Neuropsychological Assessment Evidence in Court:

Quality and Challenges to Admissibility

by

Elizabeth Mathers

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2023 by the
Graduate Supervisory Committee:

Tess M.S. Neal, Chair
Mary Burleson
Nicole Roberts

ARIZONA STATE UNIVERSITY

May 2023

ABSTRACT

When questions about a person's mental state arise in court, psychologists are often called in to help. Psychological assessment tools are routinely included in these evaluations to inform legal decision making. In accordance with the *Daubert* standard, which governs the admissibility of expert testimony, courts are obligated to exclude evidence that relies on poor scientific practice, including assessment tools. However, prior research demonstrates that psychological assessment tools with weak psychometric properties are routinely admitted in court, rarely challenged on the basis of their reliability, and if a challenge is indeed raised, often still admitted (Neal et al., 2019). Is neuropsychological assessment evidence in particular vulnerable to the same pitfalls? The present research aimed to 1) quantify the quality of neuropsychological assessment evidence used in court, 2) evaluate whether courts are calibrated to the quality of these tools through the rate and success of legal admissibility challenges raised, and 3) compare forensic mental health evaluators' experiences and practices with regard to the quality of neuropsychological versus non-neuropsychological assessment tools. Neuropsychological tools appeared to perform worse than non-neuropsychological tools in terms of psychometric quality. However, in a case law analysis, significantly fewer challenges were observed to the legal admissibility of neuropsychological tools than to non-neuropsychological tools. To protect the legitimacy of the legal system and prevent wrongful decisions, it is critical that the evidence on which psychologists' expert opinions are formed is scientifically valid, and that judges and attorneys adequately scrutinize the quality of evidence introduced in court.

DEDICATION

For my Tiger, who stayed by my side until the end. With all my heart and more.

ACKNOWLEDGMENTS

This endeavor would not have been possible without my advisor, Tess M.S. Neal. I would like to thank her for the many years of mentorship and guidance she has provided me throughout my academic career as both an undergraduate and master's student in her lab. I want to thank the members of my committee, Mary Burleson and Nicole Roberts, for accepting me into their research labs, providing feedback on my proposed research, and encouraging me throughout my studies. I am thankful for my fellow lab members, both past and present, including Emily Line for paving the way with her work on forensic evaluators' practices with psychological assessment evidence, and Pamela (Nicky) Sandberg for her invaluable legal advice and experience with case law. I also would like to mention each of my labmates and colleagues who devoted their time to helping me code legal cases for the case law analysis portion of this research, including Karissa Beesley, Albina Letniku, Alexis Mostoller, Pamela (Nicky) Sandberg, Alex Somoza, and Gaurika Shah. I would like to extend my sincere thanks to Marianna Kaneris for her assistance with various aspects of this work and for her kindness and generosity. Finally, words cannot express my gratitude to Jake Plantz for his friendship and continued support throughout this program.

TABLE OF CONTENTS

Page

LIST OF TABLES vi

LIST OF FIGURES vii

NEUROPSYCHOLOGICAL ASSESSMENT EVIDENCE IN COURT: QUALITY AND CHALLENGES TO ADMISSIBILITY 1

Psychological Assessment Evidence 2

Neuropsychological Assessments 4

Legal Admissibility 6

Efficacy of Expert Evidence Admissibility Rules 8

THE CURRENT PROJECT 11

STUDY 1: QUANTIFYING THE QUALITY OF NEUROPSYCHOLOGICAL ASSESSMENT EVIDENCE 11

Method 12

Results 13

Discussion 21

STUDY 2: DETERMINING THE DEGREE TO WHICH COURTS ARE CALIBRATED TO THE QUALITY OF NEUROPSYCHOLOGICAL ASSESSMENT EVIDENCE 23

Method 23

Results 30

Discussion 31

STUDY 3: EXPLORING FORENSIC PSYCHOLOGISTS' ASSESSMENT PRACTICES 35

Page

Method 35

Results 38

Discussion 46

GENERAL DISCUSSION 48

REFERENCES 52

APPENDIX

A NEUROPSYCHOLOGICAL AND NON-NEUROPSYCHOLOGICAL TOOL CLASSIFICATIONS 61

B LIST OF NEUROPYSCHOLOGICAL TOOL NAMES AND ACRONYMS 69

C CODING GUIDE FOR LEGAL CASE LAW ANALYSIS 72

D FORENSIC MENTAL HEALTH EVALUATOR STUDY 3 IRB APPROVAL AND APPROVED DOCUMENTS 76

E FORENSIC MENTAL HEALTH EVALUATOR SURVEY QUESTIONS 83

LIST OF TABLES

Table Page

1. Multinomial Logistic Regression Statistics for General Acceptance between Neuropsychological versus Non-Neuropsychological Tools 16

2. Multinomial Logistic Regression Statistics for Overall Summary Evaluation between Neuropsychological versus Non-Neuropsychological Tools with "No Reviews" Reference 19

3. Multinomial Logistic Regression Statistics for Overall Summary Evaluation between Neuropsychological versus Non-Neuropsychological Tools with "Generally Favorable" Reference 20

4. Categorization of Neuropsychological Tools to Code by General Acceptance and Quality 25

5. Exemplar Neuropsychological Tool Names, Acronyms, and Number of Cases Citing Each 27

LIST OF FIGURES

Figure Page

1. General Acceptance in the Field 15
2. Evidence the Tool Was Subjected to Testing 17
3. Overall Summary Evaluation of Quality 18
4. Admissibility Challenges to Evaluators and Their Colleagues.....38
5. Evaluators' Report of Experiencing Admissibility Challenges to Psychological Assessment Evidence.....39
6. Amount of Effort Preparing to Administer a Tool for the First Time 40
7. Types of Preparation Techniques Before Administering Neuropsychological versus Non-Neuropsychological Tools for the First Time in an Evaluation 42
8. Number of Preparation Techniques Before Administering Neuropsychological versus Non-Neuropsychological Tools for the First Time in an Evaluation 43
9. Ratings of Familiarity with and Importance of Legal Admissibility Criteria.....44
10. Ability to Detect Which Psychological Assessment Tools Meet Admissibility Criteria.....45

Neuropsychological Assessment Evidence in Court:

Quality and Challenges to Admissibility

The U.S. court system adjudicates a tremendous volume of legal cases each day. Over 100 million cases are filed in state trial courts every year, and an additional 400,000 in federal trial courts (Institute for the Advancement of the American Legal System, 2015). Psychological assessment is one of many factors that interact to inform decision making in these legal contexts. Experts may be called upon by attorneys or the court to provide a psychological evaluation and facilitate decision making (Melton et al., 2017). In general, psychological assessment aids in establishing facts about a person's mental state or psychological capacities in court. Although psychological tools can be beneficial in determining legally relevant mental health symptoms and functioning, the psychometric qualities of these instruments may not always be carefully examined prior to their admission in court (Neal et al., 2019).

Because psychological evaluations can be so influential in shaping legal decisions, it is important to ensure that the tools that are used in these evaluations have been well-validated. Research examining the psychometric rigor of psychological assessments that are accepted for use in legal contexts finds considerable variation in quality. In a systematic review of 364 assessment tools that forensic evaluators reported having used in legal cases, Neal and colleagues (2019) discovered that only 67% of these tools are generally accepted in the field and only 40% possess favorable reviews of their psychometric properties. Further, attorneys rarely raised legal admissibility challenges to psychological assessment evidence (only about 5% of the time) and challenges that were raised were rarely successful (Neal et al., 2019). No studies appear to have yet focused

specifically on the properties of neuropsychological assessment evidence admitted by the courts. The proposed studies are thus intended to examine the quality of neuropsychological tools in particular, the rate and success of challenges raised to neuropsychological assessment evidence, and forensic mental health practitioners' practices and experiences using neuropsychological assessment evidence in court.

