WAIS®-5 and WMS®-5
Advanced Clinical Solutions

ACS

Performance Validity and Word Choice Test Manual

James A. Holdnack

Pearson

When referencing the WAIS-5 and WMS-5 Advanced Clinical Solutions, please cite according to APA style as follows:

NCS Pearson. (2026). *WAIS-5 and WMS-5 Advanced Clinical Solutions*.

When referencing this manual specifically, please cite according to APA style as follows:

Holdnack, J. (2026). *WAIS-5 and WMS-5 Advanced Clinical Solutions: Performance validity and Word Choice manual*. NCS Pearson.

Pearson

For inquiries or reordering:
800.627.7271
www.PearsonAssessments.com

NCS Pearson, Inc. 5601 Green Valley Drive Bloomington, MN 55437

Produced in the United States of America.

v.A A103000445132

Table of Contents

List of Tables

List of Figures

1

Introduction

The *Wechsler Adult Intelligence Scale* (5th ed.; WAIS-5; Wechsler, 2024) and the *Wechsler Memory Scale* (5th ed.; WMS-5; Wechsler, 2025) are among the most widely used adult cognitive and memory measures (LaDuke et al., 2018; Rabin et al., 2016). The *WAIS-5 and WMS-5 Advanced Clinical Solutions* (ACS) comprises various methods to aid clinicians in diagnostic, clinical, and intervention decisions. It provides embedded and stand-alone performance validity measures, estimates of premorbid cognitive and memory functioning, and demographically referenced scores. The ACS can be used in assessments to evaluate performance validity, determine decline in cognitive functioning, and estimate premorbid general ability and memory functioning. The scores, reports, and analyses included in the ACS are provided in separate manuals, each dedicated to one of the methods or clinical solutions. This organization supports the idea that most clinicians will use only specific components of the ACS during a given assessment. Most clinicians will choose one or two functions of the ACS to enhance the clinical sensitivity and utility of a standard WAIS-5 and/or WMS-5 assessment.

This manual describes the development, validity, and interpretation of the WAIS-5 and WMS-5 ACS performance validity measures: the embedded WAIS-5 and WMS-5 scores and the Word Choice Test (WCT). It also details the appropriate use of the specified procedures, along with caveats on their application and interpretation of results. The performance validity score summary and analyses are provided through Q-global, Pearson's online scoring and reporting platform. All measures can also be hand-scored. At least two scores from the following options are required for the analyses: WCT, WAIS-5 Reliable Digit Span (RDS), WMS-5 Logical Memory 2 (LM 2) Recognition, WMS-5 Visual Reproduction 2 (VR 2) Recognition, WMS-5 Verbal Paired Associates 2 (VPA 2) Recognition, or WMS-5 Designs 2 (DE 2) Recognition.

If the WCT is used as one of the performance validity measures, the WCT materials are required for obtaining these scores. The materials available for the ACS Performance Validity and Word Choice Test are listed in Table 1.1, including whether materials are available in print or digitally in the Q-global Resource Library. Details on administering the WCT and obtaining the embedded WAIS-5 and WMS-5 performance validity scores are provided in Chapter 2. The ACS does not provide administration materials, manuals, or scoring for the primary or secondary WAIS-5 subtests or the primary WMS-5 subtests. The WCT can be scored manually or within the Q-global software.

Table 1.1 Materials Available for the ACS Performance Validity and Word Choice Test

Component	Print	Digital
Performance Validity and Word Choice Test Manual	✓	✓
Performance Validity/Word Choice Test Record Form	✓	
Word Choice Test Stimulus Book	✓	✓
Word Choice Test Card	✓	✓

The ACS also provides education-referenced and demographically referenced scores to enable clinicians to refine hypotheses about the degree to which a specific score is unexpected when compared to individuals with similar background characteristics (e.g., education level). These scores have specific applications and should be used only to answer appropriate clinical questions. Detailed information on the appropriate use of education-referenced and demographically referenced scores is provided in the *WAIS-5 and WMS-5 ACS Demographically Referenced Scores Manual* (Holdnack, 2026).

