Neuropsychological Test Performance in Pediatric TBI, Learning Disability, and ADHD: Difference in Reliable Digit Span

By

Emily West

****

Senior Thesis

A thesis presented in partial fulfillment
Of the requirements for the degree of
Bachelor of Science
Department of Psychology
Neuroscience Program

UNION COLLEGE

Schenectady, New York

June 2017

ABSTRACT

WEST, EMILY   Neuropsychological Test Performance in Pediatric TBI, Learning Disability, and ADHD: Difference in Reliable Digit Span

Department of Psychology, June 2017.

ADVISOR: Cay Anderson-Hanley.

Children have been found to give non-credible performances on neuropsychological evaluations and using Performance Validity Tests (PVT) can help identify such cases, but maximum effort is necessary for tests to be reliable for determining next steps. This study assessed the differences in rates of non-credible performance on the Reliable Digit Span (RDS) between traumatic brain injury (TBI), learning disability (LD) and ADHD diagnoses in a pediatric clinical sample. It was hypothesized that those with ADHD would have a higher failure rate than those with either TBI or LD. RDS data from 200 clients referred to neuropsychological testing agencies were collected and scores were analyzed to determine mean differences and failure rates using a ≤ 6 cutoff score. Results showed a significant difference ($p = 0.005$) in RDS performance between those with a diagnosis of TBI or ADHD and between those with LD or ADHD ($p = .003$), with LD having a higher rate of failure than ADHD, and ADHD having a higher failure rate than TBI. The results indicate that rates of non-credible performance on RDS vary across diagnoses, which concurs with past literature, and emphasizes the importance of examining PVTs, like RDS, when conducting neuropsychological exams with children. Additional research is needed to clarify whether this finding persists with improved methodology (e.g., balanced sample sizes, clarified diagnoses, etc.).

TABLE OF CONTENTS

1. Introduction…………………………………………………………………pg. 4
2. Methods……………………………………………………………………pg. 12
3. Results……………………………………………………………………..pg. 15
4. Discussion………………………………………………………………....pg. 19
5. Acknowledgments…………………………………………………………pg. 25
6. References…………………………………………………………………pg. 26
7. Figure 1……………………………………………………………………...pg. 29
8. Figure 2……………………………………………………………………pg. 30
9. Figure 3……………………………………………………………………pg. 31
10. Figure 4……………………………………………………………………pg. 32
11. Figure 5……………………………………………………………………pg. 33
12. Table 1……………………………………………………………………pg. 34

INTRODUCTION

Symptom Validity Tests (SVTs) and Performance Validity Tests (PVTs) are frequently used in the assessment of client effort when completing neuropsychological testing (DeRight & Carone, 2015; Institute of Medicine, 2015). Clients are referred for assessment for various reasons, each of which can involve compensation in diverse forms, requiring proper assessment of the level of cognitive dysfunction in the individual (Institute of Medicine, 2015). Results of neuropsychological tests are used to help decide the appropriate interventions and/or compensation, but it is necessary for maximum effort to be exerted for the test results to be valid and useful in determining next steps. Children have been found to give non-credible performances on neuropsychological evaluations (DeRight & Carone, 2015; Kirkwood, 2015; Kirkwood & Kirk, 2010; Kirkwood, Kirk, Blaha, & Wilson, 2010) and using SVTs and PVTs can help in the identification of clients who do not put forth such effort.

Validity Testing as a Measure

More often than not, we do not have neuropsychological test scores for clients' baseline performance or knowledge of how they can be expected to perform in the case of injury or emergence of cognitive deficits. Thus, we need measures to make sure that clients are not feigning a poor performance and that such a performance is actually the result of cognitive damage or dysfunction. Patterns in the scores of multiple PVTs are interpreted to determine the validity of a client's performance, but scores of single PVTs can give a picture of potential non-credible performance as well (Institute of Medicine, 2015). It is important to note that making the determination of a non-credible performance, or invalid results, requires the use of multiple PVTs, since passing one test does not indicate optimal effort, while similarly, failing one test does not definitively indicate suboptimal effort (Institute of Medicine, 2015; Schroder, Twumasi-Ankrah, Baade, & Marshall, 2011).

Traditionally, PVTs are designed with a specific performance cutoff that is the number of errors expected before a non-credible performance is suspected. These cutoff scores are related to below-chance performance, such that even individuals with extensive impairment are capable of receiving good scores on PVTs, meaning that it is anticipated that nearly everyone will display excellent performance (Institute of Medicine, 2015). Those who are not intentionally doing poorly should not score below these cutoffs.

Reliable Digit Span as a Measure of Effort

PVTs come in various forms, such as embedded, derived, or standalone, each with pros and cons related to different aspects of administration, interpretation, sensitivity, and specificity (Institute of Medicine, 2015). Embedded tests, like the Reliable Digit Span (RDS), calculate scores from performance on measures already being used in the neuropsychological test battery. RDS is embedded in Digit Span, and is calculated by adding the highest number of digits correctly stated by the testee in both the forward and backward trials. A specific standard cutoff score is used, and if a client's RDS score falls below this score, then the credibility of their performance may be brought into question (DeRight & Carone, 2015; Institute of Medicine, 2015).

The RDS cutoff score used is determined by the number that yields the best specificity and sensitivity related to effort. In the development of the RDS measure, a score of ≤ 7 was used to identify those who were not giving optimal effort and has since been a commonly used cutoff score (Lezak, Howieson, & Loring, 2004; Babikian, Boone, Lu, & Arnold, 2006; Etherton, Bianchini, Ciota, & Greve, 2005). Other studies have also determined that a cutoff of ≤ 6, although lowering sensitivity, is also effective in detecting suboptimal effort (Etherton, Bianchini, Greve, & Heinly, 2005; Schroder, Twumasi-Ankrah, Baase, & Marshall, 2011). While different cutoff scores work to yield varying sensitivity and specificity, it has overall been determined that RDS scores are effective in indicating when a person may be suspected of suboptimal performance across diagnoses (Babikian, Boone, Lu & Arnold, 2006; Bodin, et al., 2014; Etherton, Bianchini, Greve,& Heinly, 2005; Jasinski, Berry, Shandera, Clark, 2011; Kirkwood, Hargrave, & Kirk, 2011; Mathias, Greve, Bianchini, Houston, & Crouch, 2002; Schroder, Twumasi-Ankrah, Baade, & Marshall, 2011).

