Performance Validity Testing in Justice-Involved Adults with Fetal Alcohol Spectrum Disorder

Mullally, K., McLachlan, K., MacKillop, E., & Pei, J.

Suggested Citation: Mullally, K., McLachlan, K., MacKillop, E., & Pei, J. (2020). Performance validity testing in justice-involved adults with fetal alcohol spectrum disorder. *Journal of the International Neuropsychological Society, 26*(7), 701 – 713. https://doi.org/10.1017/S1355617720000132

Performance Validity Testing in Justice-Involved Adults with Fetal Alcohol Spectrum Disorder

Authors:

Katelyn Mullally, M.A., Department of Psychology, University of Guelph

*Kaitlyn McLachlan, Ph.D., C. Psych, Department of Psychology, University of Guelph

Emily MacKillop, Ph.D., C. Psych., ABPP-CN, Clinical Neuropsychology Service, St. Joseph's Healthcare Hamilton; Department of Psychiatry and Behavioural Neurosciences, McMaster University

Jacqueline Pei, Ph.D., R. Psych., Faculty of Education, University of Alberta

*Corresponding Author: Kaitlyn McLachlan. Assistant Professor, Department of Psychology, University of Guelph. MacKinnon 3011, 50 Stone Road East, Guelph, Ontario, Canada, N1G 2W1. kmclac02@uoguelph.ca, (519) 824-4120, ext. 56447.

Acknowledgements: Funding for the primary study was provided by the Department of Justice, Territorial Government in the study jurisdiction. There are no conflicts of interest to report for the authors.

Conflicts of Interest: The authors have no financial or other conflicts to declare.

Abstract

Objectives: A number of commonly used performance validity tests (PVTs) may be prone to high failure rates when used for individuals with severe neurocognitive deficits. This study investigated the validity of 10 PVT scores in justice-involved adults with fetal alcohol spectrum disorder (FASD), a neurodevelopmental disability stemming from prenatal alcohol exposure and linked with severe neurocognitive deficits.

Method: The sample comprised 80 justice-involved adults (ages 19 – 40) including 25 with confirmed or possible FASD and 55 where FASD was ruled out. Ten PVT scores were calculated, derived from Word Memory Test, Genuine Memory Impairment Profile, Advanced Clinical Solutions (Word Choice), the Wechsler Adult Intelligence Scale - Fourth Edition (Reliable Digit Span and age corrected scaled scores (ACSS) from Digit Span, Coding, Symbol Search, Coding - Symbol Search, Vocabulary - Digit Span), and the Wechsler Memory Scale - Fourth Edition (Logical Memory II Recognition).

Results: Participants with diagnosed/possible FASD were more likely to fail *any* single PVT, and failed a greater number of PVTs overall, compared to those without FASD. They were also more likely to fail based on Word Memory Test, Digit Span ACSS, Coding ACSS, Symbol Search ACSS and Logical Memory II Recognition, compared to controls (35% - 76%). Across both groups, substantially more participants with IQ <70 failed two or more PVTs (90%), compared to those with an IQ ≥70 (44%).

Conclusions: Results highlight the need for additional research examining the use of PVTs in justice-involved populations with FASD.

Keywords: Fetal alcohol spectrum disorder, performance validity testing, forensic assessment, neuropsychological assessment, non-credible responding, reliability, validity, prenatal alcohol exposure.

Fetal alcohol spectrum disorder (FASD) comprises a range of impairments stemming from prenatal alcohol exposure (PAE), including neurocognitive deficits, problems regulating affect and behavior, and in some cases, characteristic dysmorphic facial features and growth restriction (Cook et al., 2016; Hoyme et al., 2016; Mattson, Bernes, & Doyle, 2019). Individuals with FASD experience high rates of additional adversities and comorbid conditions (e.g., McLachlan et al., 2016; Pei, Denys, Hughes, & Rasmussen, 2011; Popova et al., 2016; Streissguth et al., 2004). North American FASD prevalence estimates range from 2 to 5%, with higher rates in forensic and criminal justice contexts, where prevalence estimates range from 10% to 36% (May et al., 2014, 2018; Popova et al., 2018; Popova, Lange, Shield, Burd, & Rehm, 2019). A number of factors have been proposed to explain the increased risk of criminal justice contact among individuals with FASD, including the nature of neurocognitive deficits, high rates of early adversities and challenges related to diagnosis (see Currie, Hoy, Legge, Temple, & Tahir, 2016; Flannigan, Pei, Stewart, & Johnson, 2018)

Cognitive impairment is a defining characteristic of FASD, though notable variability of relative deficits and strengths is often seen in this population (Ali, Kerns, Mulligan, Olson, & Astley, 2018; Mattson et al., 2019; Mattson, Crocker, & Nguyen, 2011). Specific cognitive impairments may include deficits in attention, executive functioning, language, memory, learning, communication, and intellectual functioning (see reviews by Kodituwakku & Kodituwakku, 2014; Mattson et al., 2019). FASD is often underrecognized owing to many factors, including the heterogeneous nature of the disability, sometimes preserved overall intellectual functioning, and compensatory skills that mask underlying neurocognitive deficits (Ali et al., 2018; Astley, 2010; Kodituwakku & Kodituwakku, 2014; Mattson et al., 2011).

There are currently several approaches to the diagnosis and classification of individuals with FASD in North America. Across guidelines, best practices include comprehensive assessment undertaken by a trained interdisciplinary team (e.g., Astley, 2004; Coles et al., 2016; Cook et al., 2016; Hoyme et al., 2016). However, given elevated rates of FASD in special populations, it is likely that clinicians working in both general health and forensic/correctional settings will encounter individuals with PAE and/or FASD in their usual practice. This is also likely to occur with increased frequency given the addition of FASD in the *Diagnostic and Statistical Manual of Mental Disorders*[1] (DSM-5; American Psychiatric Association, 2013, p. 86) and the *International Classification of Diseases*[2] (ICD-10; World Health Organization, 2007), and as knowledge about the disability continues to increase (e.g., Mukherjee, Hollins, & Turk, 2006)

Performance Validity Testing

Performance validity tests (PVTs) form an important aspect of neuropsychological and cognitive assessment. This is particularly true in medicolegal, correctional, and forensic contexts, where there may be greater incentive to mislead an examiner in order to increase potential financial compensation or decrease legal consequences. Correspondingly, higher base rates of non-credible responding are often observed in these populations (Ardolf, Denney, & Houston, 2007; Bush, Heilbronner, & Ruff, 2014; Bush et al., 2005; Larrabee, 2003). PVTs should be sensitive to effort, while remaining robust to the effects of cognitive impairment in order to accurately differentiate individuals with true deficits from those with non-credible responding

[1] FASD is included in the DSM-V as a condition for further study and an exemplar of "Other Specified Neurodevelopmental Disorder" (American Psychiatric Association, 2013)

[2] Fetal Alcohol Syndrome (FAS) is included in the ICD-11 as a diagnostic term under the heading Congenital Malformation Syndromes Due to Known Exogenous Causes, Not Elsewhere Classified (World Health Organization, 2007).

(Bain & Soble, 2017; Dwyer, 1996). Best practice guidelines underscore the importance of selecting PVTs with established psychometric properties for both examinee and setting (e.g., clinical, medicolegal), and suggest using multiple measures over the course of an evaluation (Bush et al., 2014, 2005). These considerations are particularly critical in legal contexts, where any psychological measures introduced during court proceedings must meet standards for evidentiary admissibility (e.g., *Daubert v. Merrell Dow Pharmaceuticals*, 1993; *R. v. Peters*, 2011; *R v. Mohan*, 1994).