The *Test of Premorbid Functioning* (2nd ed.; TOPF-2; Holdnack, in press) is included in the ACS to provide estimates of premorbid intellectual and memory ability. The TOPF-2 Manual describes how to predict premorbid abilities by applying a regression equation using the examinee's ability to read words that have atypical grapheme-to-phoneme translations and/or the examinee's demographic characteristics. Demographic information can be combined with TOPF-2 scores to increase the accuracy of the estimated scores. Comprehensive guidance on the appropriate use of TOPF-2 and premorbid cognitive estimation is provided in the *WAIS-5 and WMS-5 ACS TOPF-2 Manual*.

Assessment of Performance Validity

Psychological tests are administered and interpreted with an understanding that the test validly measures the construct of interest, such as memory, reading, or psychiatric symptomatology (Larrabee, 2014). That is, the test score can be interpreted meaningfully. For example, a low memory test score should indicate limited memory ability or a memory problem. However, examinee factors can influence performance and impact the validity of results. These include situational factors, such as fatigue, poor cooperation, or response bias. Response bias occurs when an examinee responds in a way that misrepresents performance capacities, either positively or negatively (Sweet et al., 2021). Positive response bias involves reducing or minimizing symptomatology or enhancing performance (e.g., reporting low severity of symptoms), while negative response bias involves increasing symptomatology or lowering performance (e.g., reporting severe impairment or performing below expected performance for the nature and severity of the presenting problem). Response bias may relate to evaluation outcomes; for example, an examinee may attempt to conceal or minimize psychological symptoms to support a claim of psychological well-being, even when problems exist. Conversely, an examinee may attempt to create the impression of disability when it is not present or is less severe than performance indicates. The likelihood and nature of response bias are directly related to the context of the evaluation (e.g., prehiring police officer exam, Social Security disability evaluation; Holdnack et al., 2013). When response bias is suspected, clinicians must interpret test results with caution because obtained scores may be discordant with actual ability or symptomatology.

The concept of validity testing has been extensively researched for decades (Sweet, 2025), and a full review is beyond the scope of this manual. Validity testing evaluates responses to provide information on response bias and includes both symptom validity and performance validity. Symptom validity reflects the accuracy of an examinee's self-reported symptoms on personality or psychological measures (e.g., severity of depression or anxiety; Holdnack et al., 2013;

Larrabee, 2014; Sherman et al., 2020). Performance validity reflects whether an examinee's test performance on direct measures of ability accurately represents their true ability. While these concepts are related and are often applied simultaneously in psychological evaluations, they are distinct. For example, an examinee may fail symptom validity tests (SVTs) but pass performance validity tests (PVTs), indicating that self-reports may not be accurate, but performance-based measures are valid (Copeland et al., 2016). In the ACS, performance validity is the primary focus, while symptom validity is not assessed.

Previously, the ACS for WAIS–IV and WMS–IV referred to validity testing as "suboptimal effort." However, the growing body of literature has shifted away from the term "effort" to "performance validity" (Sweet et al., 2021). The term "effort" implies that low scores on performance validity measures are solely due to the examinee's effort; however, there are multiple factors that can influence performance and the validity of obtained scores and interpretations. The WAIS-5 and WMS-5 ACS uses the term "performance validity" to align with current clinical practice and research, providing a framework for interpreting the performance validity of cognitive test scores.

Historically, psychologists have used measures of symptom validity to assess reported psychiatric symptoms and cognitive impairments; however, performance validity has increasingly become a standard part of psychological and neuropsychological assessments. Demands from the legal system for assessment of validity have increased the need for psychologists to detect feigned brain injury (Sweet, 2025). In forensic neuropsychology, performance validity assessment is considered essential (Clem & Schroeder, 2025), and most experts consider it a mandatory part of a comprehensive forensic neuropsychological assessment (Martin et al., 2025). Hirst et al. (2017) found that more than 70% of surveyed neuropsychologists believe PVTs should be included in every neuropsychological evaluation. Additionally, nearly 90% of respondents indicated they received training on the administration and interpretation of PVTs during graduate school. Furthermore, more than 90% of general clinicians and those who perform forensic evaluations agree that there is sufficient empirical evidence for the standard use of PVTs in neuropsychological assessments (Sweet, 2025). The use of psychometric indicators, such as test scores, rather than clinical judgment, is increasingly recognized as the most valid method for determining performance validity (Sweet et al., 2021).