PVT Use and Children

PVT use in children is much less common than it is in adults, with poor effort in children being interpreted by some as children's lack of interest in completing the testing or even psychologists' assumptions that children are unable of giving purposefully minimal effort. This is not true, however, as studies have shown children capable of purposeful deception on neuropsychological measures at ages as young as 6 years old and improving this ability as they get older and develop greater mental capacity (DeRight & Carone, 2015; Kirkwood, 2015; Kirkwood, Kirk, Blaha, & Wilson, 2010). Purposeful deception can also occur even when “incentives” are not obvious or salient, as can be the case when lawyers or parents encourage below-capability performance. Malingering is not only for financial gain, as may be the case with litigation and disability funding, as children attempt to avoid academic or athletic responsibility or to receive academic accommodations (DeRight & Carone, 2015; Kirkwood & Kirk, 2010). A lack of PVT use in pediatric arenas is likely the result of a lack of awareness of this capability in children and therefore an unknowing of their importance and need for use in the testing process with children (DeRight & Carone, 2015).

As cutoff scores for PVTs are used with adults, they are also used with children to assess the credibility and validity of their performance and results, respectively. While ≤ 7 has commonly been used as the RDS cutoff score in adults, that cutoff does not necessarily apply in the case of children. Kirkwood, Hargrave, and Kirk (2011) found that using the adult cutoff of ≤ 7 yielded too high a false-positive rate of failure when the children's RDS scores were compared with their performance on the MSVT and TOMM PVTs. It is recommended that a more conservative cutoff of ≤ 6 be used for determining non-credible performance in children when using the RDS measure (Kirkwood, Hargrave, & Kirk, 2011).

RDS Score with Age and IQ Relationships

Existing research reporting age and IQ relationships with PVT scores provide overall inconsistent conclusions on the significance of correlations. Where some studies have found significant correlations between PVT scores and age and/or IQ, others have found no such relationships.

In assessing the effectiveness of the RDS testing measure in children with varying diagnoses, several studies have found a positive correlation between age and RDS scores, with older children scoring higher on the test (Ernst, Kneavel, Gallo, Moylan, & Brosof, 2014; Welsh, Bender, Whitman, Vasserman, MacAllister, 2012). Other research, while using the RDS, TOMM, or MSVT as measures of effort, have found non-significant relationships between PVT scores and age (Gidley Larson, et al, 2015; Kirk, 2011; Kirkwood, Hargrave, & Kirk, 2011; Kirkwood, Yeates, Randolph, & Kirk, 2012). Gidley Larson, et al. (2015), while finding no overall correlation between MSVT score and age across the overall sample, did find a correlation within the Fetal Alcohol Spectrum Disorder (FASD) group, with significantly younger participants failing the test, a relationship that was not found within their mTBI group.

The literature assessing PVT performance in different populations shows contradictory and varying relationships between PVT and Full-Scale IQ (FSIQ) scores. In a study using RDS with a pediatric sample referred for academic problems, RDS scores were found to be non-significantly correlated with FSIQ scores (Ernst, Kneavel, Gallo, Moylan, & Brosof, 2014). In looking at other literature using the Medical Symptom Validity Test (MSVT) with a pediatric TBI population, Word Memory (WMT) with ADHD and LD college-aged population, and RDS with a pediatric and adolescent epilepsy population, however, FSIQ scores were found to be significantly correlated with the respective PVT score (Kirkwood, Yeates, Randolph, & Kirk, 2012; Sullivan, May, & Galbally, 2007; Welsh, Bender, Whitman, Vasserman, & MacAllister, 2012). Despite the differences in results regarding significance between PVT and FSIQ scores, it appears that the significant relationship is present across both test-type, including RDS, and age group.

RDS Across Diagnoses

Academic and Learning Diagnoses

Testing related to a referral for academic problems was an area of study in the literature, where LD and ADHD were diagnoses of particular interest. One such study examined performance on the RDS by pediatric referrals for academic concerns (Ernst, Kneavel, Gallo, Moylan, & Brosof, 2014). After testing was completed, some children received an appropriate diagnosis, if there was one, including LD, ADHD, or ADHD and LD. RDS scores did not significantly differ between LD, ADHD, and combined diagnoses, but the pass rate for RDS scores of roughly 92% maintains RDS as a valid measure of effort for such a population (Ernst, Kneavel, Gallo, Moylan, & Brosof, 2014). Sullivan, May, and Galbally (2007) used the WMT to also examine PVT performance between ADHD and LD, but in a college-aged sample and with the intention of analyzing rates of failure and symptom exaggeration. It was found that overall, 22% those tested for LD or ADHD showed evidence of low effort or non-credible performance, with ADHD-only diagnoses having a failure rate of 47%, LD 15%, and a combined ADHD/LD diagnosis having a 9% failure rate (Sullivan, May, & Galbally, 2007). Despite the differences in age group and PVT used, there is inconsistency in the literature when rates of failure between ADHD and LD are called into question, a topic the present study hopes to clarify.

Medical Diagnoses

Many of the studies using a pediatric sample look at TBI populations alone to assess the effectiveness of PVTs in detecting non-credible effort, but there appears to be limited literature comparing TBI failure rates with other diagnoses. From such limited research, Kirkwood (2015) found that non-credible tests are seen more often in children in rehabilitation programs for TBI and Gidley Larson, et al. (2015) found that children with mTBI displayed significantly worse performance than Fetal Alcohol Spectrum Disorder (FASD) children when using the MSVT as an effort measure.

Failure Rates Across Diagnoses

Among children and adolescents, the PVT failure rate in the literature appears to be considerably variable, particularly when examining different diagnostic groups. Children with TBI across multiple studies and use of varying PVTs show a failure rate ranging from 12 to 20% (Araujo, et al., 2014; DeRight & Carone, 2015; Gidley Larson, et al., 2015; Kirkwood, 2015; Kirkwood, Yeates, Randolph, & Kirk, 2012). Academically-based diagnoses also display a range in PVT failure rates, with LD diagnoses ranging from 9% when combined with a diagnosis of ADHD to 15% when diagnosed on its own, and ADHD ranging from 9% when combined with a diagnosis of LD to 47% when it is the only diagnosis (Sullivan, May, Galbally, 2007). It was expected that in the present study, the rate of failure would resemble the rates found in the literature for TBI, ADHD, and LD diagnoses. Also, based on the rates of failure from previous research, it was expected that those participants with a diagnosis of ADHD would have higher rates of failure when compared to those of either the TBI or the LD groups.