Psychological Assessment Evidence

Psychological experts frequently disagree as to the interpretation of psychological evidence, and on some occasions, experts may even come to opposite conclusions based on the same evidence. This is often evident in cases where a "battle of the experts" arises, where experts provide opposing testimony on the same question. There are many high profile situations where this has occurred. In John Hinckley Jr.'s trial, four experts for the defense testified that he suffered from schizophrenia and was insane at the time he attempted to assassinate President Ronald Reagan, while five experts for the prosecution argued instead that he was fully aware of his actions (Ewing & McCann, 2006). In a case against the heavy metal band Judas Priest, four experts for the prosecution argued that subliminal messaging hidden in their lyrics drove the suicide and attempted suicide of two young men, while the experts for the defense identified flaws in these psychodynamic concepts and pointed rather to preexisting depression, substance abuse, and challenging life circumstances as factors underlying their suicidality (Ewing & McCann, 2006). Seven experts, distributed among the prosecution, the defense, and the court, testified in the case of Jeffrey Dahmer, a serial killer and sex offender, disagreeing about his sanity and mental diagnoses (Ewing & McCann, 2006). What explains this lack

of consensus among trained and experienced experts? The variable quality of psychological tools used in evaluations may be at least partially responsible.

Psychometry, derived from the Greek *psyche* (soul) and *metro* (measure), was first defined by Sir Francis Galton in 1879 as the "the art of imposing measurement and number upon operations of the mind" (p. 149). Inspired by Darwin's work studying natural selection and inheritance, Galton (1869) advanced the idea that mental attributes are measurable quantities and sought to classify individuals accordingly. Much of Galton's work revolved around the quantification of phenomena, from the examination of biological characteristics such as variations in fingerprints, to audience interest levels, to changes in weather patterns, and more (Wasserman & Bracken, 2013). Modern psychometric theories have evolved from this foundation. Two frameworks dominate the present ideology in psychometry – classical test theory (CTT) and item response theory (IRT) – though test developers often apply aspects of both approaches in tandem (Nunnally & Bernstein, 1994).

Psychologists use psychological testing to inform a wide array of legal judgments, including competency, violence risk, parental fitness, sanity, disability claims, and countless other legal issues (Neal & Grisso, 2014). A wide variety of techniques are available at their disposal to gather relevant information, and most often include clinical interviews (Segal & Hersen, 2010). These approaches can be classified within three broad categories: 1) structured actuarial assessments, based on standardized psychometric tools; 2) unstructured clinical judgments, based on the expert's subjective intuition without the use of psychometric tools; and 3) structured clinical judgments, which involve aspects of both structured actuarial and unstructured clinical approaches (Douglas et al., 1999; Faust

& Ziskin, 1988; Skeem & Monahan, 2011). These assessment approaches tend to be of varying quality, with structured assessment methods typically achieving better levels of diagnostic accuracy than unstructured assessments (Miller et al., 2001).

Neuropsychological Assessments

Within the broader purview of psychological assessment fall many specialized types of assessment. The sub-field of neuropsychology focuses its scope to address the relationship between brain and behavior (e.g., APA, 2008; Hebben & Milberg, 2009). Neuropsychological assessment inspects an individual's cognitive functioning, including intellectual capabilities, reading and language comprehension, learning and memory, processing abilities, and other related domains (e.g., Harvey, 2012; Hebben & Milberg, 2009).

Neuropsychologists are often asked by the courts to provide an opinion on a person's level of cognitive function and performance. These questions frequently arise in both civil and criminal cases. In civil contexts, neuropsychological assessment can be relevant in guardianship evaluations (Rothke et al., 2019), testamentary capacity and undue influence (Mart, 2016), financial and legal competency (Cohen et al., 2019), and in other types of cases. In criminal cases, neuropsychological assessment evidence is often introduced in questions of criminal competency, determinations of insanity, and sentencing decisions, among others (LaDuke et al., 2012). Prior research suggests that legal admissibility standards for reliable and valid evidence are applied more lackadaisically in criminal trials than in civil cases (Faigman et al., 2008). Although neuropsychological measures are commonly used to inform legal issues in cases like

these, the ecological validity of many of these tests, or their ability to produce valid results in real-world settings, may be limited (Sbordone & Long, 1996).

Problematically, people may be unduly susceptible to influence by "neuro-" information, a phenomenon coined "neuroenchantment" (Ali et al., 2014). Relatedly, researchers have noted that neuroscientific explanations often reach beyond what the evidence really suggests, describing this concept as "neurohype" (Lilienfeld et al., 2018). A number of similar terms also exist to describe this notion, such as neuromania, neuropunditry, and neurobollocks (Lilienfeld et al., 2018; Satel & Lilienfeld, 2013). Essentially, this points to the idea that the prefix "neuro-" itself may superficially increase scientific credibility, rendering people in general less critical of evidence (Ali, 2012). Though this concept was defined in the context of neuroscience and neural imaging, it might also extend to neuropsychological practice and assessment.

Judges may be vulnerable to the guise of scientific evidence, and potentially to neuropsychological evidence. Although judges are generally excellent decision makers, their judgments are nonetheless susceptible to bias and error (Rachlinski & Wistrich, 2017). Judges can be persuaded by the popular reputation of scientific evidence and the qualities of the expert presenting it, rather than purely by the underlying scientific methodology (Nir & Liu, 2021; see also Brodsky et al., 2009). With respect to the admissibility of neuropsychological evidence in court, the consequences of this phenomenon are problematic if judges or other legal characters overlook the caliber of the evidence which is introduced in favor of its title and reputation.

Legal Admissibility

Standards for the admissibility of evidence in U.S. courts have evolved over the past century. *Frye v. United States* (1923) was the first case to delineate the admissibility of scientific evidence in legal settings. In this case, James Alphonzo Frye was convicted of second-degree murder after the trial court excluded expert testimony about a systolic blood pressure deception test which purported that deception is manifested physiologically from fear during examination. An appeals court later affirmed that this test had not yet gained recognition among psychological authorities and thus established the *Frye* standard – that scientific evidence must be generally accepted within its relevant scientific community in order to justify its admission in court.

The *Frye* standard was superseded by the 1993 case, *Daubert v. Merrell Dow Pharmaceuticals, Inc.* Together with two subsequent cases (*General Electric v. Joiner*, 1997; *Kumho Tire Co, Ltd. v. Carmichael*, 1999), this "*Daubert* trilogy" effectively established the current guidance for evidence admissibility in the United States. The *Daubert* case enumerated a series of factors that judges can (but are not required to) consider in evidence admissibility decisions. In addition to general acceptance, these include: whether the technique can and has been tested, whether it has undergone peer review, the existence of standards controlling its operation, and its known or potential rate of error. These criteria are codified in the Federal Rules of Evidence (FRE) Rule 702, which guides the admissibility of expert testimony.

Though the *Daubert* standard is set as the current guidance for evidence admissibility at the federal level and in the majority of states, much variability still proliferates throughout the United States regarding precisely which evidence

admissibility standards are upheld. The *Daubert* standard is often not applied uniformly even in states where it is endorsed, and many states have adopted modifications of *Daubert* criteria. For example, Iowa (which employs a modified *Daubert* standard in both civil and criminal cases) encourages but does not require adherence to *Daubert* specifications. Moreover, several states continue to rely on the *Frye* or *Frye*-plus standard (i.e., Illinois, Minnesota, New York, Pennsylvania, and Washington; Hill, 2020).