The American Academy of Clinical Neuropsychology (AACN) 2021 consensus paper reaffirmed the necessity of validating cognitive test performance in neuropsychological assessments (Sweet et al., 2021), as did the position paper of the National Academy of Neuropsychology (NAN; Bush et al., 2005). Interest in and research on PVTs have grown steadily over the years in both general clinical practice and clinical neuropsychology. While high rates of feigned cognitive impairment have long been a concern in medical-legal and forensic evaluations (Larrabee, 2014; Mittenberg et al., 2002; Young, 2015), there is increasing awareness of the potential for invalid test results across settings. Feigned responding in Social Security disability evaluations carries a significant societal cost and limits resources for individuals who genuinely require assistance (Chafetz & Underhill, 2013). Furthermore, high base rates of suspected invalid performance are observed in adult attention-deficit/hyperactivity disorder (ADHD) evaluations (Shura & Armistead-Jehle, 2024). Research involving healthy college students suggests that insufficient effort on cognitive testing occurs more frequently than expected (Ross et al., 2016). College students undergoing psychoeducational evaluations for accommodations or medication exhibit higher rates of noncredible test performance (Pella et al., 2012). Performance validity has also gained importance in the assessment of children (Kirkwood, 2015; Kirkwood et al., 2011). Conceptually, the presence of external incentives increases the likelihood of noncredible or invalid test performance (Sherman et al., 2020). Therefore, the use of PVTs has expanded beyond medical-legal and forensic settings to include nearly all assessment settings.

PVTs are designed to identify potential exaggeration or feigning of cognitive symptomatology. However, low scores on PVTs do not necessarily indicate intentional malingering because other response styles can also result in low scores on these measures (Sweet et al., 2021). Experts typically avoid the term "malingering" when discussing PVT results because it implies intent on the part of the examinee. Instead, experts recommend using statements such as "results of the assessment suggest invalid cognitive test performance" (Sherman et al., 2020). The process of detecting malingering involves evaluating multiple dimensions of underlying behavior (Sweet et al., 2021). Thus, the primary goal of measuring performance validity is to evaluate whether the examinee's test performance is consistent with their injury or clinical condition and to determine if test scores can be validly interpreted.

There is no consensus on whether clinicians should inform examinees that tests measuring performance validity will be used during the assessment. AACN guidelines state, "PVT effectiveness requires that examinees are unaware of their purposes and outcomes" (Sweet et al., 2021), indicating that examinees should not be explicitly instructed about the use of these measures. Among general clinical neuropsychologists, most respondents (59%) rarely or never inform examinees about the use of PVTs, though roughly one-third inform examinees that validity indicators are included in the evaluation (Hirst et al., 2017). Most clinicians insert PVTs throughout the battery rather than at the beginning or end (Hirst et al., 2017) to assess validity throughout the testing session. For a comprehensive review of issues in assessing performance validity, see the AACN and NAN position papers (Bush et al., 2005; Sweet et al., 2021).

Approaches to Measuring Performance Validity

External or stand-alone PVTs are designed to identify noncredible performance; they are not intended to directly measure other cognitive constructs (e.g., memory). External validity measures are the most prevalent type of PVT used by clinicians and experts (Martin et al., 2025). Clinical and typically developing populations usually obtain high scores on these measures, so their psychometric properties are highly skewed, producing nonnormal distributions. Additionally, external PVTs are not designed to measure a range of ability; rather, they are designed to identify response patterns that are unusual relative to individuals with true cognitive impairment (Sweet et al., 2021). Therefore, the cutoff scores used for determining performance validity differ from the normative scores used for traditional cognitive tests. Many external validity measures use a forced-choice memory paradigm. Forced-choice tasks appear challenging, but most examinees, including clinical populations, achieve high scores. For these tasks, the examinee is presented with multiple target stimuli. After presentation, the examinee is shown pairs of stimuli—one target and one distracter—and asked to choose the target stimuli shown previously. Forced-choice tests using words compared to other stimuli (e.g., visual images) are the most researched and widely used (Schroeder & Martin, 2021).

