obtained from borderline and impaired participants in the WAIS-III standardization project might reveal additional indicators, beyond the FSIQ, to facilitate making the distinction between borderline vs. impaired intellectual abilities. Examination of the data revealed several predictor comparisons that could be summed into a composite indicator. After establishing a cut score for the composite indicator, it was found that the usage of the indicator could result in a *two-fold increase in the likelihood of accurately detecting an individual with impaired intellectual abilities*.

While it is still possible that an individual might malinger impairment successfully, the involvement of several predictor comparisons to compute the composite indicator score makes it seem unlikely that an individual whose history and adaptive functioning suggest, at best, borderline abilities could be sophisticated enough to produce the necessary performance pattern for classification as impaired. A unique feature of the composite indicator described here is the inclusion of a regression equation to compute the WMI score if the LNS subtest is not completed.

Given the potential motivation to feign mental retardation in the criminal justice system and the fact that faking bad is facilitated on assessments of cognitive ability because the individual is told what is required in order to do well, Johnstone and Cooke (2003) have suggested that particular attention should be given to inconsistencies in patterns of responding and floor effects. The composite indicator developed here, using difference scores between indexes and a regression estimate when failure to complete an index is encountered, achieves this goal and could potentially prove valuable to clinicians called upon to make judgments regarding the level of impairment for an individual within the criminal justice system. While the composite indicator appears to have strong potential, further research using a procedure similar to that of Johnstone and Cooke (2003) to field-test the composite indicator's effectiveness in detecting individuals who are instructed to purposefully perform within the impaired range is needed to validate its predictive utility.

References

Association on Mental Retardation (2002). Mental retardation definition, classification, and systems of supports. *American Association on Mental Retardation* (10th ed.). Washington, DC.

American Civil Liberties Union (2004). *ACLU Capital Punishment Project.* Retrieved September 25, 2006. http://www.aclu.org/capital/mentalretardation/10391pub20030904.html

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

Atkins v. Virginia, 536, U.S. 304 (2002).

Bonnie, R. J. (2004). The American Psychiatric Association's resource document on mental retardation and capital sentencing: Implementing *Atkins v. Virginia*. *The Journal of the American Academy of Psychiatry and the Law, 32,* 304-308.

Brodsky, S. L., & Bennett, A. D. (2005). Psychological assessments of confessions and suggestibility in

mentally retarded suspects. *Journal of Psychiatry and Law, 33,* 359-366.
Criminal Justice Project (Summer, 2005). *Death Row U.S.A.* Washington, DC: NAACP Legal Defense and Educational Fund, Inc.
Daubert v. Merrell Dow Pharmaceuticals, 509, U.S. 579 (1993).
Drizin, S. A., & Leo, R. A. (2004). The problem of false confessions in the post-DNA world. *North Carolina Law Review, 82,* 891–1007.
Greiffenstein, M. F., Gola, T., & Baker, W. J. (1995). MMPI-2 validity scales versus domain specific measures in the detection of factitious traumatic brain injury. *The Clinical Neuropsychologist, 9,* 230–240.
Greve, K. W., Bianchini, K. J., Mathias, C. W., Houston, R. J., & Crouch J. A. (2003). Detecting malingered performance on the Wechsler Adult Intelligence Scale. Validation of Mittenberg's approach in traumatic brain injury. *Archives of Clinical Neuropsychology, 18,* 245-260.
Hall, H. H., & Pritchard, D. A. (1997). *Detecting malingering and deception: Forensic distortion analysis (FDA)*. Raton, FL: St Lucie Press.
Hawkins, K. A. (1998). Indicators of brain dysfunction derived from graphic representations of the WAIS–III/WMS–III Technical Manual clinical samples data: A preliminary approach to clinical utility. *The Clinical Neuropsychologist, 12,* 535–551.
Iverson, G. L., & Tulsky, D. S. (2003). Detecting malingering on the WAIS-III: Unusual digit span performance patterns in the normal population and in clinical groups. *Archives of Clinical Neuropsychology*, *18,* 1-9.
Johnstone, L., & Cooke, D. J. (2003). Feigned intellectual deficits on the Wechsler Adult Intelligence Scale-Revised. *British Journal of Clinical Psychology, 42,* 303-318.
Mathias, C. W., Greve, K. W., Bianchini, K. J., Houston, R. J., & Crouch, J. A. (2002). Detecting malingered neurocognitive dysfunction using the Reliable Digit Span in traumatic brain injury. *Assessment*, *9,* 301-308.
Meyers, E., & Volbrecht, M. E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology, 18,* 261–276.
Miller v. State, O-2005-1072, Okl. Cr.(April 5, 2006).
Miller, L. J., Ryan, J. J., Caruthers, C. A., & Cluff, R. B. (2004). Brief screening indexes for malingering: A confirmation of vocabulary minus digit span from the WAIS-III and the rarely missed index from the WMS-III. *Clinical Neuropsychologist*, *18*, 327-333.
Mittenberg, W., Aguila-Puentes, G., Patton, C., Canyock, E. M., & Heilbronner, R. L. (2002). Neuropsychological profiling of symptom exaggeration and malingering. *Journal of Forensic Neuropsychology, 3*, 227-240.
Oklahoma Statutes §10-1408 (2005).
Penry v. Lynaugh, 492, U.S. 302 (1989).
Schalock, R. L., Luckasson, R. A., & Shogren, K. A. (2007). The renaming of mental retardation: Understanding the change to the term intellectual disability. *Intellectual and Developmental Disabilities, 45,* 116-124.
Strauss, E., Slick, D. J., Levy-Bencheton, J., Hunter, M., MacDonald, S. W. S., & Hultsch, D. F. (2002). Intraindividual variability as an indicator of malingering in head injury. *Archives of Clinical Neuropsychology*, *17,* 423-444.
Wolfe-Christensen, C., & Callahan, J. L. (in press). The current state of standardization adherence: A reflection of competency in psychological assessment. *Training & Education in Professional Psychology.*