RDS and Other PVTs

Due to the fact that multiple PVTs exist for measuring effort on neuropsychological testing, people passing one PVT should also pass other PVTs administered in the test battery (Institute of Health, 2015).

The California Verbal Learning Test (CVLT) is another embedded PVT involving multiple trials measuring learning and memory of verbal information. Studies have suggested the use of different scores, particularly the recognition discriminability score or the forced-choice trial (with cutoff of ≤ 14), as moderately effective measures and indicators of effort and non-credible performance (Baker, Connery, Kirk, & Kirkwood, 2013; Brooks & Ploetz, 2015; Lichtenstein, Erdodi, & Linnea, 2017; Schwartz, et al., 2016). Because RDS and CVLT scores can both be used as indicators of potentially non-credible effort on tests, it can be expected that participants would perform similarly on the two measures.

In this retrospective study, data from neuropsychological testing agencies was collected and analyzed to assess the following hypothesis created based on the prior literature.

Hypotheses:

It is expected that:

1. Full-Scale IQ (FSIQ) will be positively correlated with RDS scores.
2. Age will be positively correlated with RDS scores.
3. Those with a diagnosis of ADHD will display worse performance on Reliable Digit Span than those with a Traumatic Brain Injury

(TBI) diagnosis

4. Those with ADHD will display worse performance on Reliable Digit Span than those with a Learning Disability (LD) diagnosis.
5. RDS scores will be positively correlated with CVLT scores.

METHODS

Participants

The sample (n=212) consisted of clients from the northeast aged 6-30 years who were referred for neuropsychological testing to two pediatric neuropsychologists for varying reasons related to medical and potential psychiatric, developmental, and neurological diagnoses. 27 participants were removed from the sample to be analyzed, however, as they were older in age (22-30 years) or did not have RDS score data, which were necessary for the analyses. There was a remaining pediatric sample of 185 participants aged 6-20 (M = 11.23; SD = 3.39) with both males (n=102) and females (n=83) who participated in the study.

In the present study, there is no data about which participants, if any, are involved in litigation cases, but there are potential secondary gains and incentives that could influence performance such as financial compensation if involved in litigation, prescription of medication, and academic aid, as has been discussed in the prior literature.

Procedures

The present study was approved by the Union College IRB as exempt due the lack of risk to the participants, the removal of all identifying information connecting the participants to study or the results, and the overall retrospective nature of the project.

Participant data of their performance on the Wechsler Intelligence Scale for Children - IV (WISC-IV), the Wechsler Adult Intelligence Scale -IV (WAIS- IV), and/or the California Verbal Learning Test for Children (CVLT-C) was collected from two neuropsychological testing agencies in the Northeast. To preserve anonymity of the participants, data from one agency was read aloud in-person from paper files and entered into a computerized data set with new ID numbers, and in addition, all identifying information, such as birth date, name, address, date of evaluation, etc., was withheld. Data from a second agency was collected via email with all identifying information removed from the set prior to being sent and new ID numbers being assigned once combined with data from the first agency.

Diagnoses of the participants, from both before testing occurred and based on the results of the evaluation being analyzed, were coded into six diagnostic categories: ADHD, TBI, LD, Neurological, Psychiatric, Developmental. Once the diagnoses to be analyzed (TBI, ADHD, LD) were decided upon based the prior research and the available data, columns were made to code for each group so that cases with dual- diagnoses within the ADHD/TBI and ADHD/LD comparisons were excluded from the analyses of the RDS scores. Other demographic information was also tracked for each participant.

For initial analyses, dual-diagnoses were allowed as long as the second diagnosis was not in the comparison analysis. For example, someone in the ADHD group with epilepsy was included, but someone with ADHD and TBI diagnoses would not be included in the RDS score analysis between ADHD and TBI diagnoses. After running initial analyses, the sample was further refined such that dual-diagnoses were removed for the ADHD samples in order to focus in on that specific diagnosis, leading to more balanced sample sizes and less noise in the pure sample groups.

Measures

The Wechsler Intelligence Scale for Children - IV (WISC; Wechsler, 2003) and the Wechsler Adult Intelligence Scale - IV (WAIS; Wechsler, 2014), each a psychological testing battery administered based on age to assess level of cognitive functioning and intelligence, were administered to all of the participants by pediatric neuropsychologists. IQ and RDS scores are both calculated as part of these assessments. RDS uses the sum of the longest number of forward and backward digits repeated on the Digit Span and has been shown as a valid measure of effort in both adults and children (Etherton, Bianchini, Greve, & Heinly, 2005; Ernst, Kneavel, Gallo, Moylan, & Brosof, 2014; Kirkwood, Hargrave, & Kirk, 2011; Schroder, Twumasi-Ankrah, Baade, & Marshall, 2011). The California Verbal Learning Test for Children (CVLT-C; Delis, Kramer, Kaplan & Ober, 1994) is a multi-trial word list recall and recognition test, with the final 15-item forced-choice recognition trial serving as an embedded PVT.

Statistical Analyses

Data was analyzed using the Statistical Package for the Social Sciences (SPSS v. 22) and Microsoft Excel. Independent t-tests were run to test hypotheses regarding differences in the means of RDS scores between diagnostic groups for the hypotheses described. Partial correlations were used to control for both age and IQ when examining relationships between RDS and diagnostic categories, and crosstab analyses using RDS score and diagnostic groups were run to determine number of participants scoring below the cutoff in each group.

RESULTS

Reliable Digit Span scores for the whole sample ranged from 2-14 (M = 7.35; SD = 1.76) with 67 of the 185 participants scoring below the ≤ 6 cutoff, yielding a 36% failure rate.