Embedded PVTs are procedures or scores that are derived from a test not specifically designed to measure performance validity but that have psychometric properties that may be amenable to use in this capacity (e.g., highly skewed distributions in normative and clinical populations). One of the most well-known and researched embedded PVTs is RDS, which was originally developed for the WAIS-R and has been used in subsequent revisions of the WAIS (Schroeder et al., 2012). Using embedded PVTs within a standard psychological battery is an effective model for simultaneously assessing performance validity and providing assessment of important cognitive constructs. Embedded validity indicators also provide a means of evaluating validity throughout the course of the assessment, not only at a single point in the assessment. Experts use embedded measures less frequently as PVTs than external PVTs (Martin et al., 2025).

The ACS PVTs consist of one external/stand-alone validity measure, the WCT, and five embedded validity indicators: RDS and the recognition conditions from LM 2, VPA 2, VR 2, and DE 2. DE 2 Recognition is the only new test included in the ACS performance validity solution; all other measures were included in the ACS for WAIS–IV and WMS–IV. Of the four embedded validity indicators that were retained, the most substantive change from the previous edition is to VR 2 Recognition. In the previous edition, there were seven multiple-choice items, each with one correct response and five distracters. The current version has 18 items with two responses: one correct response and one distracter. Due to new stories for Logical Memory, LM 2 Recognition has more items in the current ACS than in the previous version. VPA 2 Recognition is the same task; however, examinees receive fewer learning trials, which may impact cutoff scores. The cutoff score development and multivariate base-rate approach have not changed from the previous edition.

The WCT is increasingly used in research and clinical practice (Schroeder & Martin, 2021). Research on the WCT as a stand-alone measure in the ACS for WAIS–IV and WMS–IV demonstrates very good sensitivity and specificity compared to other PVTs (Schroeder et al., 2018). It is effective in classifying mixed clinical samples with cognitive impairment (Bain & Soble, 2019), identifying simulators (An et al., 2019; Barhon et al., 2014), and classifying traumatic brain injury (Bashem et al., 2014). The WCT and the Recognition Memory Test (Warrington, 1984), another commonly used external PVT, have similar score distributions and achieve comparable results; however, specific test cutoff scores are required to achieve these results (Erdodi et al., 2014; Erdodi et al., 2017). Additionally, the WCT and the Test of Memory Malingering (TOMM; Tombaugh, 1996, 1997) show similar sensitivity and specificity for identifying invalid performance among college-age simulators (An et al., 2019). While most clinical populations obtain relatively high scores on the WCT, research findings are mixed regarding the impact of cognitive impairment on WCT, with some analyses suggesting lower scores and others finding no effect (Bain & Soble, 2019). Considerable research over the past decade shows good to very good sensitivity and specificity for identifying invalid performance, comparability to other external validity indicators, and good utility in multiple clinical and simulator groups.

Another line of research has focused on identifying an optimal cutoff score for the WCT, as many stand-alone PVTs offer a single valid/invalid cutoff score, while the WCT offers flexibility in cutoff scores based on examinee characteristics (NCS Pearson, 2009). Bernstein et al. (2021) conducted a meta-analysis of studies that used the WCT in a variety of clinical and nonclinical populations. Using multiple samples, they determined that a WCT score of ≤42 (out of a total of 50) optimized invalid/valid performance sensitivity and specificity. In two studies using different samples, cutoff scores of ≤45 (Erdodi et al., 2014) and ≤47 optimized classification accuracy (Erdodi et al., 2017). Tyson et al. (2025) sought to cross-validate the findings from the Bernstein et al. (2021) meta-analysis in an independent mixed clinical sample. When using the proposed cutoff score of ≤42, specificity was very high (>.97), but sensitivity was low (.30–.38); a cutoff score of ≤45 yielded the most optimal diagnostic statistics. However, the authors found very little difference in applying cutoff scores ranging from 43 to 47, suggesting that specifying a single cutoff score may not be necessary (Tyson et al., 2025). In a study including individuals with severe memory impairment, the WCT differentiated valid from invalid test performance using a cutoff score of ≤41, though low scores were observed in the severe memory impairment group. Alternate cutoff scores are likely needed when severe memory impairment is present (Neale et al., 2020). In the ACS for WAIS–IV and WMS–IV, cutoff scores were established at various base rates of occurrence in an overall clinical sample as well as in several low-ability groups (e.g., intellectual disability-mild severity [IDMI], WAIS–IV General Ability Index [GAI] ≤69). A cutoff score of ≤45 on the WCT corresponds to the 15% base rate, and a cutoff score of ≤42 is just under the 10% base rate (i.e., ≤43). Base rates of 15% and 10% in a mixed-diagnosis clinical group are recommended for general use; however, for individuals with severe cognitive impairment, the base rates for the IDMI or the GAI ≤69 comparison groups provide guidance for setting an appropriate cutoff score (NCS Pearson, 2009).