Under FRE 702, judges are designated as the gatekeepers of scientific evidence. In other words, judges hold the responsibility to ensure that the evidence admitted into court is both relevant and reliable. Despite having defined admissibility standards, judges are granted a high degree of discretion and flexibility in their decisions. *General Electric v. Joiner* (1997), part of the *Daubert* trilogy, further solidified this freedom in defining the "abuse of discretion" standard as the proper standard in appellate review. Under this standard, when an appellate court reviews a lower court's evidentiary rulings (including the decision to admit or exclude expert testimony), deference is given to the trial judge's decision unless error is apparent or the discretion exercised is unjustifiable.

In practice, however, the court typically relies on the capabilities of the adversarial process, rather than the independent action of judges, to filter out questionable evidence through "vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof" (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 1993, section 596). Indeed, in Nir and Liu's examination of judicial perceptions of scientific evidence, one judge commented "The adversarial system fails its purpose when the attorney is not skilled enough to challenge the expert" (2021, p. 9). The adversarial structure of the U.S. criminal justice system incentivizes prosecutors to maximize

convictions (and thus, to introduce expert testimony in favor of judgment), and defense attorneys to challenge expert testimony (Damon-Moore, 2017; see also Fisher, 1988). In reality, however, challenges to admissibility – and the court's ability to screen out poor evidence – are limited by practical obstacles. Criminal defense attorneys, especially public defenders, grapple with intense caseloads, insufficient compensation, and sparse knowledge regarding the shortcomings of evidence (Damon-Moore, 2017). Markedly, one analysis found that out of the near 15 million state criminal cases filed between August 1999 through August 2000, only 50 *Daubert* admissibility challenges to forensic science evidence were reported (Neufeld, 2005).

Efficacy of Expert Evidence Admissibility Rules

Are *Daubert* and similar standards successful in filtering out invalid and unreliable evidence? There are widespread concerns about the scientific merit and replicability of the methods underlying various forms of evidence and testimony. This concern is not limited to psychological assessment, but rather spans across forensic disciplines, such as latent print identification, hair and fiber analysis, toxicology, and DNA analysis, to name a few. Many disciplines lack cohesion in oversight, accreditation, certification, standards of operation, and ethics (National Research Council, 2009). In 2016, the President's Council of Advisors on Science and Technology (PCAST) released a report commenting on the validity and reliability of standard forensic science techniques, presenting their concerns about the unknown error rates of many techniques, subjective nature of procedures commonly perceived to be reliable (e.g., DNA and fingerprint analyses), and the presence of confirmation and contextual biases in forensic examiners.

The field of psychology cannot escape the same criticisms. The reliability of psychological assessment tools differs between the lab and the field; generally, tools perform worse in clinical practice than they do in controlled experimental settings (Edens & Boccaccini, 2017). As noted previously, evaluators often reach different conclusions on the basis of the same evidence. Some of the disagreement among "expert" opinions might be explained by the limited training and certification requirements for forensic evaluators (Guarnera et al., 2017). Fewer than half of states provide state-level certification in forensic assessment, and those that do may only have weak standards (Gowensmith et al., 2015). In the absence of standardized training, forensic psychologists may utilize a wide variety of evaluation methods (structured assessment tools, reliance on clinical judgment, or a combination of both), further contributing to disagreement among evaluators (Neal & Grisso, 2014).

With flaws such as these embedded deeply within the practice of forensic science and psychology, it may not be surprising that so-called "junk" science is frequently admitted as evidence in court. In 2019, Neal and colleagues found that only about 67% of the 364 psychological assessment tools included in their review of tools used in legal cases are generally accepted in the field. Further, legal challenges to the admissibility of psychological assessment evidence occurred in only about 5% of the sample. Even when challenges were raised, the evidence in question was still likely to be admitted (Neal et al., 2019). Evidence challenged on the basis of reliability (by which the courts mean validity) is also more likely to be admitted in court than evidence where non-reliability challenges are raised, or those based on relevance to the legal issue, violations of procedural rules, or qualifications of the expert (O'Brien, 2018).

Why is evidence with weak psychometric characteristics accepted in court? One explanation may be that judges are largely untrained in science (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 1995, p. 1316). Judges may struggle to understand and apply concepts necessitated by *Daubert* and similar criteria. In a survey by Gatowski and colleagues (2001), only about 5% of state court judges expressed a clear understanding of factors like error rate and falsifiability, relying instead on the principle of general acceptance to guide their admissibility decisions. Judges and lawyers are also uninfluenced by variations in the reliability and validity of psychological assessments (Chorn & Kovera, 2019). Although judges and attorneys are interested in hearing testimony about mental health, they often fail to consider the underlying research and statistical analyses (Redding et al., 2001).

Jurors, too, struggle to differentiate between high and low quality evidence. Jurors are not sensitive to variations in the validity of psychological science, specifically the impact of control groups, confounds, and experimenter bias (McAuliff et al., 2009), nor are they sensitive to differences in the quality of scientific evidence (Neal, 2020). In a trial context, mock jurors are vulnerable to persuasion by weak scientific evidence because they assume that judges have effectively fulfilled their role in barring low quality science by screening the evidence in advance (Schweitzer & Saks, 2009). Psychological assessment evidence has the potential to wield a powerful influence over the judge and jury. When legal decision makers do not understand it well enough to differentiate between the strong and weak, the ability of the judicial system to assess cases fairly and accurately is compromised. These issues could be exacerbated for neuropsychological

assessment evidence if a neuroenchantment bias is also at play, rendering the courts particularly vulnerable to these threats.

The Current Project

Research has yet to examine whether courts are sensitive to the quality of neuropsychological assessment evidence in particular. Given the high frequency with which neuropsychological evidence is proffered in both civil and criminal cases (e.g., Cohen et al., 2019; Mart, 2016; Rothke et al., 2019; LaDuke et al., 2012), this constitutes a significant gap in the literature. In a series of three studies, we aimed to (1) quantify the quality of the neuropsychological assessment tools that are commonly used by experts in court, (2) determine the degree to which courts are calibrated to the quality of neuropsychological assessment evidence, and (3) understand forensic mental health evaluators' perspectives on and experiences using neuropsychological versus non-neuropsychological assessment tools to inform their expert judgments in court.

Study 1: Quantifying the Quality of Neuropsychological Assessment Evidence

Are the neuropsychological instruments used to inform legal decisions in court psychometrically strong? Neuropsychological assessment tools are likely to be vulnerable to the same problems as psychological assessment evidence in general. In line with differences in rigor and norms between the fields, we hypothesized that neuropsychological instruments would have stronger psychometric qualities than non-neuropsychological assessment tools. Specifically, we anticipated that neuropsychological tools would achieve higher rates of general acceptance within the community, higher rates of testing, and more favorable peer review evaluations of their psychometric quality than non-neuropsychological tools. We expected this because a

greater percentage of neuropsychologists are board certified than forensic psychologists (Lin et al., 2017).

Method

Procedure. All hypotheses and methods were preregistered on the Open Science Framework (https://osf.io/7anw3/). For this study, secondary data analysis was conducted on Neal and colleagues' (2019) systematic review of psychological assessment tools routinely used in courts. In their study, a list of 364 psychological assessments was generated from 22 surveys of forensic mental health practitioners. These tools were coded by Neal and colleagues for general acceptance in the field, subjection to testing, and information about peer review evaluations of psychometric quality.