Hypothesis 1

There was a significant positive correlation between RDS score and FSIQ, r = .47, p < .001 (see Figure 1), such that those with higher FSIQ scores (M = 94, SD = 17.53) performed better, or had higher scores, on the RDS (M = 7.35; SD = 1.76). IQ scores ranged from 57-141 as determined on the WISC-IV or WAIS-IV. Of the 118 participants that had IQ data, 46 participants, or 39%, scored below the cutoff on the RDS and 8 of those who failed, or 17%, were within borderline deficiency or below range for IQ.

Hypothesis 2

There was a significant positive correlation between age and RDS score, r = .39, p < .001 (see Figure 2), such that older children (M = 11.23, SD = 3.39) performed better, or had higher scores, on the RDS (M = 7.35; SD = 1.76).

Hypothesis 3

When ADHD and TBI were both allowed to include dual-diagnoses other than each other, a two-tailed t-test revealed a significant difference in RDS scores between the ADHD (n=79) and TBI (n=15) groups, $t(92) = -2.52$, $p = 0.01$, such that ADHD scored lower, or displayed worse performance, ($M = 7.44$, $SD = 1.65$) than TBI ($M = 8.67$, $SD = 2.06$). Due to the fact that dual-diagnoses can create noise in the data, a refined and more pure ADHD group was identified by including those with only a diagnosis of ADHD (removing dual-diagnoses). A two-tailed t-test showed that there was not a significant difference in RDS scores between the ADHD (n=21) and TBI (n=15) groups, $t(34) = -1.63$, $p = .11$. Although not a significant difference, the TBI group demonstrated a trend toward better performance/higher scores ($M = 8.67$, $SD = 2.06$), than the ADHD group ($M = 7.71$, $SD = 1.45$; See Figure 3). When controlling for age and IQ, partial correlations revealed an even weaker relationship suggesting that the variability in RDS scores was not significantly related to a diagnosis of ADHD or TBI once age ($r = .01$ $p = .942$) or IQ ($r = .09$, $p = .718$) was taken into consideration.

Crosstab analysis with the ADHD pure/TBI mixed sample and RDS score showed that 28.57% of the ADHD group and 13.3% of the TBI group scored below the ≤ 6 RDS cutoff score, a significant difference as shown by a chi-square analysis ($p = .005$; See Figure 4).

Post hoc and exploratory considerations resulted in t-test analyses using a pure sample for both ADHD (n =21) and TBI (n = 8) groups to remove as much noise from the data as possible. A t-test analysis revealed that there was still a nonsignificant difference in RDS scores between the two groups, $t(27) = -.89$, $p = .381$. As in the prior analyses between ADHD and TBI, although there was not a significant difference, the

TBI group trended towards better performance/higher scores (*M* = 8.38, *SD* = 2.50) than the ADHD group (*M* = 7.71, *SD* = 1.45). Partial correlations again showed weakened relationships between RDS score and TBI or ADHD diagnosis when controlling for age ($r = -.12$, $p = .53$) and IQ ($r = .14$, $p = .60$). A crosstab analysis with the pure ADHD and TBI samples with RDS score revealed that 29% of the ADHD group and 25% of the TBI group gave a failing performance on the RDS. Due to the small sample size, a chi square analysis could not be run to determine the significance of this difference.

Hypothesis 4

Similar to the ADHD vs TBI analysis, the first ADHD vs LD analysis allowed for dual-diagnoses in each of the groups. A two-tailed t-test showed a nonsignificant difference between the ADHD (n=68) and LD (n=23) group RDS scores, $t(89) = 1.30$, $p = .20$. Although there was not a significant difference in the scores, the ADHD group had higher RDS scores, or trended towards performing better (*M*= 7.57, *SD* = 1.60), than the LD group (*M* = 7.09, *SD* = 1.41). Again, dual-diagnoses can create noise in the sample, so the ADHD sample was filtered to exclude all dual-diagnoses while the LD sample maintained them in the second analysis. A two-tailed t-test revealed that the difference in RDS scores between the ADHD (n=21) and LD (n=23) groups was still nonsignificant, $t(42) = 1.45$, $p = .15$. The ADHD group trended toward better/higher RDS scores (*M* = 7.71, *SD* = 1.45) than the LD group (*M* = 7.09, *SD* = 1.4; See Figure 5). Partial correlations revealed that variability in RDS scores across ADHD and LD diagnoses were not significant when controlling for age ($r = -.28$, $p = .07$) and IQ ($r = -.35$, $p = .10$), but trended towards significance.

Crosstab analysis was also run for the ADHD pure vs LD mixed comparison group. Because the ADHD pure group did not change between the TBI and LD comparisons, there was still a 28% failure among those participants, while there was a 39% failure among the LD group, a difference found to be significant by chi square analysis (p = .003; See Figure 4).

As with the ADHD/TBI analyses, a post hoc exploratory t-test was run using pure data, or no dual-diagnoses, with the ADHD (n = 21) and LD (13) samples. It was revealed that there was no significant difference on the RDS performance between the ADHD and LD groups, $t(32) = .68$, $p = .51$. The ADHD group trended towards better performance/higher scores (M = 7.71, SD = 1.45) than the LD group (M = 7.38, SD = 1.26). When controlling for variables using partial correlation, a weak relationship was maintained between RDS scores and a diagnosis of ADHD or LD for age (r = -.09, p = .63) and IQ (r = -.28, p = .27) covariates. Crosstab analysis with ADHD and LD diagnoses with RDS scores showed that 29% of the ADHD group and 31% of the LD group scores below the ≤ 6 RDS cutoff score. Similar to the ADHD/TBI post hoc analysis, the sample size was too small to determine the significance of the difference between failure rates using a chi-square test.

Hypothesis 5

There was no significant correlation between RDS (M = 7.80, SD = 1.99) and CVLT (M = 13.56, SD = 2.09) scores, r = .16, p = .31.

DISCUSSION

As expected, those diagnosed with ADHD displayed worse performance than those diagnosed with TBI. Although the difference was not significant regarding RDS score means, there was a highly significant difference in the failure rates between the groups, with the ADHD group having a higher rate. Unexpectedly, however, the LD group showed worse performance than the ADHD group. As with the ADHD/TBI comparison, the difference between group means was not significant, but differences in the failure rates between the LD and ADHD groups was highly significant. Overall, three of the four hypotheses were supported, as there were significant correlations between RDS score & age and RDS score & IQ. The failure rates seen in the ADHD and TBI groups resembled the failure rates presented in the literature, while the LD failure rate was higher than would be expected from past research (See Table 1).