Many PVTs have demonstrated sound psychometric properties in both clinical and forensic settings. However, their classification accuracy tends to be lower in populations marked by severe neurocognitive deficits, including individuals with intellectual disability (ID), dementia, Alzheimer's disease, and traumatic brain injury (TBI) (Bain & Soble, 2017; Bain et al., 2019; Bigler, 2012; Dean, Victor, Boone, & Arnold, 2008; Glassmire, Wood, Ta, Kinney, & Nitch, 2019; Merten, Bossink, & Schmand, 2007; Zenisek, Millis, Banks, & Miller, 2016). For instance, participants with confirmed neurocognitive deficits often present with lower overall scores and high failure rates using established cutoff scores on PVTs, including Reliable Digit Span (RDS), Word Memory Test (WMT), Coding age corrected scaled score (CD ACSS), Symbol Search ACSS (SS ACSS), and Coding-Symbol Search ACSS (CD - SS ACSS) (Dean et al., 2008; Erdodi et al., 2017; Merten et al., 2007; Zenisek et al., 2016). They also tend to fail more PVTs when multiple measures are administered, compared to individuals without severe neurocognitive deficits (Dean et al., 2008; Merten et al., 2007; Zenisek et al., 2016). Consistent with this, commonly used PVTs, such as RDS and Logical Memory II Recognition (LM-II-R), have often demonstrated inadequate specificity (defined as less than 90%; Boone, 2013) and/or sensitivity (above 40% but closer to 70%; Boone, 2013) for identifying non-credible responding

in individuals with neurocognitive deficits (Bain et al., 2019; Dean et al., 2008; Schroeder, Twumasi-Ankrah, Baade, & Marshall, 2012).

There is some evidence to suggest that PVT performance is associated with overall intellectual ability, where samples with low IQ perform worse on commonly used PVTs and produce more failing scores, compared to those with preserved overall intellectual functioning (Dean et al., 2008; Glassmire et al., 2019; Merten et al., 2007; Zenisek et al., 2016). However, low IQ on its own may be insufficient to explain failure on PVTs, suggesting that other mechanisms may contribute to poor performance in populations with high failure rates, such as a cumulative or interactive combination of cognitive deficits (Flaro, Green, & Robertson, 2007; Green & Flaro, 2015; Love, Glassmire, Zanolini, & Wolf, 2014; Shandera et al., 2010; Simon, 2007). False positives have adverse clinical and practical implications, as they may influence diagnostic accuracy, and consequently, prevent access to appropriate treatment and services. As a result, some guidelines suggest that certain clinical groups, including individuals with ID and dementia, be exempt from PVTs (Boone, 2013).

Performance Validity Testing and FASD

Given the severe neurocognitive deficits linked with FASD, there may be an increased risk of improperly identifying individuals with the disability as non-credible responders when using PVTs. While experts caution against the use of PVTs for individuals with neurocognitive deficits, limited knowledge about FASD among practitioners, coupled with high rates of 'missed diagnosis' and the relative invisibility of the disability, suggest that clinicians may unknowingly use PVTs with this population (Astley, 2010; Cox, Clairmont, & Cox, 2008; May et al., 2018; Sokol, Delaney-Black, & Nordstom, 2003). However, to our knowledge, there is limited evidence to support PVT validity among adults with FASD, in addition to justice-involved

adults, despite high rates of cognitive impairment and frequent PVT use in these populations (Farrer & Hedges, 2011; Hellenbach, Karatzias, & Brown, 2017; LaDuke, Brodale, & Rabin, 2016). In the FASD diagnostic context, inaccurate identification of invalid performance may lead to missed diagnosis, poor understanding of cognitive deficits and needs, and limited access to appropriate services. In legal contexts, consequences may be particularly serious and could include lengthier incarceration terms or inability to access appropriate defense or legal safeguards (e.g., fitness to stand trial). Research examining PVTs in cognitively impaired populations provides a helpful starting point for considering FASD PVT validity. However, FASD may be distinguished from other neurodevelopmental and cognitive disorders based on phenotypic variability, often preserved overall intellectual ability, high comorbidity with physical and mental health conditions, and high rates of criminal justice-involvement, highlighting the need for focused study in this group (Pei et al., 2011; Popova et al., 2016; Streissguth et al., 2004).

A limited number of studies have examined PVT performance patterns among individuals with FASD. Two studies focusing on the WMT and Medical Symptom Validity Test (MSVT, Green 2003) have shown that both children and adults with FASD performed better on these measures and failed less frequently compared to those with mild TBI (Green, Montijo, & Brockhaus, 2011; Larson et al., 2015). Similarly, within the WMT standardization sample, a subset of 19 youth with fetal alcohol syndrome scored above the clinical cutoff on all three effort subtests, on average, suggesting valid performance (Green, 2003). While these findings may provide preliminary support for the valid use of the WMT in children and adolescents with FASD, research examining PVT performance in adults with FASD is limited. Moreover, given

the importance of using multiple measures to assess PVT validity, additional research is needed to disentangle performance patterns in this population across a wider range of measures.

Current Study

The current study sought to evaluate the validity of 10 commonly used PVT scores in a sample of justice-involved adults diagnosed with FASD or possible FASD, compared to a control group of adults in the criminal justice system (CJS) who did not have FASD. Based on findings from other neurocognitively compromised populations, we expected that individuals with diagnosed and possible FASD would show worse performance and higher failure rates on PVTs, compared to those without FASD (e.g., Dean et al., 2008; Merten et al., 2007).

Method

Participants

Data for this study was drawn from a larger project (McLachlan et al., 2019). Participants included 80 justice-involved adults from a Northern Canadian correctional jurisdiction. Participants were consecutively recruited from both jail and community-based criminal justice settings and each had current legal involvement, or were in custody either pre- or post-adjudication. Recruitment occurred over an 18-month period, using information sessions, posters, and direct referrals by probation officers and case managers. In total, 174 prospective participants were approached by the research team, 45 declined participation, and 50 were deemed ineligible, primarily due to age >40 years or discontinued contact with the research team. Individuals who were under a review board supervision order, or considered medically or psychiatrically unstable were also excluded from this study.

Participants provided written informed consent and study procedures were reviewed and approved by the Children's and Women's Research Ethics Board at the University of British

Columbia, and Research Ethics Board at the University of Guelph. Although participants were provided an incentive commensurate with time spent participating in the study, their performance on this assessment was not tied to a clinical or legal outcome, thereby potentially diminishing external gains associated with intentional non-credible responding.

Overall, the sample was predominantly male and ranged from 18 to 40 years (M = 29.38, SD = 5.34) (Table 1). Participants were primarily assessed in a correctional facility (n = 70, 87.5% incarcerated) and two-thirds were awaiting adjudication at the time of study (n = 50, 62.5%). In total, 14 participants (17.5%) were diagnosed with FASD, and FASD was ruled out in 55 cases (69%). Another 11 individuals (13.8%) presented with neurocognitive impairment consistent with FASD, however, an FASD diagnosis could not be confirmed owing to inadequate information concerning PAE in most cases (required for diagnosis). Given similar neurocognitive presentations (e.g., cognition, academic skills, attention, memory, executive function, adaptive skills) between the confirmed and possible FASD groups, they were combined in the current study. Therefore, the FASD group includes 25 individuals with a confirmed or possible FASD diagnosis, while the criminal justice (CJ) group includes 55 individuals for whom a diagnosis of FASD was ruled out.

Procedure

Participants completed a comprehensive evaluation for FASD undertaken by a multidisciplinary team that adhered to the 2005 Canadian Diagnostic Guidelines for FASD[3] (Chudley et al., 2005). This included a semi-structured interview canvassing personal, social, and medical history, analysis of three-digit facial photographs for sentinel facial features, medical

[3] FASD includes fetal alcohol syndrome (FAS) and partial fetal alcohol syndrome (pFAS), which are diagnosed when sentinel facial features, growth retardation, and neurodevelopmental impairment are present, and alcohol-related neurodevelopmental disorder (ARND), which is diagnosed when there is neurodevelopmental impairment in the absence of physical indicators (Chudley et al., 2005).

assessment, and a comprehensive psychological assessment completed by psychologists with supervision from expert neuropsychologists. Features of FASD (e.g., growth restriction, facial features, neurocognitive deficits, and PAE) were ranked and identified according to recommended cutoff scores, and diagnostic decisions were made following an interdisciplinary case conference (Chudley et al., 2005). The larger cognitive test battery included ten PVT scores from both stand alone and embedded measures, including the WMT, Genuine Memory Impairment Profile (GMIP), RDS, Digit Span age corrected scaled score (DS ACSS), CD ACSS, SS ACSS, CD-SS ACSS, Vocabulary-Digit Span ACSS (VC-DS ACSS), LM-II-R and Word Choice (WC).