In the systematic analysis, coders identified whether the 364 assessment tools were generally accepted or endorsed by clinicians in the field. Nine surveys regarding experts' views of assessment tools were searched for information regarding the general acceptance of the tools (Archer et al., 2016; Elhai et al., 2005; LaDuke et al., 2018; Lally, 2003; McLaughlin & Kan, 2014; Neal & Grisso, 2014; Rabin et al., 2005; Ryba et al., 2003; Slick et al., 2004). These surveys were all published within the 15 years prior to the systematic review and included responses from a total of 2,384 experienced mental health experts. On average, respondents had 15 years of experience and about 15% of the sample was board certified. Coders categorized whether "the information indicated that the tool was generally accepted in the field," "it was clear that the tool was not generally accepted," "the information about general acceptance was conflicting or unclear," or "there was not enough information available to determine whether a tool was generally accepted" (Neal et al., 2019).

Evidence for whether a tool had been subjected to testing was also gathered. First, coders assessed the availability of the tools and their manuals (if they had one). Coders determined whether the tool was commercially published, whether it had a commercially published manual, whether it was available for free online, and whether it had a noncommercially published manual. Coders then indicated if they were able to find evidence of testing, such as through information available in test manuals, peer-reviewed publications, comprehensive review sources (e.g., *Mental Measurements Yearbook*), online data sources, or other references.

Reviews of psychometric quality were compiled from primary comprehensive sources, including the *Mental Measurements Yearbook* (Carlson et al., 2017), Strauss and colleagues' *Compendium of Neuropsychological Tests* (2006), and Grisso's *Evaluating Competencies: Forensic Assessments and Instruments* (2003). Coders provided an overall summary evaluation of psychometric quality (generally favorable, generally unfavorable, or mixed summary evaluations) from these professional reviews of psychological assessment tools, based on a similar method by Cizek and colleagues (2012).

For the current project, the list of 364 psychological assessment tools was split into neuropsychological assessments (n = 57) and non-neuropsychological assessments (n = 307) based on classifications assembled by King and colleagues (2017; see Appendix A for full list of tool classifications). These two categories were then compared based on their levels of general acceptance, testing rates, and peer review summary ratings.

Results

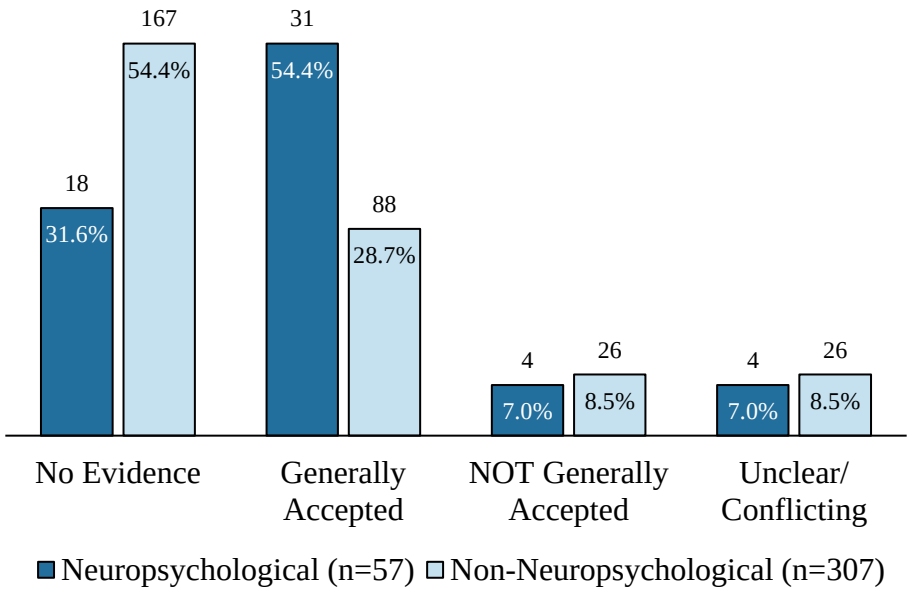
Logistic regression was conducted to assess whether there were differences between neuropsychological and non-neuropsychological tools with a 95% confidence

interval. The neuropsychological classification of the tools was entered as a categorical independent variable (two levels: neuropsychological or non-neuropsychological), and the following three criteria were entered separately as categorical dependent variables: 1) general acceptance (four levels: no evidence, generally accepted, not generally accepted, or unclear/conflicting evidence), 2) subjection to testing (two levels: yes or no), and 3) peer review summary evaluations (four levels: no reviews, generally favorable, generally unfavorable, or mixed reviews).

Over half (54.4%) of the neuropsychological tools and less than one-third (28.7%) of the non-neuropsychological tools in the sample were generally accepted among psychologists in the field (see Figure 1). There were some neuropsychological and non-neuropsychological tools which had clear evidence of being *not* generally accepted in the field, and some that had unclear or conflicting indications of general acceptance. Approximately one-third of the neuropsychological tools (31.6%) and over half (54.4%) of the non-neuropsychological tools lacked any evidence of general acceptance.

Figure 1

General Acceptance in the Field



Note. This figure describes the general acceptance of neuropsychological and non-neuropsychological tools. Percentages and frequencies are displayed.

Multinomial logistic regression was used to analyze differences in the general acceptance of neuropsychological versus non-neuropsychological tools with a reference category of "no evidence of general acceptance" (see Table 1). In comparison to neuropsychological tools, non-neuropsychological tools were 0.31 times as likely to be generally accepted in the field than to have no evidence of general acceptance (or in other words, non-neuropsychological tools were 3.23 times more likely to have no evidence of general acceptance than they were to be generally accepted; b = -1.18, SE = 0.32, *Wald*

$X^2(1) = 13.34$, $OR = 0.31$, $p < .001$). No other contrasts regarding general acceptance were statistically significant.

Table 1

Multinominal Logistic Regression Statistics for General Acceptance between Neuropsychological versus Non-Neuropsychological Tools

Outcome Group	*B*	*SE*	*Wald*	*df*	*p*	*OR*	*95% CI*
NOT Generally Accepted							
Intercept	-1.50	0.55	7.40	1	.007		
NeuroClass	-0.36	0.59	0.36	1	.55	0.70	[0.22, 2.23]
Unclear							
Intercept	-1.50	0.55	7.40	1	.007		
NeuroClass	-0.36	0.59	0.36	1	.55	0.70	[0.22, 2.23]
Generally Accepted							
Intercept	0.54	0.30	3.36	1	.07		
NeuroClass	-1.18	0.32	13.33	1	<.001	.31	[0.16, 0.58]