Children have been found capable of deception at ages as young as 6, with multiple potential motivations for such poor performance as compensation, prescription of medication, academic aid, lessened academic and athletic responsibility, and/or parental or lawyer coaching. Such incentives may be obvious in certain circumstances, particularly in cases of litigation, but they are not always (DeRight & Carone, 2015; Kirkwood, 2015; Kirkwood, Kirk, 2010; Kirkwood, Kirk, Blaha, & Wilson, 2010).

While there are potential outside motivations for those being assessed for ADHD and TBI to purposefully perform poorly on neuropsychological tests, ADHD motivations and outcomes theoretically appear to be long term, relatively speaking. Unless there is severe brain damage associated with the brain injury and given the overall nature of the two disorders, a person would only be exempt from their normal activities or receive accommodations for TBI for a limited amount of time. Essentially, there appears to be a greater long-term incentive to receive a diagnosis of ADHD than to perform poorly given a diagnosis of TBI. Such a hypothesis would support the findings of the present study, where ADHD diagnoses showed a significantly higher failure rate than those with a TBI diagnosis, potentially the result of greater incentives.

As with ADHD and TBI, it can be said that a diagnosis of ADHD would yield greater incentives, particularly prescription of medication, than that of LD (Sullivan, May & Galbally, 2007). The unexpected failure rates between the ADHD and LD groups, where LD had a significantly higher failure rate than ADHD, could be due to multiple factors including the small sample size used and the inclusion of those with other diagnoses for the LD group. The difference seen could also be the result of additional parental coaching that may have been coincidentally high for the particular LD sample used here. Although the LD group's higher failure rate was surprising, it still demonstrates a difference in performance that may be present between the two academic-related disorders, a discrepancy that is debated (Ernst, Kneavel, Gallo, Moylan, & Brosof, 2014; Sullivan, May, Galbally, 2007).

As the comparison between failure rates in the ADHD vs LD analysis was unexpected (LD had higher rate than ADHD), so too was the correlation between RDS and CVLT scores. The two different measures have both been shown to effectively measure effort in children (Baker, Connery, Kirk, & Kirkwood, 2013; Brooks & Ploetz, 2015; Kirkwood, Hargrave, & Kirk, 2011; Lichtenstein, Erdodi, & Linnea, 2017; Schwartz, et al., 2016), meaning the scores should be correlated. It is possible that there was a nonsignificant correlation as a result of the small number of people with CVLT data to compare to RDS scores. Another possible explanation for the nonsignificant correlation could be related to the fact that the RDS and CVLT-C are both embedded PVTs, which can have a greater chance of yielding false positives (DeRight & Carone, 2015). Perhaps either the RDS or CVLT scores did have a high number of false-positives and the scores could have been correlated if RDS data had been compared with standalone PVT scores, like the TOMM, instead.

Limitations

One of the limitations facing the present study was the overall retrospective nature of the sample used. By conducting such a study, we were only able to work with the data that was available rather than designing our own study and recruiting participants to collect data for the specific analysis areas desired. The sample used was a clinical one, meaning that participants were not recruited on the basis that they were referred for neuropsychological testing for a single diagnosis; there were many dual diagnoses present in the sample, creating a potential confound and noise in each additional diagnosis. Using a retrospective data set and having certain data available also meant that there was a limited amount of PVT data to use in comparisons with RDS. While some participants had CVLT scores, they were limited in quantity (n=40) compared to the overall sample size with RDS scores (n=185). Ideally, every participant would have had scores for the TOMM, MSVT, or both, in addition to their RDS scores to compare their performance across effort measures.

As a result of the retrospective sample and only certain data being available, the sample sizes were limited for the actual comparison analyses of RDS performance between the different diagnoses. The number of dual diagnoses present due to the data being clinical in nature also contributed to the limited sample size, as these cases were removed wherever possible without making samples too small; we removed the dual-diagnoses for the ADHD group, but not LD or TBI as the samples would have hovered near ten participants.

Another of the limitations also stems from the retrospective data set, as timing of diagnoses were not readily made clear; it was not always known if a diagnosis had been made prior to the test scores being analyzed, or if the diagnosis was made as a result of the test scores in question.

Future Directions

Future research should look to replicate the present study, but with a larger sample of ‘pure’ diagnoses, rather than the small sample with dual-diagnoses discussed here. This would help to have a cleaner and more generalizable sample and set of results with which to discuss PVT performance and failure rates across different diagnoses. It would also be beneficial to examine performance on multiple PVTs, both embedded and standalone since embedded PVTs have the potential for a high false-positive rate (DeRight & Carone, 2015), for individual diagnoses, as that is the set up for an actual test battery as it would be administered to someone being assessed for a potential diagnosis or dysfunction. Important to note about the present research is that no control, or non-clinical, group was used with which to compare RDS scores. Future research should aim to include control groups as comparison when using clinical samples.

More generally looking at future research utilizing the RDS as a measure of effort, a computerized version of digit span administration may be considered. Woods, et al. (2011) found that a computerized version of the digit span test showed greater consistency during delivery of the numbers, allowed for randomized digit lists, and allowed for important information to be gathered by looking at the variability in the performance near the maximum bound of ability when a two-error termination was not used. Overall, there was a greater reliability and precision with the computerized form of digit span that could be utilized in future research for calculating RDS when measuring effort for comparison with the traditional voice-read digit span or other PVTs.

In addition to new methods of administration of the digit span test to calculate RDS scores, future research can consider looking into new measures of effort. Egeland and Langfjaeran (2007) used the Stroop interference scores and the Trail Making Test-ratio in an attempt to differentiate malingerers from those with actual brain dysfunction. The two measures on their own were not sufficient in terms of specificity and sensitivity (Egeland & Langfjaeran, 2007), but have potential to be a viable measure of effort in combination with other PVTs should future research explore this area.

As children are capable of deception and incentives may not be obvious (DeRight & Carone, 2015; Kirkwood, 2015; Kirkwood, Kirk, 2010; Kirkwood, Kirk, Blaha, & Wilson, 2010), the field of pediatric neuropsychology should continue to question the potential motivations for a poor performance in the future. Test administrators should consider the referral source and what could be gained from receiving particular diagnoses when examining testing scores and attempting to determine potential non-credible performance.