Measures

Word Memory Test (WMT; Green, 2003). The WMT is a stand-alone PVT comprising multiple effort indicators in the context of a verbal memory task. In the current study, three WMT subtests were administered, including Immediate Recall (IR), Delayed Recall (DR) and Consistency (CNS). IR and DR measure an individual's ability to remember a list of 20 word-pairs immediately after exposure (IR) and at a 30-minute delay (DR), whereas CNS provides a measure of response consistency from IR to DR. A score <82.5% correct on IR, DR, or CNS is classified as failure (Green, 2003). Research suggests that the primary WMT classification decision is relatively insensitive to neurological diseases and memory impairment (Green, 2003). In the current study, we applied the standard <82.5% cutoff score for IR, DR, or CNS. However, this was not used to calculate the total number of measures failed, and was instead replaced with GMIP.

Genuine Memory Impairment Profile (GMIP; Green et al., 2011). GMIP is an alternative WMT criterion designed to reduce false positives in cognitively impaired populations

by differentiating performance below standard cutoff scores (Alverson, O'Rourke, & Soble, 2019; Rienstra, Twennaar, & Schmand, 2013). Criteria for an invalid GMIP profile involves failure on ≥1 effort subtest (IR, DR, CNS) and a discrepancy ≥30 between the means of WMT effort and memory subtests (multiple choice, paired associates, free recall) (Green et al., 2011). Research suggests that the GMIP results in lower failure rates than the WMT, and adequate specificity and sensitivity in clinical samples with mild cognitive impairment (Alverson et al., 2019; Green et al., 2011). In the current study, we applied the standard cutoff criterion involving <82.5% on IR, DR, or CNS and ≥30 discrepancy, which was used to calculate the total number of measures failed. We also applied adjusted discrepancy criteria of ≥35, 40 and 45.

Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV; Wechsler, 2008). The WAIS-IV is an overall measure of intellectual functioning for adults. In the current study, six commonly used embedded WAIS-IV PVT scores were considered, including RDS, DS ACSS, CD ACSS, SS ACSS, CD-SS ACSS and VC-DS ACSS. Digit Span (DS) provides a measure of attention and working memory, from which a commonly used PVT can be derived by combining the total number of digits correctly recalled on two successive trials of both DS Forward and DS Backward. Previously, RDS ≤7 was considered indicative of invalid performance (Axelrod, Fichtenberg, Millis, & Wertheimer, 2006; Schroeder et al., 2012; Zenisek et al., 2016). More recent research suggests that a cutoff ≤6 yields improved specificity, though caution is advised in populations with ID and severe memory impairment (Schroeder et al., 2012; Webber & Soble, 2018). Using a mean cut off score of 7.1 across 24 studies, RDS has shown moderate sensitivity (63%), good specificity (86%) and an overall hit rate of 76% in distinguishing valid from suboptimal effort (Jasinski, Berry, Shandera, & Clark, 2011). In the current study, we applied the

≤7 cutoff , which was used to calculate the total number of measures failed among participants, as well as the ≤6 cutoff for exploratory purposes.

DS ACSS has also been applied as a PVT, with scaled scores ≤5 suggesting invalid performance (Webber & Soble, 2018). Evidence suggests that DS ACSS may be as effective as RDS, and potentially superior among older (e.g., 39-69) clinical groups and those at higher risk of neurocognitive impairment (Jasinski et al., 2011; Reese, Suhr, & Riddle, 2012; Spencer et al., 2013, 2017; Webber & Soble, 2018). In a sample of veterans referred to a neuropsychological clinic, DS ACSS significantly predicted group membership (e.g., valid vs. invalid performance) with an AUC of .85 (Webber & Soble, 2018). In the current study, we applied the ≤5 cutoff score for DS ACSS.

Additional WAIS-IV ACSS have been evaluated as embedded PVTs, including Coding (CD) and Symbol Search (SS), both measures of processing speed. CD ACSS ≤5 has shown good specificity (.90 – 1.00), but low and variable sensitivity (.04 - .64) for identifying invalid performance in mixed clinical samples (Erdodi et al., 2017; Erdodi & Lichtenstein, 2017). SS ACSS scores ≤6 have shown similarly variable sensitivity (.38 - .64) and good specificity (.88 - .93) in a mixed clinical sample (Erdodi et al., 2017), though other studies have found that the SS ACSS failed to reach minimum specificity against other validated PVTs (Erdodi & Lichtenstein, 2017). CD-SS ACSS is another embedded WAIS-IV PVT calculated by taking the difference between the the CD and SS ACSS. Difference scores ≥3 on this measure have been shown to yield adequate specificity (.94) in forensic patients with schizophrenia, including a subset with a General Ability Index (GAI) between 70 and 79 (.97) and GAI ≤69 (.88) (Glassmire et al., 2019). In the current study, we applied the ≤5 criteria for CD ACSS; ≤ 6 for SS ACSS; and ≥3 for CD-SS ACSS.

The WAIS-IV Vocabulary (VC) subtest measures word knowledge and verbal concept formation and has been compared with DS as a possible PVT, with VC-DS ACSS differences ≥3 reflecting invalid performance. Mittenberg et al. (1995) found that VC-DS ACSS accurately classified 71% of cases instructed to provide invalid performance. Moreover, in a sample of 151 adults referred for neuropsychological assessment, Greve and colleagues (2003) found that the measure had good sensitivity (.67) and specificity (.80) for identifying invalid performance by participants with FSIQ ≥ 85, but poor sensitivity for those with FSIQ < 85. In the current study, we applied the standard ≥3 cutoff score for VC-DS ACSS.

Logical Memory II Recognition (WMS-IV; Wechsler, Holdnack, & Drozdick, 2009). The Wechsler Memory Scale – Fourth Edition (WMS-IV) is a neuropsychological test designed to assess memory in adults and comprises several PVTs, including one used in the current study. LM-II-R provides a measure of delayed verbal memory using a dichotomous recognition format, and most neurologically healthy examinees perform well on this task. LM-II-R has been applied as a PVT, with unexpectedly low raw scores (≤20) indicating invalid performance (Bortnik et al., 2010). Within the WMS-IV standardization sample, fewer than 25% of the clinical sample achieved a score indicative of poor effort, resulting in an accuracy rate of 67% (Pearson Assessment, 2009). In the current study, we applied the ≤20 cutoff criterion.

Word Choice (ACS; Pearson Assessment, 2009). Advanced Clinical Solutions (ACS) is a test battery designed to enhance the clinical utility of the WAIS-IV and WMS-IV. The ACS WC subtest is a stand-alone PVT that uses a forced-choice recognition memory paradigm. Using a criterion of ≤42 (raw score) on WC has shown good classification accuracy (86%) for individuals without cognitive impairment (Bain & Soble, 2017; Barhon, Batchelor, Meares, Chekaluk, & Shores, 2015; Miller et al., 2011). However, classification accuracy is thought to be

lower for individuals with cognitive impairment (69%), owing to reduced sensitivity (Bain & Soble, 2017; Davis, 2014). In the current study, we applied the standard ≤42 cutoff criterion.

Data Analysis

Participant characteristics and between-group comparisons on PVTs for dichotomous scores (pass/fail) were compared using *t*-tests and chi-square analyses. Failure on each of the PVTs was established using standard cutoff scores (see Measures). The total number of PVTs failed was calculated by summing the number of 'failure' classifications across nine PVT scores, excluding WMT (see Measures). Practice guidelines suggest that failing ≥2 PVTs within a battery of measures is indicative of invalid performance, though others recommend a more stringent criteria of ≥3 (Erdodi et al., 2018; Larrabee, 2008; Victor, Boone, Serpa, Buehler, & Ziegler, 2009). As a result, we examined the number of participants failing ≥1, 2, 3, and 4 PVTs within both participant groups. We also undertook exploratory GMIP analyses to identify a potential alternative percentage point difference for classifying suboptimal effort in this sample (e.g., differences ≥35, 40, 45). Failure rates are also presented for participants with IQ ≥70 and <70, based on clinical diagnostic criteria for ID (Carr & O'Reilly, 2016). This dichotomy was used to draw comparisons between individuals with low and higher IQ, but was understood to be clinically artificial given that many other considerations factor into a diagnosis of ID. Effect sizes are reported for all analyses, including Cohen's *d* (small = .2, medium = .5, large = .8), and phi (small = .1, medium = .3, large = ≥.35) (Cohen, 1988). 95% confidence intervals are also reported. Statistical analyses were conducted using IBM SPSS version 25.0 for Mac.