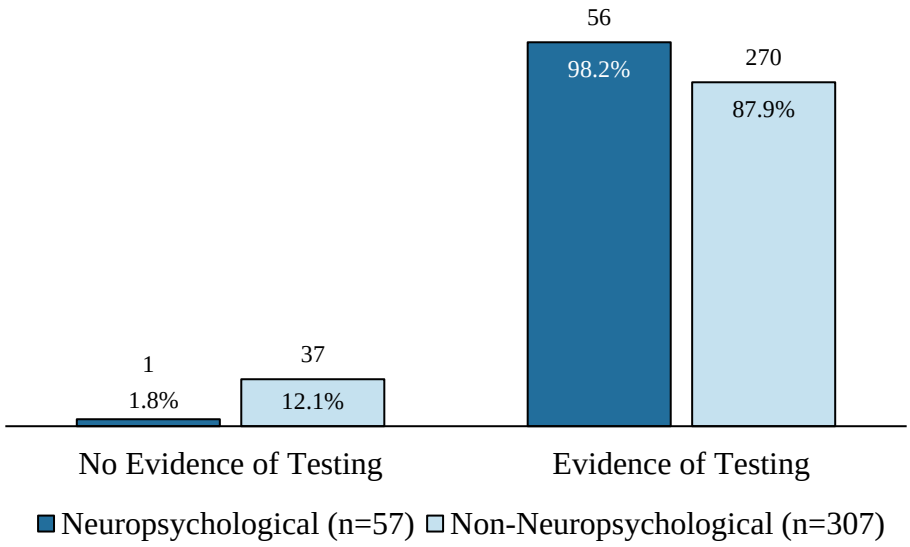
Note. This table displays the multinomial logistic regression statistics for general acceptance in the field. The reference category is no evidence of general acceptance. NeuroClass refers to the tool's classification as a non-neuropsychological tool (in comparison to neuropsychological tools). OR is the odds ratio. The dependent variable is one of four mutually exclusive outcome groups (no evidence of general acceptance, unclear evidence of general acceptance, not generally accepted, generally accepted).

Out of the 57 neuropsychological tools in the sample, all but one (98.2%) had been subjected to empirical testing (see Figure 2). In contrast, no evidence of testing could be found for approximately 12% of the non-neuropsychological tools. Binary

logistic regression demonstrated that neuropsychological tools were 7.67 times more likely to have been tested than non-neuropsychological tools, (b = 2.04, SE = 1.02, *Wald* $X^2(1) = 3.96$, $OR = 7.67$, $p = .047$).

Figure 2

Evidence the Tool Was Subjected to Testing



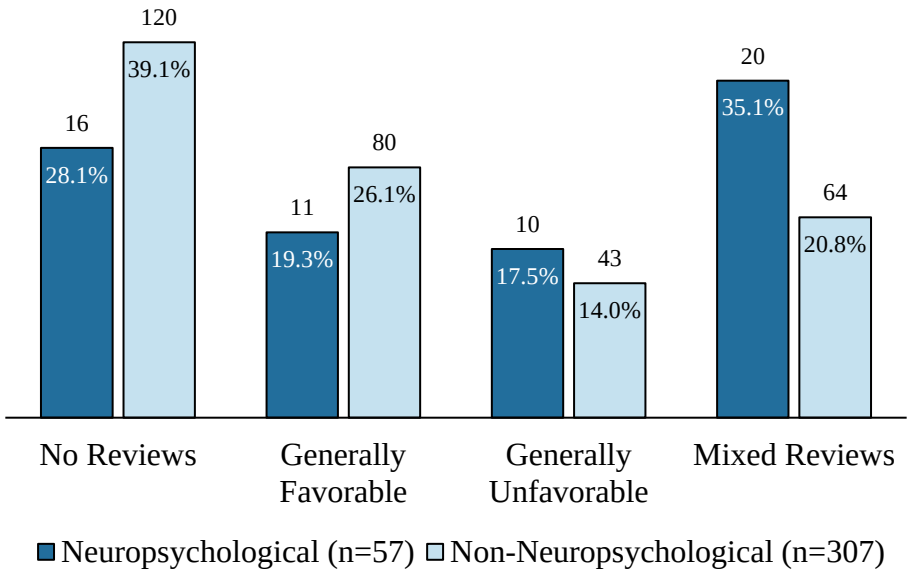
Note. This figure describes whether evidence of testing was available for neuropsychological and non-neuropsychological tools. Percentages and frequencies are displayed.

Many of the tools included in the sample had no peer reviews of their psychometric quality (28% of neuropsychological tools and nearly 40% of non-neuropsychological tools; see Figure 3). Those that did ranged across the board in terms

of quality. Only about 19% and 26% of neuropsychological and non-neuropsychological tools, respectively, had generally favorable reviews of their psychometric properties. A portion of both neuropsychological and non-neuropsychological tools had generally unfavorable reviews, as well as some with mixed reviews of quality.

Figure 3

Overall Summary Evaluation of Quality



Note. This figure describes the overall peer review summary evaluation of neuropsychological and non-neuropsychological tools. Percentages and frequencies are displayed.

We ran two separate models using multinomial logistic regression for peer review summary evaluations of quality. First, a reference category of "no reviews" was used (see

Table 2). Second, the reference category was updated to "generally favorable reviews" (see Table 3). In comparison to non-neuropsychological tools, neuropsychological tools were 2.34 times more likely to have mixed peer review evaluations than to have no reviews (b = .85, SE = .37, $Wald\ X^2(1)$ = 5.32, OR = 2.34, p = .02), but 2.27 times more likely to have no reviews than to have generally favorable reviews (b = .82, SE = .41, $Wald\ X^2(1)$ = 3.99, OR = 2.27, p = .046. No other contrasts regarding peer review summary evaluations were statistically significant.

Table 2

Multinomial Logistic Regression Statistics for Overall Summary Evaluation between Neuropsychological versus Non-Neuropsychological Tools with "No Reviews" Reference

Outcome Group	*B*	*SE*	*Wald*	*df*	*p*	*OR*	*95% CI*
Mixed Reviews							
Intercept	0.22	0.34	0.44	1	.51		
NeuroClass	-0.85	0.37	5.32	1	.02	0.43	[0.21, 0.88]
Generally Unfavorable							
Intercept	-0.47	0.40	1.36	1	.24		
NeuroClass	-0.56	0.44	1.59	1	.21	0.57	[0.24, 1.36]
Generally Favorable							
Intercept	-0.38	0.39	0.92	1	.34		
NeuroClass	-0.03	0.42	0.01	1	.94	0.97	[0.43, 2.20]

Note. This table displays the multinomial logistic regression statistics for overall summary evaluation. The reference category is no reviews. NeuroClass refers to the tool's classification as a non-neuropsychological tool (in comparison to neuropsychological tools). OR is the odds ratio. The dependent variable is one of four

mutually exclusive outcome groups (no reviews, mixed reviews, generally unfavorable reviews, generally favorable reviews).

Table 3

Multinomial Logistic Regression Statistics for Overall Summary Evaluation between Neuropsychological versus Non-Neuropsychological Tools with "Generally Favorable" Reference

Outcome Group	*B*	*SE*	*Wald*	*df*	*p*	*OR*	*95% CI*
No Reviews							
Intercept	0.41	0.14	7.89	1	.005		
NeuroClass	-0.03	0.42	0.01	1	.94	0.97	[0.43, 2.20]
Mixed Reviews							
Intercept	-0.22	0.17	1.77	1	.18		
NeuroClass	0.82	0.41	3.99	1	.046	2.27	[1.02, 5.09]
Generally Unfavorable							
Intercept	-0.62	0.19	10.78	1	.001		
NeuroClass	0.53	0.48	1.22	1	.27	1.69	[0.67, 4.30]

Note. This table displays the multinomial logistic regression statistics for overall summary evaluation. The reference category is generally favorable reviews. NeuroClass refers to the tool's classification as a non-neuropsychological tool (in comparison to neuropsychological tools). OR is the odds ratio. The dependent variable is one of four mutually exclusive outcome groups (no reviews, mixed reviews, generally unfavorable reviews, generally favorable reviews).

Discussion

Neuropsychological assessment tools were expected to generally fare better than non-neuropsychological assessment tools in terms of their general acceptance, rates of testing, and peer review evaluations. Even so, both high and low quality neuropsychological tools were expected to still be admitted in court. This hypothesis was partially supported.