Conclusion

In the present study, a significant difference was found between the failure rates across diagnoses on the RDS, where LD had the highest failure rate followed by ADHD and TBI. The RDS is one such PVT that can be used to measure level of effort. It is important to note, however, that there is a need for a combination of PVT scores in order to absolutely determine if a person is feigning their symptoms and is purposefully putting forth less than their best effort. The use of PVTs in pediatric realms is essential for ensuring valid testing and future research in the area should expand on the literature regarding failure rates across both diagnoses and tests.

ACKNOWLEDGEMENTS

1. Thank you to all participants
2. Thank you to my advisor, Cay Anderson-Hanley, PhD. for her continuous support and help throughout the length of my thesis.
3. Thank you to Dr. Mina Dunnam, Dr. Paula Zuffante, and Dr. Marie McCabe for providing all of the data used in this study and offering help and suggestions throughout the research process.
4. Thank you to Dr. Mike McCann for providing feedback and aiding in the poster presentation of this thesis at the annual meeting of the American Academy of Clinical Neuropsychology on June 8th in Boston, MA.
5. A version of these results will be presented at the annual meeting of the American Academy of Clinical Neuropsychology on June 8th in Boston, MA.

REFERENCES

Araujo, G.C., et al. (2014). The relationship between suboptimal effort and post-concussion symptoms in children and adolescents with mild traumatic brain injury. *The Clinical Neuropsychologist, 28,* 786-801.

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

Baker, D.A., Connery, A.K., Kirk, J.W., & Kirkwood, M.W. (2013). Embedded performance validity indicators within the California Verbal Learning Test, Children's version. *The Clinical Neuropsychologist, 28,* 116-127.

Brooks, B.L. & Ploetz, D.M. (2015). Embedded performance validity on the CVLT-C for youth with neurological disorders. *Archives of Clinical Neuropsychology, 30,* 200-206.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1994). *CVLT-C: California Verbal Learning Test*.

DeRight, J. & Carone, D. A. (2015). Assessment of effort in children: A systematic review. *Child Neuropsychology, 21,* 1-24.

Egeland, J. & Langfjaeran, T. (2007). Differentiating malingering from genuine cognitive dysfunction using the Trail Making Test-ratio and Stroop Interference scores. *Journal of Applied Neuropsychology, 14,* 113-119.

Ernst, W., Kneavel, M., Gallo, A., Moylan, R., & Brosof, A. (2014). C-89 reliable digit span performance in children referred primarily due to academic problems. *Archives of Clinical Neuropsychology, 29,* 606.

Etherton, J.L., Bianchini, K.J., Ciota, M.A., & Greve, K.W. (2005). Reliable digit span is unaffected by laboratory-induced pain. *Sage Journals, 12,* 101-106.

Etherton, J.L., Bianchini, K.J., Greve, K.W., & Heinly, M.T. (2005). Sensitivity and specificity of reliable digit span in malingered pain-related disability. *Assessment, 12,* 130-136.

Gidley Larson, J.C., et al. (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,* 192-199.

Institute of Medicine. (2015). Cognitive tests and performance validity tests. In *Psychological Testing in the Service of Disability Determination* (5). Retrieved from https://www.nap.edu/read/21704/chapter/7

Jasinski, L.J., Berry, D.T.R., Shandera, A.L., & Clark, J.A. (2011). Use of the Weschler Adult Intelligence Scale Digit Span subtest of malingering detection: A meta-analytic review. *Journal of Clinical and Experimental Neuropsychology, 33,* 300-314.

Kirkwood, M.W. (2015). Pediatric validity assessment. *NeuroRehabilitation, 36,* 439-450.

Kirkwood, M.W., Hargrave, D.D. & Kirk, J.W. (2011). The value of the WISC-IV digit span subtest in detecting non-credible performance during pediatric neuropsychological examinations. *Archives of Clinical Neuropsychology, 26,* 377-384.

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

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

Kirkwood, M.W., Yeates, K.O., Randolph, C., & Kirk, J.W. (2012). The implications of symptom validity test failure for ability-based test performance in a pediatric sample. *Psychological Assessment, 24,* 36-45.

Lezak, M.D., Howieson, D.B, & Loring, D.W. (2004). Neuropsychological assessment. New York: Oxford University Press, Inc.

Lichtenstein, J.D., Erdodi, L.A., & Linnea, K.S. (2017). Introducing a forced-choice recognition task to the California Verbal Learning Test - Children's version. *Child Neuropsychology, 23,* 284-299.

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. *Sage Journals, 9,* 301-308.

Schroder, R.W., Twumasi-Ankrah, P., Baade, L.E. & Marshall, P.S. (2011). Reliable Digit Span: A systematic review and cross-validation study. *Sage,* 1-10. DOI: 10.1177/1073191111428764

Schwartz, E.S., Erdodi, L., Roderiguez, N., Ghosh, J.J., Curtain, J.R., Flashman, L.A., & Roth, R.M. (2016). CVLT-II forced choice recognition trial as an embedded validity indicator: A systematic review of the evidence. *Journal of the International Neuropsychological Society, 22,* 851-858.

Sullivan, B.K., May, K., & Galbally, L. (2007). Symptom exaggeration by college adults in attention-deficit hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology, 14,* 189-207.

Wechsler, D. (2014). Wechsler adult intelligence scale–fourth edition (WAIS–IV).

Wechsler, D. (2003). *Wechsler intelligence scale for children-WISC-IV*. Psychological Corporation.

Welsh, A.J., Bender, A.H., Whitman, L.A., Vasserman, M., & MacAllister, W.S. (2012). Clinical utility of reliable digit span in assessing effort in children and adolescents with epilepsy. *Archives of Clinical Neuropsychology, 27,* 735-741.

Woods, D. L., et al. (2011). Improving digit span assessment of short-term verbal memory. *Journal of Clinical Experimental Neuropsychology, 33,* 101-111.

Figure 1. Graph showing the positive correlation between RDS score and full-scale IQ score, $r = .47$, $p < .001$, with children having higher IQ scores scoring higher on the RDS.