There has been significantly less research utilizing the embedded PVTs from the ACS for WAIS–IV and WMS–IV; however, the RDS has been researched the most. A study using RDS and LM 2 Recognition found these two embedded measures did not outperform the WCT in isolation or in combination (Bain & Soble, 2019). Schroeder et al. (2012) performed a comprehensive meta-analysis of RDS covering nearly two decades of research. The results indicate that the commonly used cutoff score of ≤7 yields high rates of false positives, and a cutoff score of ≤6 was more appropriate for general clinical populations, although it still yields a high number of false positives in individuals with significant cognitive impairment. This is consistent with other research, which found that RDS cutoff scores of ≤7 produce a high false positive rate in older adults (Zenisek et al., 2016). RDS was one of the best embedded performance validity measures when combined with an external validity indicator (Miele et al., 2012). New RDS measures that include the Digit Span Sequencing condition were useful for detecting invalid test performance (Reese et al., 2012; Resch et al., 2023). Research with RDS supports its use as a PVT; however, it may not work as expected in some populations, particularly when used in isolation.

Although other embedded ACS validity indicators are not heavily researched, existing research suggests that the embedded PVTs from the ACS for WAIS–IV and WMS–IV perform very well. Miller et al. (2011) used all four embedded PVTs and the WCT with a sample of individuals who had a history of traumatic brain injury (TBI) and simulators. The combined WCT and embedded PVTs were very effective in differentiating the groups. Henry (2023) observed a very high degree of accuracy in identifying invalid performance for three embedded measures—Logical Memory Delayed Recall (not in the ACS for WAIS–IV and WMS–IV), Logical Memory Delayed Recognition, and Visual Reproduction Recognition—in a sample of mild TBI litigants. Verbal Paired Associates 2 Recognition and Visual Reproduction 2 Recognition show good sensitivity and specificity as PVTs in a mixed clinical sample (Crisan et al., 2024). Soble et al. (2019) evaluated LM 2 Recognition as a PVT in a mixed clinical sample. Several individuals identified as having normal recognition scores using age-referenced cumulative percentages were also identified as having invalid performance when LM 2 Recognition was used as a single PVT. However, many examinees in the study were administered the Older Adult battery version of Logical Memory, which is not the version used in the ACS for WAIS–IV and WMS–IV to establish the recommended cutoffs (NCS Pearson, 2009). ACS for WAIS–IV and WMS–IV PVTs show acceptable false positive rates, consistent with other data sets (Larrabee, 2014).

There is substantial independent research support for using the ACS PVTs in clinical practice. There is a high degree of support for the WCT as a stand-alone/external PVT and the RDS, which has a long history of use in clinical practice. The other embedded measures are less studied, and rarely are all the measures used in a single study. Furthermore, the system has not been fully vetted in the research literature, including the establishment of cutoff scores suited to the specific characteristics of the sample under assessment.

If a high-ability college sample is used, a higher cutoff score may be required. Alternatively, for individuals with low scores on cognitive tests (e.g., GAI 69 or lower) or for individuals with low levels of education, a more restrictive cutoff score may be required. Not surprisingly, some studies report that significant cognitive impairment produces false positives, especially in the embedded measures. The ACS PVTs were designed to help clinicians consider the effect of various factors when selecting cutoff scores for these measures.