Overall, substantial variation was present among the 364 assessment tools that psychologists reported using in their forensic evaluations. Neuropsychological tools appeared to perform better than non-neuropsychological tools on some indicators of quality, but not exclusively. Encouragingly, most of the tools in the total sample had been subjected to empirical testing. Only one neuropsychological tool did not have evidence of testing, while 12.1% (n = 37) of non-neuropsychological tools did not have evidence of testing. This indicator suggests that fewer non-neuropsychological tools are subjected to testing than neuropsychological tests.

Neuropsychological tools were significantly more likely than non-neuropsychological tools to be generally accepted than to lack evidence of general acceptance. Approximately 54% of neuropsychological tools were generally accepted, compared to only 28.7% of non-neuropsychological tools. Additionally, only about 32% of neuropsychological tools had no evidence of general acceptance, compared to 54.4% of non-neuropsychological tools. The percentage of tools that were either *not* generally accepted or had unclear evidence of general acceptance was similar for both neuropsychological and non-neuropsychological tools.

Neuropsychological tools did not appear to perform better than non-neuropsychological tools on the overall summary evaluation of quality from peer reviews. A considerable percentage of both neuropsychological tools (28.1%) and non-neuropsychological tools (39.1%) did not have any reviews of quality. Of the tools that did have reviews, a similar percentage of neuropsychological and non-neuropsychological tools had either generally favorable reviews (19.3% of neuropsychological tools, 26.1% of non-neuropsychological tools) or generally unfavorable reviews (17.5% of neuropsychological tools, 14.0% of non-neuropsychological tools). Finally, more neuropsychological tools (35.1%) had mixed reviews than non-neuropsychological tools (20.8%).

This first study provides a descriptive outlook on the current state of neuropsychological assessment tools that are admitted in court. It is problematic that any tools which are not generally accepted or which have unfavorable reviews of their psychometric properties are applied in legal settings. Psychological testimony is often very persuasive in driving the decisions of legal fact finders (e.g., Melton et al., 2017; Zapf et al., 2004), and relying on unreliable, invalid, or otherwise substandard psychological assessments and methods can thus jeopardize the fairness of the justice system.

We questioned whether neuropsychological tools might be scrutinized differently in court than non-neuropsychological tools on the basis of their quality. If courts are well calibrated to the quality of assessment tools, admissibility challenges should be raised more frequently to those which are of poorer quality, and less frequently to those which

are of higher quality. The next study in this sequence sought to further explore the legal admissibility of neuropsychological assessment tools.

Study 2: Determining the Degree to Which Courts are Calibrated to the Quality of Neuropsychological Assessment Evidence

How frequently do attorneys raise challenges to questionable neuropsychological assessment tools when they are introduced as evidence in court, and are they successful when they do? In line with Neal and colleagues' prior findings (2019), we expected that admissibility challenges to neuropsychological assessment evidence would be infrequent and often unsuccessful.

Method

Procedure. A legal case law review was conducted to determine the rate of admissibility challenges raised to neuropsychological assessment evidence in court, as well as how often these challenges are successful. In this case law analysis, we searched through published cases where the results of assessment tools were introduced as evidence by psychological experts. In Neal et al.'s (2019) review, 30 exemplars were selected to represent variation in referral questions and legal issues, as well as variation in general acceptance (no evidence, generally accepted, debated, not generally accepted) and psychometric quality (no reviews, generally unfavorable reviews, mixed reviews, and generally favorable reviews). Three of the 30 psychological assessment tools included as exemplars were neuropsychological (Iowa Gambling Task [IGT], Paced Auditory Serial Addition Test [PASAT], and Wisconsin Card Sorting Task [WCST]).

To expand on this for the current study, 27 additional neuropsychological tools were selected as exemplars (for a total of 30 neuropsychological exemplar tools) to

broadly capture variation in the general acceptance (no evidence, generally accepted, debated, not generally accepted) and quality (no reviews, generally unfavorable reviews, mixed reviews, and generally favorable reviews) of tools (see Table 4). Tools were selected within each combination of quality and acceptance, if possible. We also aimed to provide information on some tools which are commonly used (e.g., the Repeatable Battery for the Assessment of Neuropsychological Status [RBANS], Mini-Mental State Evaluation [MMSE], etc.). An updated search was conducted for the three neuropsychological tools analyzed in the previous review (IGT, PASAT, and WCST), given that the current case law review analyzed a different time range and body of cases.

Table 4

Categorization of Neuropsychological Tools to Code by General Acceptance and Quality

	Generally favorable reviews	Mixed reviews	Generally unfavorable reviews	No professional reviews
Generally accepted	NAB RAVLT **RBANS** ROCF/RCFT **WMS**	**Bender-Gestalt** BNT CAT/HCT/BCT Cognistat **CVLT** D-KEFS Grip Strength Grooved Peg HRNTB MFD **MMSE** Stroop **Trails A&B**	**PASAT** **Finger Tap** **SDMT** **TOVA** **TPT** **VOT**	**COWA/COWAT** FAS **MoCA** SPE SRT SSPT **WCST/WCST-64**
General acceptance debated	**PPVT**	**BVRT** CVMT	**LNNB**	
Not generally accepted		**IGT**		**ANIMALS** **Lat. Dominance** **Reitan-Indiana**
No evidence of general acceptance	**BRIEF** **BVMT** CogScreen KABC PAL	CCT CVLT-C **NART** PPC	**Beery VMI** LNNB-C **SCT**	Austin Maze CAVLT GNDS **Halstead-Wepman** **TFR** **TOPF**

Note. This table classifies all of the neuropsychological tools in the overall sample (N = 57 of the 364 assessment tools) according to their ratings of general acceptance and peer review summary evaluations of psychometric quality. Columns are organized by overall summary evaluation rating of quality. Rows are organized by general acceptance in the field. Tools in bold font (n = 30) were included in the case law analysis. The full name of each tool is available in Appendix B.

NexisUni, an academic legal search engine, was used to identify cases in which neuropsychological assessment evidence was introduced. NexisUni indexes business, news, and legal sources dating back to the year 1790, with legal content including administrative codes and regulations; administrative materials; briefs, pleadings, and motions; cases; law reviews and journals; legal news; and statutes and legislation (https://www.lexisnexis.com/). For this project, only cases were searched, which included all federal and state cases in the United States for the time period of interest.

The name of each of the 30 exemplar neuropsychological tests, as well as its acronym, were searched for all U.S. federal and state cases published within the last three calendar years (2020, 2021, and 2022). Cases published in languages other than English were excluded from analysis. If the acronym of the tool was a common word (e.g., "BRIEF"), additional terms were used to refine the search to relevant records (such as "expert" or "psych!" to include cases where the root "psych" was mentioned). Up to 30 cases were coded for each tool. If fewer than 30 cases were available for a given tool, all were coded. If more than 30 cases were available, a random subset of 30 cases was selected using a random choice generator (gigacalculator.com; see Table 5 for a list of the exemplar neuropsychological tools and the number of cases which cited the tool). Complete information about the cases coded is available on the Open Science Framework (https://osf.io/7anw3/).