Results

Sample characteristics

There were few demographic differences between the groups, although participants with confirmed or possible FASD were, on average, three years younger, and presented with lower average IQ than the CJ group. In addition, substantially more participants in the FASD group (n = 19, 79%) had IQ <70 on a standard measure of intellectual functioning compared to the CJ group (n = 10, 19%) $\chi^2(1) = 25.58$, p <.001, φ=.58. The average IQ for the CJ group was approximately one-standard deviation below the general population average (M = 83.08, SD = 12.29).

PVT Performance

Participants in the FASD group performed substantially worse on most PVTs compared to both the CJ group, and in reference to published scores for individuals with and without severe cognitive impairment (Table 2). Scores for the CJ group also tended to be lower compared to published neurotypical scores and were comparable to populations with severe cognitive deficits (Table 2). Failure rates in the FASD group were highest on DS ACSS (76%), LM-II-R (68%) and CD ACSS (60%), and lowest on WC (4%) and RDS (13%).

Participants with FASD failed more PVTs (M = 3.52, SD = 1.29, range 1-6) compared to those in the CJ group (M = 1.51, SD = 1.37, 0-5), $t(78) = -6.18$, p <.001, d = 1.51, 95% CI = -2.66, -1.36 (Figure 1). All participants in the FASD group failed ≥1 PVT, and all but two failed ≥2 (n = 23, 92%). In contrast, substantially fewer participants in the CJ group failed any single PVT indicator (n = 39, 71%), $\chi^2(1) = 9.09$, p = .003, φ = .34. They were also less likely to fail ≥2 PVTs (n = 25, 46%) compared to the FASD group, $\chi^2(1) = 15.52$, p < .001, φ =.44. Using more stringent criteria, 80% of participants in the FASD group (n = 20) failed ≥3 PVTs, compared to

only 22% of those in the CJ group (n = 12), χ^2 (1) = 24.24, p < .001, φ =.55. Over half of participants in FASD group (n = 13, 52%) failed ≥4 PVTs, compared to only 9% (n = 5) in CJ group, χ^2 (1) = 18.15, p <.001, φ =.48. Last, a greater proportion of the FASD group (e.g., 35% - 76%) was classified in the 'fail' range on five PVTs (WMT, DS ACSS, CD ACSS, SS ACSS, LM-II-R), compared to the CJ group (9% - 35%) (Table 2).

Using the GMIP, failure rates decreased from 35% to 20% for the FASD group, and from 9% to 7% for the CJ group. Increasing the difference criterion for the GMIP from 30 to 35 resulted in a reduced failure rate of 16% (n = 4) for the FASD group, and no change for the CJ group (7%, n = 4). Further increasing the difference range to 40 points resulted in an additional lowering of the failure rate for both the FASD (12%, n = 3) and CJ (0%) groups. Increasing the difference criterion to 45 points resulted in only a marginal change in failure rate for the FASD group (e.g., 8%, n = 2).

Examining failure rates for individuals with low and high IQ scores, we found that substantially more participants with IQ <70 failed ≥1 PVT(s) (n = 29, 100%), compared to those with IQ ≥70 (n = 33, 69%) χ^2 (1) = 11.26, p = .001, φ =.38. Similarly, 90% (n = 26) of participants with IQ <70 failed ≥2 PVTs, compared to fewer than half of those with IQ ≥70 (n = 21, 44%) χ^2 (1) = 16.02, p < .001, φ =.46. Of those with IQ ≥70 who failed ≥2 PVTs, 19% (n = 4) had diagnosed/possible FASD, and 81% (n = 17) were not diagnosed with FASD.

Discussion

Assessing PVT validity is critical in the context of neuropsychological and cognitive evaluation, particularly in forensic and medicolegal contexts (Bush et al., 2005; Larrabee, 2003). The current study undertook a novel investigation of ten commonly used PVT scores in justice-involved adults with diagnosed/possible FASD. Consistent with studies evaluating PVT validity

in adults with a range of neurocognitive deficits, we found worse performance across multiple PVT indicators for individuals with diagnosed and possible FASD, compared to CJ controls (Dean et al., 2008; Merten et al., 2007; Zenisek et al., 2016). Almost all participants in the FASD group met criteria for non-credible responding based on the 'two-or-more' guideline, and more than half met criteria for suboptimal effort based on the 'three-or-more' guideline (Larrabee, 2008; Victor et al., 2009). Thus, participants with FASD were more likely to be identified as having provided invalid performance based on a series of nine PVT scores. This is consistent with a large body of research examining PVT use in groups with severe cognitive impairment, who show higher failure rates compared to unimpaired populations, and inadequate sensitivity and specificity for identifying invalid performance in these populations (Bain et al., 2019; Dean et al., 2008; Merten et al., 2007; Soble et al., 2018; Zenisek et al., 2016). This finding also highlights the importance of considering the relation between PVTs when multiple tests are administered, in order to avoid over-administration of similar measures, which may result in inflated failure rates (Berthelson, Mulchan, Odland, Miller, & Mittenberg, 2013; Odland, Lammy, Martin, Grote, & Mittenberg, 2015). Moreover, it is noteworthy that individuals with FASD performed worse on PVTs compared to CJ controls. This finding suggests that the deficits associated with FASD may increase the likelihood of PVT failure even when compared to other cognitively impaired populations.

Consistent with findings that the WMT may have inadequate specificity and classification accuracy in the context of cognitive impairment, participants in the FASD group were more likely to fail this measure using standard cutoff criterion, compared to the CJ group (Allen, Bigler, Larsen, Goodrich-Hunsaker, & Hopkins, 2007; Allen, Wu, & Bigler, 2011; Greve, Ord, Curtis, Bianchini, & Brennan, 2008; Merten et al., 2007). This finding stands in

contrast to that of Larson and colleagues (2015), who found low failure rates on the MSVT in a sample of children and adolescents with FASD. Several possible factors may account for this difference, including higher rates of poor health and cognitive impairment in the current justice-involved sample, compared to children and adolescents referred to a private practice for neuropsychological assessment (Larson et al., 2015). On the other hand, it is possible that the MSVT is more robust to cognitive impairment than its original counterpart. Indeed, findings from this study suggest that the GMIP may be more appropriate for use in FASD populations, given comparable failure rates to the CJ group, and lower failure rates compared to traditional WMT failure indicators. Nonetheless, in the current sample, it took considerable adjustment to the GMIP difference criterion in order to achieve lower failure rates. Moreover, it is not possible to know whether those identified as providing inadequate effort using the traditional GMIP represent false positives given the absence of an external criterion to validate classification. Thus, our findings suggest that using adjusted criteria developed for other clinically impaired populations, such as the GMIP, may not adequately protect against the risk of false positives for individuals with FASD, and further research is encouraged.

Participants in the FASD group were also more likely to fail a number of embedded PVTs, including DS ACSS, CD ACSS, SS ACSS and LM-II-R. This finding is consistent with suggestions that embedded PVTs may be less robust to cognitive impairment compared to stand-alone PVTs (Zenisek et al., 2016). For instance, findings regarding the utility of LM-II-R appear mixed, and recent studies have shown that it has inadequate classification accuracy compared to other commonly used PVTs (e.g., WC) and mixed specificity and sensitivity in clinical samples, including those with cognitive impairment (Bain et al., 2019; Erdodi, Tyson, et al., 2018; Greve et al., 2008; Miller et al., 2011; Soble et al., 2018; Webber & Soble, 2018). Thus, the current

findings highlight the need for further research examining performance patterns using embedded PVTs in populations with severe cognitive impairment, given their inextricable relation to cognitive ability. Moreover, caution may be warranted when using these measures in the context of severe cognitive impairment until further research is undertaken to explore their utility in these populations.

In evaluating the extent to which overall intellectual functioning contributes to PVT performance in CJ adults with and without FASD, the majority of participants with IQ <70 failed ≥2 PVTs. However, nearly half of participants with IQ ≥70 also failed ≥ 2 PVTs. Functionally, this may suggest that low IQ alone is not a sufficient predictor of poor PVT performance in individuals with FASD. This is consistent with previous findings wherein patterns of failure and false positives have varied substantially between participants of varying cognitive abilities (Flaro et al., 2007; Green & Flaro, 2015; Love et al., 2014; Smith et al., 2014).