Attaining a low score on a single PVT is not unusual, especially with embedded measures, and clinical guidelines on the use of PVTs emphasize using two or more PVTs to optimize identification of valid or invalid performance (Bilder et al., 2014; Bush et al., 2005; Heilbronner et al., 2009; Soble et al., 2020; Sweet et al., 2021). Rhoads et al. (2021) evaluated the impact of single PVT failure on classification accuracy versus multiple PVTs and showed that using a single measure to classify invalid performance resulted in poor to acceptable classification accuracy, whereas using two measures created stable classifications. When using PVTs, it is important to use multiple measures and make judgments on the validity of the performance within the full context of the evaluation.

Population Comparisons

The context in which an evaluation occurs is an important consideration when using PVTs. External incentives, defined as perceived benefits obtained from poor performance, increase the potential for response bias (Larrabee, 2014; Sherman et al., 2020; Sweet et al., 2021). Examples of perceived benefit include obtaining a financial award, desired medication, accommodations, or reduced punishments such as in court sentencing. However, the presence of external gain does not necessarily indicate response bias, nor does the absence of apparent external gain eliminate the potential for response bias (Sherman et al., 2020; Sweet et al., 2021). Historically, identifying invalid performance has relied on comparing an examinee's obtained test score to an established cutoff score based on the performance in mixed-diagnosis clinical populations (Boone et al., 2002) and to specific clinical populations of interest, such as individuals with TBI (Boone et al., 2002; Greve et al., 2007). Sensitivity and specificity are subsequently determined by evaluating different cutoff scores across clinical samples, simulators, and suspected malingerers. Sensitivity and specificity are optimized when clinicians apply multiple validity indicators and require at least two scores to fall below the prescribed cutoff score. However, a single PVT score well below chance may indicate significant response bias (Holdnack et al., 2013; Larrabee, 2014; Sherman et al., 2020; Sweet et al., 2021).

PVTs are not standardized like traditional cognitive measures. Instead, cutoff scores are derived from the raw score distribution of a mixed-diagnosis clinical sample. Cutoff scores are based on base rates within the sample (e.g., the score at which 90% of known clinical cases perform better, corresponding to a 10% false positive rate). This model aids in interpreting low scores because such scores are considered atypical even among individuals with a clinical diagnosis. The research literature contains numerous studies using mixed-diagnosis clinical samples to identify false positive rates for various PVT cutoff scores. These data are important for understanding the operational characteristics of the tests. However, the composition of the clinical sample, including both diagnostic and demographic factors, can influence the results. For example, a meta-analysis of the WCT reported an optimal cutoff score of ≤42 (Bernstein et al., 2021). However, Tyson et al. (2025) found that a cutoff score of ≤45 was optimal, with cutoff scores between ≤47 and ≤43 also proving effective. These values align closely with cutoff scores reported in the ACS manual (NCS Pearson, 2009). This variability applies to all PVTs. For example, Deloria et al. (2021) examined several PVTs, such as the TOMM (Tombaugh, 1997), Dot Counting Test (Binks et al., 1997), Victoria Symptom Validity Test (Slick et al., 1997), and RDS, in various combinations. They found that more conservative cutoff scores were needed to achieve similar base rates compared to those previously established. However, having two or more PVT failures at the 10% base rate, a well-established criterion (Larrabee, 2014), demonstrated very good sensitivity and specificity. It is difficult to establish a single cutoff appropriate for every clinical group; therefore, applying cutoff scores tailored to a specific client population (e.g., individuals with very low cognitive ability or high education levels) can reduce false positives and improve sensitivity.

Research on specific clinical samples shows that high false-positive rates may occur when using established cutoff scores with individuals who have a history of intellectual disability, epilepsy, neurodegenerative disorders (McWhirter et al., 2020), schizophrenia/schizoaffective disorder (Hunt et al., 2014; McWhirter et al., 2020), moderate to severe brain injury (Finley et al., 2023; McWhirter et al., 2020), and mild cognitive impairment (McWhirter et al., 2020; Roor et al., 2024). Individuals diagnosed with Huntington's disease generally exhibit valid performance; those whose PVT indicated invalid performance had severe cognitive, adaptive, and motor impairments (Sieck et al., 2013). Low intellectual ability has been associated with higher rates of PVT failures (Chafetz et al., 2011; Finley et al., 2023); however, when individuals with low intellectual ability have an incentive to perform well (e.g., job seeking, child custody), they generally do not show inflated PVT failures (Chafetz et al., 2011). In a sample of individuals with a history of mild TBI, higher rates of invalid PVT performance were associated with litigation status (Meyers et al., 2011), and both litigation status and failed PVTs are associated with lower overall test scores (Meyers et al., 2011). In less cognitively impaired clinical populations, low PVT scores are less common. For example, there is no evidence that depression or anxiety increases the probability of invalid PVT findings (Sweet et al., 2021). While most individuals within a clinical population perform adequately on PVTs, those with more severe impairments, such as individuals with dementia or intellectual disability, may score in the invalid range of performance. Caution must be used when determining the validity of these scores (Milanovich et al., 1996).