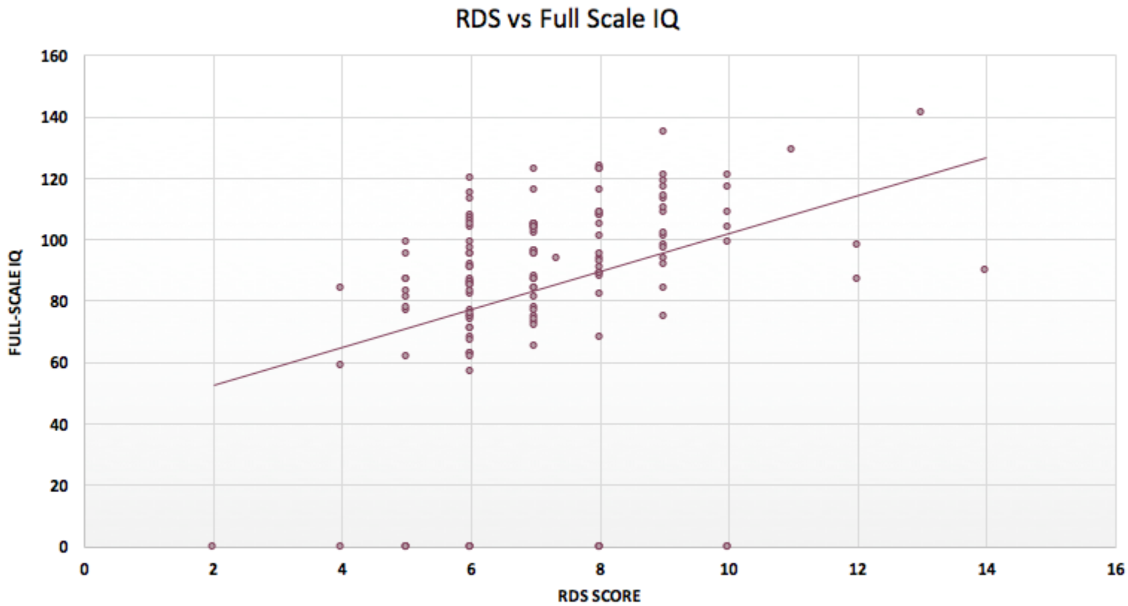

Figure 2. Graph showing the positive correlation between RDS score and age, $r = .39$, $p < .001$, with older children scoring higher than younger children.

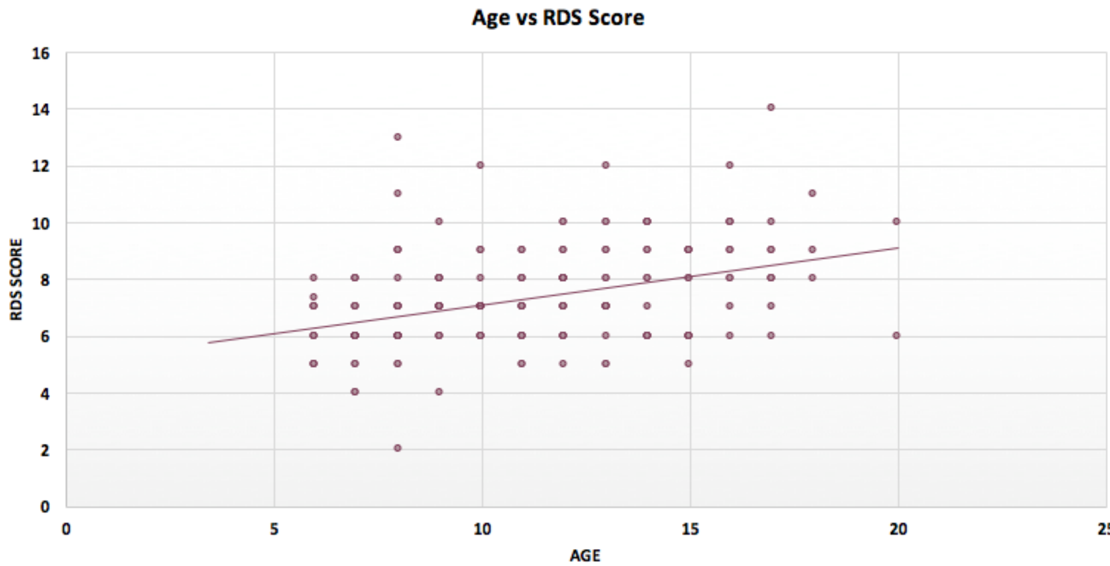

Figure 3. Graph showing the average RDS score for TBI, ADHD pure, and ADHD mixed samples, where there was a significant difference in means between TBI and ADHD mixed, but not TBI and ADHD pure. Error bars show RDS score standard deviations for each sample.

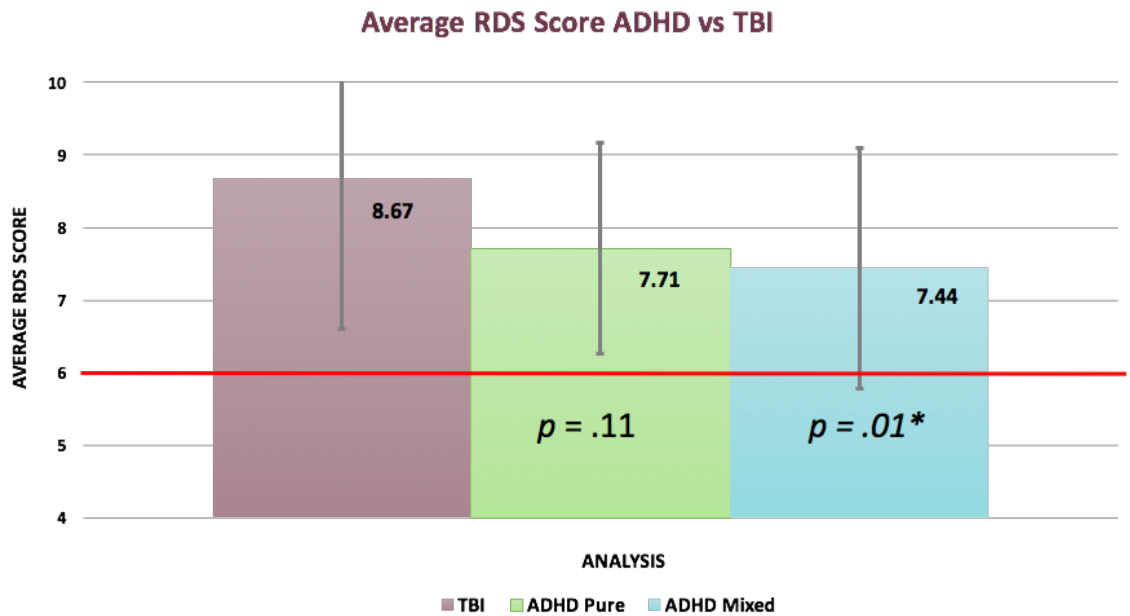

Figure 4. Graph showing the failure rates of LD, ADHD pure, and TBI, where failure rates differed significantly between ADHD & LD and ADHD & TBI.