While the finding that individuals with FASD performed worse compared to those without FASD may be unsurprising in the context of a large body of research that cautions against the use of PVTs in cognitively impaired populations, the challenges associated with identification and diagnosis for individuals with FASD suggest that clinicians may be unknowingly using PVTs with this population. In the context of limited knowledge concerning FASD among clinicians, coupled with high rates of undiagnosed cases, the risk of potentially invalid PVT interpretation in this population may be significant and lead to inaccurate conclusions regarding invalid performance. Given the lack of external incentives linked with participants' performance in the context of this study as well as the variability of performance across measures, it is possible that the high failure rates for individuals with possible and diagnosed FASD could represent false positives.

Limitations and Future Directions

The current study uniquely contributes to the literature on the use of common, clinically-normed PVTs in adults with and without FASD recruited from the CJS. However, some limitations should be noted. The study sample size was small and geographically unique, suggesting that results best generalize to similar correctional jurisdictions and warrant further study before being applied to clinical, non-CJ populations. Although a comprehensive "gold standard" FASD evaluation was completed for all participants, this study did not control for additional diagnoses or impairments that may have impacted individuals' performance on PVTs, which may be particularly relevant in correctional populations (Farrer & Hedges, 2011; Hellenbach et al., 2017). Therefore, there is a need for further study in larger samples and exploring contributing mechanisms. Finally, the unique research design was thought to limit external incentives associated with performance, and therefore, motivation for non-credible responding. However, participants did not necessarily have incentive to perform the best of their ability, and may still have had motivation to perform poorly (An, Kaploun, Erdodi, & Abeare, 2017; Erdodi et al., 2018). As a result, additional research involving individuals with *bona fide* and feigned impairments associated with FASD is also needed, in order to assess the predictive validity and psychometrics of PVTs for this unique population. Moreover, future studies should aim to explore and propose alternative cutoff scores for PVT interpretation for use with individuals who may have FASD.

Implications

This study represents an important step towards understanding whether and how PVTs should be used for individuals with FASD in the criminal justice context, in addition to furthering the literature on PVT use for individuals with severe neurocognitive deficits. Ensuring

valid interpretation of PVTs is critical given the negative potential consequences associated with failure, particularly in criminal and civil legal contexts. For example, mislabelling individuals with true cognitive deficits as having provided invalid effort may result in incorrect diagnosis and prevent access to appropriate treatment opportunities or resources. In forensic and medicolegal contexts, this may extend to finding that an examinee was uncooperative or engaging in overt misrepresentation of true functioning. In turn, this may result in a range of adverse legal outcomes, including conviction, restriction from injury benefits, or restriction from legal safeguards, such as fitness to stand trial. The current findings highlight the need for further research examining PVT use in this unique population. In addition, developing practice guidelines may prove helpful in informing PVT interpretation in adults with FASD, particularly in legal and forensic contexts. There is also a critical need for increased FASD training among professionals in order to prevent misdiagnosis and ensure that clinicians understand the complex relationship between neurocognitive impairment, and potentially, FASD and PVTs, to support appropriate treatment and intervention practices.

References

Ali, S., Kerns, K. A., Mulligan, B. P., Olson, H. C., & Astley, S. J. (2018). An investigation of intra-individual variability in children with fetal alcohol spectrum disorder (FASD). *Child Neuropsychology*, *24*(5), 617–637. https://doi.org/10.1080/09297049.2017.1302579

Allen, M., Bigler, E. D., Larsen, J., Goodrich-Hunsaker, N., & Hopkins, R. (2007). Functional neuroimaging evidence for high cognitive effort on the word memory test in the absence of external incentives. *Brain Injury*, *21*(13–14), 1425–1428. https://doi.org/10.1080/02699050701769819

Allen, M., Wu, T. C. K., & Bigler, E. D. (2011). Traumatic Brain Injury alters Word Memory Test performance by slowing response time and increasing cortical activation: An fMRI study of a symptom validity test. *Psychological Injury and Law*, *4*(2), 140–146. https://doi.org/10.1007/s12207-011-9105-4

Alverson, A. W., O'Rourke, J. F., & Soble, J. R. (2019). The Word Memory Test Genuine Memory Impairment Profile discriminates genuine memory impairment from invalid performance in a mixed clinical sample with cognitive impairment. *The Clinical Neuropsychologist*, 1–16. https://doi.org/10.1080/13854046.2019.1599071

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington, DC: American Psychiatric Publishing.

An, K. Y., Kaploun, K., Erdodi, L. A., & Abeare, C. A. (2017). Performance validity in undergraduate research participants: a comparison of failure rates across tests and cutoffs. *Clinical Neuropsychologist*, *31*(1), 193–206. https://doi.org/10.1080/13854046.2016.1217046

Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and

malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *Clinical Neuropsychologist*, *21*(6), 899–916. https://doi.org/10.1080/13825580600966391

Ashendorf, L., Clark, E. L., & Sugarman, M. A. (2017). Performance validity and processing speed in a VA Polytrauma sample. *Clinical Neuropsychologist*, *31*(5), 857–866. https://doi.org/10.1080/13854046.2017.1285961

Astley, S. J. (2004). *Diagnostic guide for fetal alcohol spectrum disorders:the 4-digit diagnostic code* (3rd ed.). Seattle, WA: University of Washington Publication Services.

Astley, S. J. (2010). Profile of the first 1,400 patients receiving diagnostic evaluations for fetal alcohol spectrum disorder at the Washington State Fetal Alcohol Syndrome Diagnostic & Prevention Network. *Journal of Population Therapeutics and Clinical Pharmacology*, *17*(1), 132–164.

Axelrod, B. N., Fichtenberg, N. L., Millis, S. R., & Wertheimer, J. C. (2006). Detecting incomplete effort with digit span from the Wechsler Adult Intelligence Scale - Third Edition. *Clinical Neuropsychologist*, *20*(3), 513–523. https://doi.org/10.1080/13854040590967117

Bain, K. M., & Soble, J. R. (2017). Validation of the Advanced Clinical Solutions Word Choice Test (WCT) in a mixed clinical sample: Establishing classification accuracy, sensitivity/specificity, and cutoff scores. *Assessment*, 1–9. https://doi.org/10.1177/10731911177251

Bain, K. M., Soble, J. R., Webber, T. A., Messerly, J. M., Bailey, K. C., Kirton, J. W., & McCoy, K. J. M. (2019). Cross-validation of three Advanced Clinical Solutions performance validity tests: Examining combinations of measures to maximize classification of invalid

performance. *Applied Neuropsychology:Adult*, *0*(0), 1–11. https://doi.org/10.1080/23279095.2019.1585352

Barhon, L. I., Batchelor, J., Meares, S., Chekaluk, E., & Shores, E. A. (2015). A comparison of the degree of effort involved in the TOMM and the ACS word choice test using a dual-task paradigm. *Applied Neuropsychology:Adult*, *22*(2), 114–123. https://doi.org/10.1080/23279095.2013.863775

Berthelson, L., Mulchan, S. S., Odland, A. P., Miller, L. J., & Mittenberg, W. (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury*, *27*(7–8), 909–916. https://doi.org/10.3109/02699052.2013.793400

Bigler, E. D. (2012). Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*(04), 632–642. https://doi.org/10.1017/S1355617712000252

Boone, K. B. (2013). *Clinical practice of forensic neuropsychology*. New York, NY: Guilford.

Bortnik, K. E., Boone, K. B., Marion, S. D., Amano, S., Ziegler, E., Victor, T. L., & Zeller, M. A. (2010). Examination of various WMS-III Logical Memory Scores in the assessment of response bias. *Clinical Neuropsychologist*, *24*(2), 344–357. https://doi.org/10.1080/13854040903307268

Brockhaus, R., & Merten, T. (2004). Neuropsychologische diagnostik suboptimalen leistungsverhaltens mit dem Word Memory Test. *Nervenarzt*, *75*(9), 882–887. https://doi.org/10.1007/s00115-003-1592-4

Bush, S. S., Heilbronner, R. L., & Ruff, R. M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and*

Law, *7*(3), 197–205. https://doi.org/10.1007/s12207-014-9198-7

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., … Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, *20*(4), 419–426. https://doi.org/10.1016/j.acn.2005.02.002

Carr, A., & O'Reilly, G. (2016). Diagnosis, classification and epidemiology. In *The handbook of intellectual disability and clinical psychology practice.* London, UK: Routledge.