Table 5

Exemplar Neuropsychological Tool Names, Acronyms, and Number of Cases Citing Each

Test/tool name	Acronym(s)	No. of cases citing tool	Reference
Animal Naming [verbal fluency test/task]	ANIMALS	1	-
Beery-Buktenica Developmental Test of Visual-Motor Integration	Beery VMI	9	Beery et al., 2010
Bender Visual-Motor Gestalt Test	Bender-Gestalt	22	Bender, 2003
Behavior Rating Inventory of Executive Function	BRIEF	12	Gioia et al., 2015
Brief Visuospatial Memory Test	BVMT	4	Benedict, 1997
Benton Visual Retention Test	BVRT	1	Benton-Sivan, 1991
Controlled Oral Word Association Test	COWA/COWAT	8	Benton et al., 1994
California Verbal Learning Test	CVLT	19	Delis et al., 2000
Finger Tap/Tapping/ Oscillation Test	FTT	10	Reitan & Wolfson, 1993
Halstead-Wepman Aphasia Screening Test		0	Halstead & Wepman, 1949
Iowa Gambling Task	IGT	0	Bechara, 2007
Lateral Dominance Examination		1	Reitan & Wolfson, 1993
Luria-Nebraska Neuropsychological Battery	LNNB	2	Golden et al., 1985
Mini-Mental State Examination	MMSE	30	Folstein et al., 1975
Montreal Cognitive Examination	MoCA	29	Nasreddine et al., 2005
National Adult Reading Test	NART	0	Nelson & Willison, 1991
Paced Auditory Serial Addition Test	PASAT	1	Gronwall, 1977
Peabody Picture Vocabulary Test	PPVT	9	Dunn & Dunn, 2007
Repeatable Battery for the Assessment of	RBANS	29	Randolph, 2012

Test/tool name	Acronym(s)	No. of cases citing tool	Reference
Neuropsychological Status			
Reitan-Indiana Aphasia Screening Test		3	Reitan & Wolfson, 1993
Short Category Test	SCT	1	Reitan & Wolfson, 1993
Symbol Digit Modalities Test	SDMT	4	Smith, 1973
Tactile Form Recognition Test	TFR	1	Reitan & Wolfson, 1993
Test of Pre-Morbid Functioning	TOPF	5	Holdnack & Drozdick, 2009
Test of Variables of Attention	TOVA	7	Greenberg et al., 2018
Tactual Performance Test	TPT	3	Reitan & Wolfson, 1993
Trails A & B/Trail Making Test/Comprehensive TMT	TMT/CTMT	30	Reitan & Wolfson, 1993
Hooper Visual Organization Test	VOT	1	Hooper, 1983
Wisconsin Card Sorting Test	WCST/WCST-64	24	Heaton et al., 1993
Wechsler Memory Scale	WMS	30	Wechsler, 2009

Note. This table displays the 30 neuropsychological exemplar tools used in the legal case law analysis, their acronyms, the number of cases which were coded for each tool, and the tool's reference, if available. A random subsection of 30 cases was selected for the five tools which returned more than 30 case results: MMSE, MoCA, RBANS, Trails, and WMS. The MoCA and RBANS results each included a case which mentioned the tool but the tool was not actually used in the case and so excluded from analysis.

Screening and Coding Cases. Each case was screened to ensure that the neuropsychological tool was used, analyzed for whether an admissibility challenge was

raised, and if so, whether the challenge was successful along with additional details about the challenge. Each case was double coded by two of seven total coders to achieve a measure of interrater reliability. Additional information about the cases in which the tools were introduced was also coded, including the jurisdiction (federal, state), the level of the court (trial, appellate), and the type of case (criminal, civil).

A list of 1,638 legal cases was generated from the original search of tool names and acronyms. After preliminarily screening these cases to ensure the actual tools were included (e.g., discarding cases which referred to hiking trails rather than the Trail Making Test), 786 legal cases remained. Five tools produced more than 30 case results (i.e., WMS, Trails, RBANS, MOCA, and MMSE); for these five tools, random subsets of 30 cases for each were selected, narrowing the sample to 301 legal cases. Five additional cases were later screened out in which the tool was mentioned in some way (for example, citing a precedential case which included the tool) but not used in the case. In some cases, test batteries including multiple different assessment instruments were administered. For example, in the case E.M. v. Kijakazi (2022), the Beery VMI, Trails, and RBANS tests were all administered. Our coders coded this case three separate times – once while attending to how the court treated the Beery VMI, once for Trails, and once for RBANS. This left a total of 218 *unique* legal cases for us to code. In total, accounting for cases in which relevant batteries of tests were administered, 296 cases were coded for the final sample.

Coders first identified whether a legal admissibility challenge was raised to the neuropsychological tool in question. If a challenge was raised, coders summarized the basis of the challenge, including "fit" or relevance to the legal issue, reliability or

validity, general or peer acceptance, qualifications of the expert, helpfulness, prejudicial impact, or another issue. Coders indicated whether the evidence was ruled inadmissible (which we considered a "successful" admissibility challenge; see Appendix C for full coding guide).

Results

Regarding descriptive statistics for the 218 unique legal cases we coded, most of the cases included in the review were federal (91.7%) and heard at the trial level (89.4%). The majority of cases (88.1%) involved civil issues, and out of these, most (90.6%) related to Social Security or other disability benefits. Criminal cases most often pertained to questions of competence (50.0% of the criminal cases in the sample) or habeas relief (26.9%).

Interrater reliability was evaluated both by assessing for percent agreement among ratings, and also through a conservative reliability statistic called Krippendorff's alpha, which is robust for any level of measurement, any number of interchangeable raters, various sample sizes, and missing data (Hayes & Krippendorff, 2007). With regard to the percent agreement, coders initially agreed on whether a challenge was raised in 91.9% of cases. However, the Krippendorff statistic, with 5,000 bootstrapped estimates and treating the data nominally, showed a poor rate of agreement ($\alpha = .046$). This low Krippendorff's alpha variable is likely due to the base rate of admissibility challenges being very low, so even rare discrepancies between raters weighed heavily in the reliability statistic. Given that our percent agreement was high, and that the issue we coded for was a low base-rate event, we moved forward with consensus meetings to resolve discrepancies. Those discrepancies were also reviewed by an attorney for accuracy. The results section below

reports the consensus data after the meetings to resolve discrepancies, and after the attorney reviewed for accuracy. Given the percent agreement, consensus process, and attorney reviewer, we believe that the results of this project are sufficiently reliable for drawing meaningful inferences.

Legal admissibility challenges were raised to neuropsychological assessment tools in none of the 296 cases included in the analysis. In contrast, challenges were raised in 5.1% of cases using non-neuropsychological tools (n = 372) in Neal and colleagues' original review. Neuropsychological tools were thus challenged on the basis of their legal admissibility significantly less frequently than non-neuropsychological tools, $t(371) = -4.47$, $p < .001$, $d = -0.31$.

Discussion

We expected that legal admissibility challenges to neuropsychological assessment evidence would be raised infrequently by attorneys. Out of the 296 legal cases included in the sample, not a single admissibility challenge was brought forward. This was significantly lower than the challenge rate observed by Neal and colleagues' prior review (2019) of psychological assessment evidence in court, where only 19 challenges occurred out of 372 legal cases. Their review included three neuropsychological tools, but these challenges were raised exclusively to the non-neuropsychological tools examined.

Many of the exemplar tools included in our current case law analysis objectively did not meet *Daubert* and related admissibility criteria. For example, the ANIMALS test is not generally accepted in the psychological community and there are no reviews of its psychometric quality available. Yet, we observed no admissibility challenges to this test or to any of the questionable neuropsychological tools included in our review.