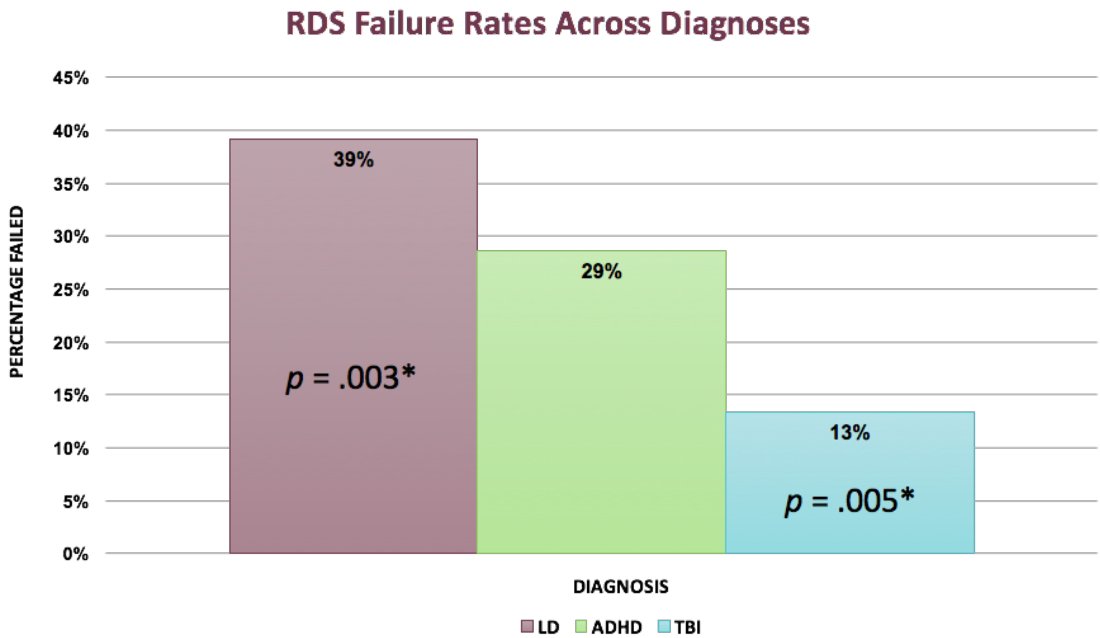

Figure 5. Graph showing the average RDS score for the LD, ADHD mixed, and ADHD pure samples, where there was a nonsignificant difference between the means for both the ADHD pure vs LD and ADHD mixed vs LD analyses. Error bars show standard deviations for each sample.

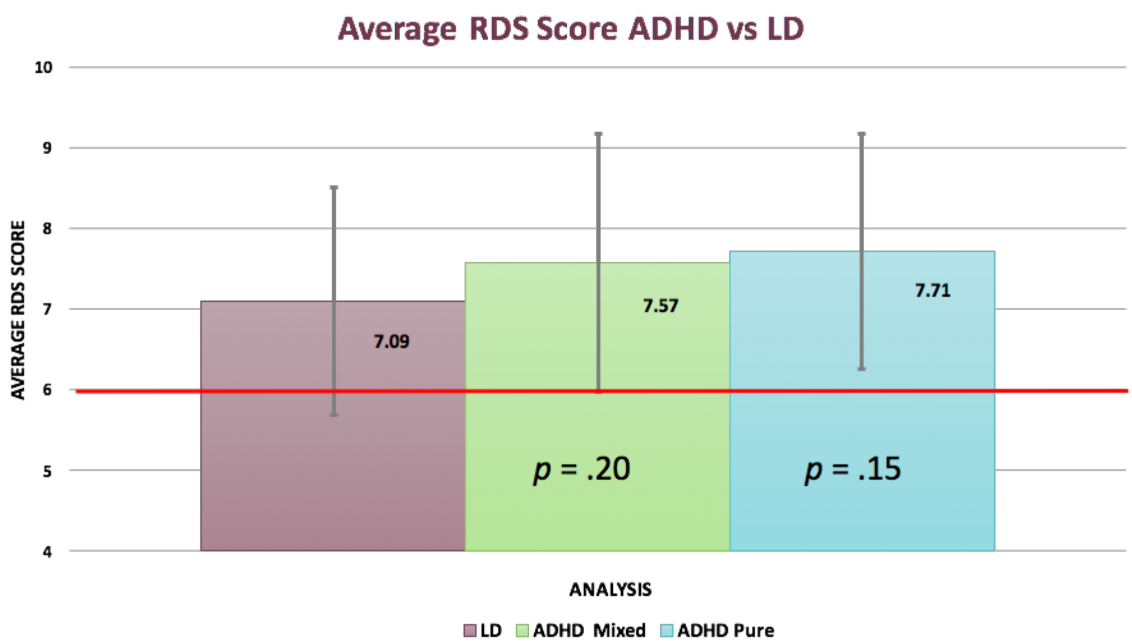

Table 1. Table showing the range in PVT failure rates for different diagnoses in the literature and the present study (1 = Sullivan, et al., 2007; 2 = Arujo, et al., 2014; DeRight & Carone, 2015; Gidley Larson, et al., 2015; Kirkwood, 2015; Kirkwood, et al., 2012).

		% Failing PVT		
		LD	ADHD	TBI
Literature	Mixed	9%[1]	9%[1]	
	Pure	15%[1]	48%[1]	12-20%[2]
Current Sample	Mixed	39%		13%
	Pure	31%	29%	25%