The use of PVTs is most frequently associated with medical-legal and forensic cases; however, there is increased interest in incorporating PVTs into psychoeducational and ADHD evaluations, particularly when an external incentive is involved. College students with an external incentive for low cognitive test performance (e.g., academic accommodations, medication evaluations) demonstrate a substantially higher PVT failure rate than students with no obvious external incentive (Pella et al., 2012). Generally, lower test scores are observed in the presence of an external incentive (Pella et al., 2012). Failed PVT performance during adult ADHD evaluations is associated with significantly greater impairment on a continuous performance test compared to individuals with ADHD and psychiatric controls (Suhr et al., 2011). Results are mixed when using embedded measures in ADHD evaluations: Some embedded measures of sustained attention may produce elevated false-positive rates even among valid ADHD cases, whereas others do not show this effect (Finley et al., 2025). The choice of validity indicators may be important to consider in ADHD evaluations (Finley et al., 2025). PVTs can be helpful in adult psychoeducational and ADHD evaluations, especially when external incentives for low test scores are present.

The use of standard cutoff scores may yield higher-than-anticipated false positives in some groups. Cultural, demographic, and language factors may influence PVT performance, although Soble et al. (2023) found that demographic factors accounted for very little variance in PVT performance in a diverse outpatient clinical sample. English language proficiency may affect performance on PVTs (Robles et al., 2015; Roor et al., 2024). Daugherty et al. (2022) developed culturally appropriate cutoff scores for three embedded PVTs, including RDS. Research suggests that nonverbal PVTs may be more appropriate for Spanish-speaking populations (Burton et al., 2012). Examinees with low educational attainment are at increased risk for invalid PVT performance (Davis & Millis, 2014; Karr et al., 2024; Roor et al., 2024). False-positive rates on PVTs may also vary by race/ethnicity, although effect sizes are small (Karr et al., 2024). Low functional status and reduced engagement in community activities are associated with higher rates of invalid PVT findings (Davis & Millis, 2014; Lippa et al., 2014). Low verbal ability (Davis & Millis, 2014) and low overall cognitive ability may impact the sensitivity and specificity of embedded PVTs (Dunham et al., 2014). It is important to consider whether an examinee's English proficiency, education level, race/ethnicity, functional status, and general cognitive ability might have adversely affected PVT scores.

The level of performance on a PVT influences the confidence with which test scores can be interpreted as valid or invalid. Bigler (2014) observes that PVTs are easiest to interpret when performance falls below chance levels; however, when scores are below established cutoff scores but above chance level, interpretation becomes more complicated. Williams (2011) stated that while identifying "malingered cognitive deficits" is complex, using validated measures increases the likelihood of detection. Furthermore, Erdodi (2023) notes that while most patients do not perform at or below chance level on forced-choice tasks, even one error on specific forced-choice measures can suggest suspect performance. Clearly, performing below chance level is strongly indicative of response bias; however, obtaining a score below a cutoff score by one point might not indicate invalid overall performance. Using multiple measures and applying clinical judgment based on the specific characteristics and scores of an individual can improve the interpretation of PVT performance.

In addition to comparing an examinee to a mixed-diagnosis clinical population and to specific clinical populations, pure guessing and simulator samples provide additional information for determining performance invalidity. In a pure guessing condition, examinees respond to test items without ever seeing or hearing the test stimuli (Flowers et al., 2008). Flowers et al. (2008) found that examinees in a pure guessing condition performed at a level overlapping with low-functioning normal controls and simulators. Moreover, many simulators performed *more poorly* than the examinees in the pure guessing condition, indicating pronounced feigning of impairment. However, interpretation of these results warrants caution because apparent suboptimal effort (e.g., performing at chance level), at least on a single indicator, is observed in normal controls without an incentive to perform poorly (Flowers et al., 2008). Having a single score within the range of guessing may or may not indicate performance invalidity; however, performing below chance levels strongly suggests an atypical response style during that test.