Chudley, A. E., Conry, J., Cook, J. L., Loock, C., Rosales, T., & Leblanc, N. (2005). Fetal alcohol spectrum disorder: Canadian guidelines for diagnosis. *Canadian Medical Association Journal*, *172*, 1–21.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.).* (2nd ed.). New York, NY: Academic Press.

Coles, C. D., Gailey, A. R., Mulle, J. G., Kable, J. A., Lynch, M. E., & Jones, K. L. (2016). A comparison among 5 methods for the clinical diagnosis of fetal alcohol spectrum disorders. *Alcoholism: Clinical and Experimental Research*, *40*(5). https://doi.org/10.1111/acer.13032

Cook, J. L., Green, C. R., Lilley, C. M., Anderson, S. M., Baldwin, M. E., Chudley, A. E., … Rosales, T. (2016). Fetal Alcohol Spectrum Disorder: A guideline for diagnosis across the lifespan. *Canadian Medical Association Journal*, *188*(3), 191–197.

Cox, L. V, Clairmont, D., & Cox, S. (2008). Knowledge and attitudes of criminal justice professionals in relation to fetal alcohol spectrum disorder. *Canadian Society for Clinical Pharmacology*, *15*(2), e306-13.

Currie, B. A., Hoy, J., Legge, L., Temple, V. K., & Tahir, M. (2016). Adults with fetal alcohol spectrum disorder: Factors associated with positive outcomes and contact with the criminal

justice system. *Journal of Population Therapeutics and Clinical Pharmacology*, *23*(1), e37–e52.

Daubert v. Merrell Dow Pharmaceuticals. (1993). 509 US 579.

Davis, J. J. (2014). Further consideration of Advanced Clinical Solutions Word Choice: Comparison to the Recognition Memory Test-Words and classification accuracy in a clinical sample. *The Clinical Neuropsychologist*, *28*(8), 1278–1294. https://doi.org/10.1080/13854046.2014.975844 Further

Dean, A. C., Victor, T. L., Boone, K. B., & Arnold, G. (2008). The relationship of IQ to effort test performance. *Clinical Neuropsychologist*, *22*(4), 705–722. https://doi.org/10.1080/13854040701440493

Dwyer, C. A. (1996). Cut scores and testing: Statistics, judgment, truth, and error. *Psychological Assessment*, *8*(4), 360–362. https://doi.org/10.1037/1040-3590.8.4.360

Erdodi, L. A., Abeare, C. A., Lichtenstein, J. D., Tyson, B. T., Kucharski, B., Zuccato, B. G., & Roth, R. M. (2017). Wechsler adult intelligence scale-fourth edition (WAIS-IV) processing speed scores as measures of noncredible responding: The third generation of embedded performance validity indicators. *Psychological Assessment*, *29*(2), 148–157. https://doi.org/10.1037/pas0000319

Erdodi, L. A., Hurtubise, J. L., Charron, C., Dunn, A., Enache, A., McDermott, A., & Hirst, R. B. (2018). The D-KEFS trails as performance validity tests. *Psychological Assessment*, *30*(8), 1082–1095. https://doi.org/10.1037/pas0000561

Erdodi, L. A., & Lichtenstein, J. D. (2017). Invalid before impaired: An emerging paradox of embedded validity indicators. *Clinical Neuropsychologist*, *31*(6–7), 1029–1046. https://doi.org/10.1080/13854046.2017.1323119

Erdodi, L. A., Tyson, B. T., Abeare, C. A., Zuccato, B. G., Rai, J. K., Seke, K. R., … Roth, R. M. (2018). Utility of critical items within the Recognition Memory Test and Word Choice Test. *Applied Neuropsychology:Adult*, *25*(4), 327–339. https://doi.org/10.1080/23279095.2017.1298600

Farrer, T. J., & Hedges, D. W. (2011). Prevalence of traumatic brain injury in incarcerated groups compared to the general population: A meta-analysis. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *35*(2), 390–394. https://doi.org/10.1016/j.pnpbp.2011.01.007

Flannigan, K., Pei, J., Stewart, M., & Johnson, A. (2018). Fetal Alcohol Spectrum Disorder and the criminal justice system: A systematic literature review. *International Journal of Law and Psychiatry*, *57*, 42–52. https://doi.org/10.1016/j.ijlp.2017.12.008

Flaro, L., Green, P., & Robertson, E. (2007). Word Memory Test failure 23 times higher in mild brain injury than in parents seeking custody: The power of external incentives. *Brain Injury*, *21*(4), 373–383. https://doi.org/10.1080/02699050701311133

Glassmire, D. M., Wood, M. E., Ta, M. T., Kinney, D. I., & Nitch, S. R. (2019). Examining false-positive rates of Wechsler Adult Intelligence Scale (WAIS-IV) processing speed-based embedded validity indicators among individuals with schizophrenia spectrum disorders. *Psychological Assessment*, *31*(1), 120–125. https://doi.org/10.1037/pas0000650

Green, P. (2003). *Manual for the Word Memory Test for Windows*. Edmonton, Alberta: Green's Publishing.

Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the word memory test and medical symptom validity test in groups with severe verbal memory impairment. *Applied Neuropsychology*, *18*(2), 86–94. https://doi.org/10.1080/09084282.2010.523389

Green, P., & Flaro, L. (2015). Results from Three Performance Validity Tests (PVTs) in Adults with Intellectual Deficits. *Applied Neuropsychology:Adult*, *22*(4), 293–303. https://doi.org/10.1080/23279095.2014.925903

Green, P., Lees-Haley, P. R., & Allen, L. M. (2003). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology*, *2*(3–4), 97–124. https://doi.org/10.1300/J151v02n03_05

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*(3), 245–260. https://doi.org/10.1016/S0887-6177(02)00137-3

Greve, K. W., Ord, J., Curtis, K. L., Bianchini, K. J., & Brennan, A. (2008). Detecting malingering in traumatic brain injury and chronic pain: A comparison of three forced-choice symptom validity tests. *Clinical Neuropsychologist*, *22*(5), 896–918. https://doi.org/10.1080/13854040701565208

Hellenbach, M., Karatzias, T., & Brown, M. (2017). Intellectual disabilities among prisoners: Prevalence and mental and physical health comorbidities. *Journal of Applied Research in Intellectual Disabilities*, *30*, 230–241.

Hoyme, H. E., Kalberg, W. O., Elliott, A. J., Blankenship, J., Buckley, D., Marais, A.-S., … May, P. A. (2016). Updated clinical guidelines for diagnosing fetal alcohol spectrum disorders. *Pediatrics*, *138*(2), e20154256–e20154256. https://doi.org/10.1542/peds.2015-4256

Jasinski, L. J., Berry, D. T. R., Shandera, A. L., & Clark, J. A. (2011). Use of the Wechsler Adult Intelligence Scale Digit Span subtest for malingering detection: A meta-analytic review.

Journal of Clinical and Experimental Neuropsychology, *33*(3), 300–314. https://doi.org/10.1080/13803395.2010.516743

Kodituwakku, P., & Kodituwakku, E. (2014). Cognitive and behavioral profiles of children with fetal alcohol spectrum disorders. *Current Developmental Disorders Reports*, *1*(3), 149–160. https://doi.org/10.1007/s40474-014-0022-6

LaDuke, C. C., Brodale, D., & Rabin, L. (2016). Common test utilization in forensic neuropsychological assessment: A survey of National Academy of Neuropsychology (NAN) and International Neuropsychology Society (INS) members. *Archives of Clinical Neuropsychology*, *31*, 573–583. https://doi.org/10.1093/arclin/acw042

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D)*, *17*(3), 410–425. https://doi.org/10.1076/clin.17.3.410.18089

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *Clinical Neuropsychologist*, *22*(4), 666–679. https://doi.org/10.1080/13854040701494987

Larson, J. C., Flaro, L., Peterson, R. L., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2015). The medical symptom validity test measures effort not ability in children: A comparison between mild TBI and fetal alcohol spectrum disorder samples. *Archives of Clinical Neuropsychology*, *30*(3), 192–199. https://doi.org/10.1093/arclin/acv012