There is a well-established body of literature demonstrating that healthy adults may obtain one or more low scores in a battery of cognitive tests (Brooks et al., 2008; Brooks et al., 2011). Moreover, individuals with low education levels or low intellectual ability often obtain multiple low scores in an otherwise valid assessment (Brooks et al., 2008). When age-scaled scores or other score referencing (e.g., education-referenced scores) is used, it is highly probable that a healthy person will obtain a score below a scaled score of 7 within a standard cognitive battery. This phenomenon is one reason the WAIS-5 and WMS-5 ACS PVTs do not use age-referenced scaled scores to determine performance validity. The multivariate base rate model is an effective means of reducing misdiagnosis of cognitive impairment (Brooks et al., 2011; Holdnack et al., 2017). This model provides a means for setting a base rate that considers the impact of using multiple test scores. For example, using a cutoff score of a 15% base rate across all five PVTs in the ACS for WAIS–IV and WMS–IV results in 22% of the overall clinical sample having at least one PVT score at or below that criterion. The multivariate base rate effect inflates a single instance of the cutoff score to a greater-than-expected false-positive rate. However, only 7% of the overall clinical sample, 9% of the TBI sample, and 13% of individuals with a GAI score of 69 or less (NCS Pearson, 2009) obtain two PVT scores at or below that criterion. The WAIS-5 and WMS-5 ACS PVTs apply the same approach. This enables clinicians to establish a general cutoff rule for multiple measures to estimate the overall false-positive rate for a variety of comparison groups.

User Responsibilities

The ethical and legal responsibilities of test users for the WAIS-5 and WMS-5 Advanced Clinical Solutions are consistent with those outlined in the WAIS-5 and WMS-5 administration and scoring manuals.

In light of the complexities of test administration, diagnosis, and assessment, WAIS-5 and WMS-5 Advanced Clinical Solutions users should have training and experience in administration and interpretation of standardized clinical instruments. They should also have training or experience testing individuals whose ages; linguistic backgrounds; and clinical, cultural, or educational histories are similar to those of the examinees they will be evaluating.

In most cases, WAIS-5 and WMS-5 ACS users should have completed formal graduate- or professional-level training in psychological assessment. Although a trained technician can administer the subtests of the WAIS-5 and WMS-5 and score the responses under supervision, results should always be interpreted by individuals with appropriate training in assessment. Furthermore, test users should follow the *Standards for Educational and Psychological Testing* (*Standards*; American Educational Research Association [AERA] et al., 2014).

In line with Pearson Clinical Assessment legal policies (please refer to the local Pearson Clinical Assessment website for the most up-to-date content), it is the responsibility of the test user to ensure that test materials, including completed assessment protocols, remain secure and are released only to professionals who will safeguard their proper use. Although review of test results with examinees or parents/caregivers is appropriate and encouraged when legally and ethically permitted, this review should not include disclosure or copying of test items, protocols, or other test materials that would compromise the security, validity, or value of the WAIS-5 and WMS-5 Advanced Clinical Solutions as a measurement tool. Under no circumstance should test materials be resold or displayed in locations where unqualified individuals can purchase or view partial or complete portions of the test. This restriction includes personal and educational Internet websites and Internet auction sites. Because all test items, norms, and other testing materials are copyrighted, Pearson must approve, in writing, the copying or reproduction of any test materials. One exception to this requirement is the copying of a completed record form for the purpose of conveying an individual's records to another qualified professional. In addition, copying and printing of specific pages of test materials marked with a reproducible copyright notice is permitted. No part of the test materials may be collected, used, or reproduced in any manner or for any purpose in connection with the development, deployment, or use of artificial intelligence technologies or similar technologies. These user responsibilities, copyright restrictions, and test security issues are consistent with the guidelines set forth in the *Standards* (AERA et al., 2014) and are required by the WAIS-5 and WMS-5 ACS licensing agreement.