Love, C. M., Glassmire, D. M., Zanolini, S. J., & Wolf, A. (2014). Specificity and false positive rates of the Test of Memory Malingering, Rey 15-Item Test, and Rey Word Recognition Test among forensic inpatients with intellectual disabilities. *Assessment*, *21*(5), 618–627. https://doi.org/10.1177/1073191114528028

Marshall, P., & Happe, M. (2007). The performance of individuals with mental retardation on cognitive tests assessing effort and motivation. *Clinical Neuropsychologist*, *21*(5), 826–840. https://doi.org/10.1080/13854040600801001

Mattson, S. N., Bernes, G. A., & Doyle, L. R. (2019). Fetal alcohol spectrum disorders: A review of the neurobehavioral deficits associated with prenatal alcohol exposure. *Alcoholism: Clinical and Experimental Research*, 1–17. https://doi.org/10.1111/acer.14040

Mattson, S. N., Crocker, N., & Nguyen, T. T. (2011). Fetal alcohol spectrum disorders: Neuropsychological and behavioral features. *Neuropsychology Review*, *21*(2), 81–101. https://doi.org/10.1007/s11065-011-9167-9

May, P. A., Baete, A., Russo, J., Elliott, A. J., Blankenship, J., Kalberg, W. O., … Hoyme, H. E. (2014). Prevalence and characteristics of fetal alcohol spectrum disorders. *Pediatrics*, *134*(5), 855–866. https://doi.org/10.1542/peds.2013-3319

May, P. A., Chambers, C. D., Kalberg, W. O., Zellner, J., Feldman, H., Buckley, D., … Hoyme, H. E. (2018). Prevalence of fetal alcohol spectrum disorders in 4 US communities. *JAMA - Journal of the American Medical Association*, *319*(5), 474–482. https://doi.org/10.1001/jama.2017.21896

McLachlan, K., McNiel, A., Pei, J., Brain, U., Andrew, G., & Oberlander, T. F. (2019). Prevalence and characteristics of adults with fetal alcohol spectrum disorder in corrections: A Canadian case ascertainment study. *BMC Public Health*, *19*(43). https://doi.org/https://doi.org/10.1186/s12889-018-6292-x

McLachlan, K., Rasmussen, C., Oberlander, T. F., Loock, C., Pei, J., Andrew, G., … Weinberg, J. (2016). Dysregulation of the cortisol diurnal rhythm following prenatal alcohol exposure and early life adversity. *Alcohol*, *53*, 9–18. https://doi.org/10.1016/j.alcohol.2016.03.003

Merten, T., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology*, *29*(3), 308–318. https://doi.org/10.1080/13803390600693607

Miller, J. B., Millis, S. R., Rapport, L. J., Bashem, J. R., Hanks, R. A., & Axelrod, B. N. (2011). Detection of insufficient effort using the advanced clinical solutions for the wechsler memory scale, fourth edition. *Clinical Neuropsychologist*, *25*(1), 160–172. https://doi.org/10.1080/13854046.2010.533197

Mittenberg, W., Theroux-Fichera, S., Zielinski, R. E., & Heilbronner, R. L. (1995). Identification of malingered head injury on the Wechsler Adult Intelligence Scale-Revised. *Professional Psychology: Research and Practice*, *26*(5), 491–498. https://doi.org/10.1037/0735-7028.26.5.491

Mukherjee, R. A. S., Hollins, S., & Turk, J. (2006). Fetal alcohol spectrum disorders: An Overview. *Journal of the Royal Society of Medicine*, *99*(6), 298–302.

Odland, A. P., Lammy, A. B., Martin, P. K., Grote, C. L., & Mittenberg, W. (2015). Advanced administration and interpretation of multiple validity tests. *Psychological Injury and Law*, *8*(1), 46–63. https://doi.org/10.1007/s12207-015-9216-4

Pearson Assessment. (2009). *Advanced Clinical Solutions for the WAIS-IV/WMS-IV*. San Antonio, TX: Pearson.

Pei, J., Denys, K., Hughes, J., & Rasmussen, C. (2011). Mental health issues in fetal alcohol spectrum disorder. *Journal of Mental Health*, *20*(5), 473–483. https://doi.org/10.3109/09638237.2011.577113

Popova, S., Lange, S., Chudley, A. E., Reynolds, J. N., Rehm, J., May, P. A., & Riley, E. P.

(2018). World Health Organization International Study on the prevalence of fetal alcohol spectrum disorder (FASD). In *Centre for Addition and Mental Health*. https://doi.org/10.1111/j.1755-9294.2010.01074.x

Popova, S., Lange, S., Shield, K., Burd, L., & Rehm, J. (2019). Prevalence of fetal alcohol spectrum disorder among special subpopulations: a systematic review and meta-analysis. *Addiction*, 1150–1172. https://doi.org/10.1111/add.14598

Popova, S., Lange, S., Shield, K., Mihic, A., Chudley, A. E., Mukherjee, R. A. S., … Rehm, J. (2016). Comorbidity of fetal alcohol spectrum disorder: A systematic review and meta-analysis. *The Lancet*, *387*(10022), 978–987. https://doi.org/10.1016/S0140-6736(15)01345-8

R. v. Peters. (2011). SKQB 419 (CanLii).

R v. Mohan. (1994). 2 SCR 9.

Reese, C. S., Suhr, J. A., & Riddle, T. L. (2012). Exploration of malingering indices in the wechsler adult intelligence scale-fourth edition digit span subtest. *Archives of Clinical Neuropsychology*, *27*(2), 176–181. https://doi.org/10.1093/arclin/acr117

Rienstra, A., Twennaar, M. K., & Schmand, B. (2013). Neuropsychological characterization of patients with the WMT dementia profile. *Archives of Clinical Neuropsychology*, *28*(5), 463–475. https://doi.org/10.1093/arclin/act026

Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable Digit Span: A systematic review and cross-validation study. *Assessment*, *19*(1), 21–30. https://doi.org/10.1177/1073191111428764

Shandera, A. L., Berry, D. T. R., Clark, J. A., Schipper, L. J., Graue, L. O., & Harp, J. P. (2010). Detection of malingered mental retardation. *Psychological Assessment*, *22*(1), 50–56.

https://doi.org/10.1037/a0016585

Simon, M. J. (2007). Performance of mentally retarded forensic patients on the Test of Memory Malingering. *Journal of Clinical Psychology*, *63*(4), 339–344. https://doi.org/10.1002/jclp

Smith, K., Boone, K. B., Victor, T., Miora, D., Cottingham, M., Ziegler, E., … Wright, M. (2014). Comparison of credible patients of very low intelligence and non-credible patients on neurocognitive performance validity indicators. *The Clinical Neuropsychologist*, *28*(6), 1048–1070. https://doi.org/10.1080/13854046.2014.931465

Soble, J. R., Bain, K. M., Bailey, K. C., Kirton, J. W., Marceaux, J. C., Critchfield, E. A., … O'Rourke, J. J. F. (2018). Evaluating the accuracy of the Wechsler Memory Scale-Fourth Edition (WMS-IV) logical memory embedded validity index for detecting invalid test performance. *Applied Neuropsychology:Adult*, 1–8. https://doi.org/10.1080/23279095.2017.1418744

Sokol, R. J., Delaney-Black, V., & Nordstom, B. (2003). Fetal alcohol spectrum disorder (FASD). *Journal of American Medical Association*, *290*(22), 22–25.

Spencer, R. J., Axelrod, B. N., Drag, L. L., Waldron-Perrine, B., Pangilinan, P. H., & Bieliauskas, L. A. (2013). WAIS-IV Reliable Digit Span is no more accurate than age corrected scaled score as an indicator of invalid performance in a veteren sample undergoing evaluation for mTBI. *The Clinical Neuropsychologist*, *27*(8), 1362–1372. https://doi.org/http://dx.doi.org/10.1080/13854046.2013.845248 WAIS-IV

Spencer, R. J., Waldron-Perrine, B., Drag, L. L., Pangilinan, P. H., Axelrod, B. N., & Bieliauskas, L. A. (2017). Neuropsychological test validity in Veterens presenting with subjective complaints of "very severe" cognitive symptoms following mild traumatic brain injury. *Brain Injury*, *31*(1), 32–38.

https://doi.org/http://dx.doi.org/10.1080/02699052.2016.1218546 ORIGINAL

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*(5), 423–444. https://doi.org/10.1016/S0887-6177(01)00126-3

Streissguth, A. P., Bookstein, F. L., Barr, H. M., Sampson, P. D., O'Malley, K., & Young, J. K. (2004). Risk factors for adverse life outcomes in Fetal Alcohol Sydnrome and Fetal Alcohol Effects. *Journal of Developmental and Behavioral Pediatrics*, *25*(4), 228–238.

Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *Clinical Neuropsychologist*, *23*(2), 297–313. https://doi.org/10.1080/13854040802232682

Webber, T. A., & Soble, J. R. (2018). Utility of various WAIS-IV Digit Span indices for identifying noncredible performance validity among cognitively impaired and unimpaired examinees. *Clinical Neuropsychologist*, *32*(4), 657–670. https://doi.org/10.1080/13854046.2017.1415374

Wechsler, D., Coalson, D.L., & Raiford, S.E. (2008). Wechsler Adult Intelligence Test: Fourth Edition Technical and Interpretive Manual. San Antonio, TX: Pearson.

Wechsler, D., Holdnack, J.A., & Drozdick, L.W. (2009). Wechsler Memory Scale: Fourth Edition Technical and Interpretive Manual. San Antonio, TX: Pearson.

World Health Organization. (2007). *International Classification of Diseases and Related Health Problems* (10th ed.). Geneva, Switzerland.

Zenisek, R., Millis, S. R., Banks, S. J., & Miller, J. B. (2016). Prevalence of below-criterion Reliable Digit Span scores in a clinical sample of older adults. *Archives of Clinical Neuropsychology*, *31*(5), 426–433. https://doi.org/10.1093/arclin/acw025

Figure Legend

Figure 1. Number of Participants Failing 0 – 7 PVTs ($N = 80$)

Table 1: Demographic Characteristics

	FASD	CJ		
	(*n* = 25)	(*n* = 55)	X^2 (*t*)	φ (*d*)
Age (*M*, *SD*)	31.20 (4.63)	28 (5.47)	(-2.11)*	(.59)
Gender (*n*, % male)	23 (92%)	45 (81.8%)	1.40	-.13
Education (*n*, %)				
< Gr. 12/GED	22 (88%)	39 (70.9%)	2.77	-.19
> Gr. 12/GED	3 (12%)	16 (29.1%)		
Setting (*n*, %)				
Custody	24 (96%)	46 (85.2%)	1.98	-.16
Community	1 (4%)	8 (14.8%)		
Legal Status (*n*, %)				
Pre-adjudication	20 (80%)	30 (54.5%)	4.72*	-.24
Post-adjudication	5 (20%)	25 (45.5%)		
Cognitive Findings (*n*/*M*, %/*SD*)				
WAIS-IV FSIQ[a]	66.83 (5.21)	83.08 (12.29)	(6.21)**	(1.72)
IQ <70	19 (79.2%)	10 (18.9%)	25.58**	.58

Note. WAIS-IV = Wechsler Adult Intelligence Scale-IV (Wechsler, Coalson & Raiford, 2008). IQ = Full Scale Intelligence Quotient.
[a] = *N* = 77 as raw IQ scores were not available for three participants.
*p<.05, **p<.001.

Table 2: PVT Performance

	Current Sample						Published Comparisons	
		Total Scores *M* (SD) [range]		Failure Rates *n* (%)		X^2 (φ)	Total Scores *M* (*SD*)	
	Cut off	FASD	CJ	FASD	CJ		Neurotypical	Clinical
WMT	≤ 82.5%			8 (34.8%)	5 (9.4%)	7.27 (.31)**		
IR	on any[a]	88.02 (17.32) [27.5 - 100]	94.81 (6.28) [65 - 100]				98 (2.8)[a]	93.3 (5.2)[o]
DR		91.30 (11.38) [52.5 - 100]	95.09 (6.81) [57.5 - 100]				98.6 (2.4[a]	96.2 (4.7)[o]
CNS		82.39 (23.10) [10 - 100]	93.11 (7.07) [67.5 - 100]				96.8 (3.8)[a]	91.2 (6.8)[o]
GMIP	See measures[b]	29.47 (18.65) [3.3 - 70]	23.78 (12.82) [3.3 - 49.2]	5 (20%)	4 (7.3%)	2.79 (.19)	22.77 (10.5)[k]	37.51 (12.60)[k]
ACS								
WC	≤ 42/50[c]	47.65 (3.10) [36 - 50]	48.82 (2.30) [37 - 50]	1 (4%)	1 (1.8%)	.34 (.07)	48.36 (1.70)[l]	46.5 (4.4)[p]
WAIS-IV								
RDS	≤ 7[d]	8.65 (1.15) [6 - 11]	11.05 (2.07) [7 - 17]	3 (13%)	1 (2.3%)	3.12 (.22)	11.77 (2.20)[m]	8.2 (2.2)[p]
	≤ 6[e]			1 (4.3%)	0	1.94 (.17)		
DS ACSS	≤ 5[f]	4.6 (1.32) [2 - 7]	7.49 (2.10) [3 - 12]	19 (76%)	11 (22.4%)	19.69 (.52)**	8.9 (2.4)[f]	8.1 (2.9)[f]
CD ACSS	≤ 5[g]	5.08 (1.98) [1 - 9]	8.16 (2.55) [3 - 14]	15 (60%)	5 (10.2%)	20.81 (.53)**	7.8 (2.8)[n]	8.6 (2.9)[g]
SS ACSS	≤ 6[g]	6.8 (2.65) [2 - 12]	8.47 (2.43) [2 - 13]	13 (52%)	10 (20.4%)	7.71 (.32)**	-	8.8 (3.1)[g]
CD – SS ACSS	≥ 3[h]	1.96 (1.54) [0 - 6]	1.94 (1.56) [0 - 7]	9 (36%)	12 (24.5%)	1.08 (.12)	-	2.0 (1.8)[g]
VC – DS ACSS	≥ 3[i]	2 (1.44) [0 - 5]	2.35 (1.64) [0 - 7]	8 (32%)	21 (42.9%)	.82 (.11)	-	-
WMS-IV								
LM-II-R	≤20[i]	18.96 (3.92) [12 - 27]	21.72 (4.23) [2 - 28]	17 (68%)	19 (34.5%)	7.77 (.31)**	-	19.35 (3.62)[q]

Note. FASD = Fetal alcohol spectrum disorder. CJ = criminal justice sample. WMT = Word Memory Test (Green, 2003). GMIP = Genuine Memory Impairment Profile (Green et al., 2011); IR = Immediate Recall; DR = Delayed Recall; CNS = Consistency; ACS = Advanced Clinical Solutions (Pearson Assessment, 2009); WC = Word Choice; WAIS-IV = Wechsler Adult Intelligence Scale, 4th edition (Wechsler, 2008); RDS = Reliable Digit Span; DS ACSS = Digiti Span age-corrected scaled score; CD ACSS = Coding age-corrected scaled score; SS ACSS = Symbol Search age-corrected scaled score; CD-SS ACSS = Coding – Symbol Search age-corrected scaled score; VC-DS ACSS = Vocabulary – Digit Span age-corrected scaled score; WMS-III = Wechsler Memory Scale, Third Edition (Wechsler, 2009); LM II Recognition = Logical Memory II Recognition. [a]Green (2003), [b]Green, Lees-Haley, & Allen (2003), [c]Pearson Assessment (2009), [d]Axelrod, Fichtenberg, Millis, & Wertheimer (2006), [e]Schroeder, Twumasi-Ankrah, Baade, & Marshall (2012), [f]Webber & Soble (2018), [g]Erdodi et al. (2017), [h]Glassmire, Wood, Ta, Kinney, & Nitch (2019), [i]Mittenberg, Theroux-Fichera, Zielinski, & Heilbronner (1995), [j]Bortnik et al. (2010), [k]Alverson, O'Rourke, & Soble (2019), [l]Bain & Soble (2017), [m]Strauss et al. (2002), [n]Ashendorf, Clark, & Sugarman (2017), [o]Brockhaus & Merten (2004), [p]Miller et al. (2011), [q]Marshall & Happe (2007).
n = 67 – 80 due to missing data.
*p<.05, **p<.001.

Figure 1. Number of participants failing 0-6 PVTS ($N = 80$)