There are several reasons why neuropsychological tools might be challenged less frequently than their non-neuropsychological counterparts. First, neuropsychologists often use a broad number of distinct tools in their evaluations, and it might be difficult for attorneys to pinpoint specific tools which have characteristics that may warrant an admissibility challenge when these larger batteries of assessments are introduced. If other non-neuropsychologists introduce a limited number of tools in their evaluations, it might be easier for attorneys to discern where to direct their time and attention and raise an admissibility challenge. In addition, it is possible that attorneys and judges are vulnerable to the strong reputation of neuropsychological evidence and generally uncritical of its potential shortcomings. The title alone of this category of evidence (i.e., "neuro-") might contribute to an overall deficit of scrutiny in comparison to other types of evidence. However, as supported by the results of Study 1, there is clear evidence that not all neuropsychological assessment tools are generally accepted in their field and not all have good psychometric properties. Therefore, a critical examination of the tools that are introduced as evidence in court is necessary. Future research might survey judges and attorneys about their perceptions of the strengths and limitations of neuropsychological assessment evidence to better gauge whether the phenomenon of neuroenchantment is active, and if so, how best to temper it and encourage judges and attorneys to assign import to the power of the science rather than the power of the prefix.

A limitation we encountered in this case law analysis is the variation in what might be considered a formal admissibility challenge to the evidence proffered by psychological experts. Prior to consensus meetings and review by an attorney, our coders disagreed on whether an admissibility challenge occurred in 24 of the 296 legal cases. In

nearly all of these instances, a discussion arose in the case regarding the *weight* of the evidence, but not the *admissibility* of the evidence. In some situations, a challenge to the *weight* of the evidence resulted in the judge attributing zero weight to the testimony of the psychologist who conducted the evaluation. This, however, is a distinct process from that of an *admissibility* challenge, where the psychological evidence is explicitly barred from the courtroom. In resolving our coding discrepancies, we focused strictly on the *admissibility* of the evidence, regardless of whether the evidence was given no weight by the judge. In future case law research investigating legal challenges, it may be valuable to consider both the weight of the evidence as well as its legal admissibility, as there is practical importance to knowing when this type of evidence is discounted in legal settings.

We assessed interrater reliability in two ways: 1) Krippendorff's alpha and 2) percent agreement between coders. Krippendorff's alpha is a robust measure of reliability and can handle multiple coders, various levels of measurement, and missing data (Krippendorff, 2017). However, our rating of interrater reliability using Krippendorff's alpha was quite low. The original alpha was 0.046, and fell to -0.003 after resolving discrepancies. When data is binary and one of the categories is rarely present, Krippendorff's alpha can return very low regardless of the agreement between coders (De Swert, 2012). This was evident in our circumstances where admissibility challenges to neuropsychological evidence were coded for very infrequently. The percent agreement between raters, however, was high. Originally, coders agreed on 91.9% of cases. After consensus meetings and review by an attorney, the percent agreement between rose to 99.0%. With this in mind, we proceeded with our analyses and interpretations.

Because we did not come across any legal admissibility challenges, we are unable to provide data on the success rate of these challenges. However, prior research demonstrates that admissibility challenges to psychological assessment evidence are rare and their success even rarer (e.g., de Vogel et al., 2022; Neal et al., 2019). We might therefore extrapolate that admissibility challenges to neuropsychological evidence are unlikely to be successful, but we do not have data to support whether this is indeed the case. It is also possible that attorneys might be deterred from proposing an admissibility challenge to this type of evidence if they anticipate a low likelihood of success. Rather than depleting their already limited time and resources to raise a challenge which is unlikely to succeed, it might instead be more strategic for attorneys to focus on other aspects of the case more amenable to their persuasion and control. Further research is necessary to elucidate what processes specifically are driving the low rate of challenges to psychological assessment evidence.

The lack of admissibility challenges to neuropsychological assessment evidence suggests that judges and attorneys may be overlooking the quality of the assessment tools psychologists use in their evaluations or may not understand the potential for (or consequences of) admitting poor psychological science in court. This problem does not originate in the courtroom, but rather with the psychologists who select which tools to use in their evaluations. Do psychologists scrutinize the quality of the tools they choose to use in their evaluations before it enters the arena of attorneys and judges? In Study 3, we surveyed forensic psychologists about their practices in selecting assessment tools as well as their experiences with legal admissibility challenges.

Study 3: Exploring Forensic Psychologists' Assessment Practices

Are there differences in how thoroughly experts vet the psychological instruments they choose to use in their evaluations? We expected that experts who use neuropsychological tools in their evaluations would put more effort into preparing and use more resources when selecting tools to use in comparison to those who use non-neuropsychological tools. We also anticipated that experts who use neuropsychological tools would report experiencing fewer admissibility challenges, in line with our previous hypotheses.

Method

Participants. Participants for this study were contacted in three separate waves, the first two of which were collected for a previous student's master's thesis (Line, 2020). In the original effort, participants were contacted from our lab's database of forensic mental health professionals which at the time contained 2,685 psychologists. First, 700 invitations including a one dollar bill were mailed out to randomly selected individuals, followed two weeks later by a postcard reminder. This resulted in 57 completed surveys. Next, email invitations were sent to the 797 professionals for whom there was an email address in the database, excluding those who had already been sent a mail invitation, followed two weeks later by an email reminder. This resulted in an additional 62 completed responses. In the present recruitment effort, email invitations for an online survey were sent to a new subset of our lab's database with contact information for an additional 2,382 psychologists, followed two weeks later by a reminder email. None of this third batch were included in either of the first two waves of data collection.

In total, 227 practicing licensed psychologists specializing in forensic psychology (*N* = 116 from 2020; *N* = 111 from 2022) participated in the survey. Those who were not licensed to practice psychology in the United States were excluded from analyses. The mean age of participants was 58.61 years (*SD* = 13.38 years). Participants had an average of 22.99 years of experience (*SD* = 13.38 years) in the field and practiced across 44 states. The sample was 42% female, 57% male, and one participant identified as another gender. Fifteen percent of the sample was board certified in forensic psychology (ABPP). An additional item was added to the 2022 survey asking whether participants were board certified in neuropsychology (ABCN). Only three participants (*N* = 111) were board certified in neuropsychology, and none were certified in both disciplines.

Procedure and Materials. Participants completed an informed consent form upon accessing the survey (see Appendix D for IRB approval and approved documents). Participants indicated whether they had ever experienced legal challenges about the psychological tools they used in their evaluations, as well as the justification for and outcome of those challenges. Participants also reported the types of information they look for when selecting an instrument to use in their evaluations as well as whether they consider legal admissibility criteria. In addition, participants indicated whether they believed that other evaluators, judges, and lawyers could detect when psychological assessment evidence meets legal admissibility criteria (see Appendix E for complete survey).

Participants were asked the following question regarding their use of assessment tools in their evaluations:

"Think back to the last tool you used for a forensic evaluation. What was it? If the evaluation included multiple tools, please enter only the last tool administered for that evaluation."

We focused specifically on participants' most recent evaluation for several reasons. Asking participants about a specific instance rather than their general practices has the primary advantage of limiting bias in recollection. Thus, participants' responses might reflect their practices more accurately, rather than their *perceptions* of their practices, and may be less susceptible to errors in memory which could have arisen from inquiring about past experiences spanning across a longer frame of time. In addition, the original waves of data collection were not specific to neuropsychological assessment, so this item allowed the inclusion of prior participants' responses in analyses. Finally, questioning participants about their use of specific assessment tools enabled us to maintain the same classification scheme as the prior two studies in this series of research regarding whether we considered the tools to be neuropsychological or not. However, this approach also has limitations. It is possible, and perhaps likely, that participants' most recent evaluations were not reflective of their typical cases. As such, we may have overlooked relevant information not captured by this specific instance.

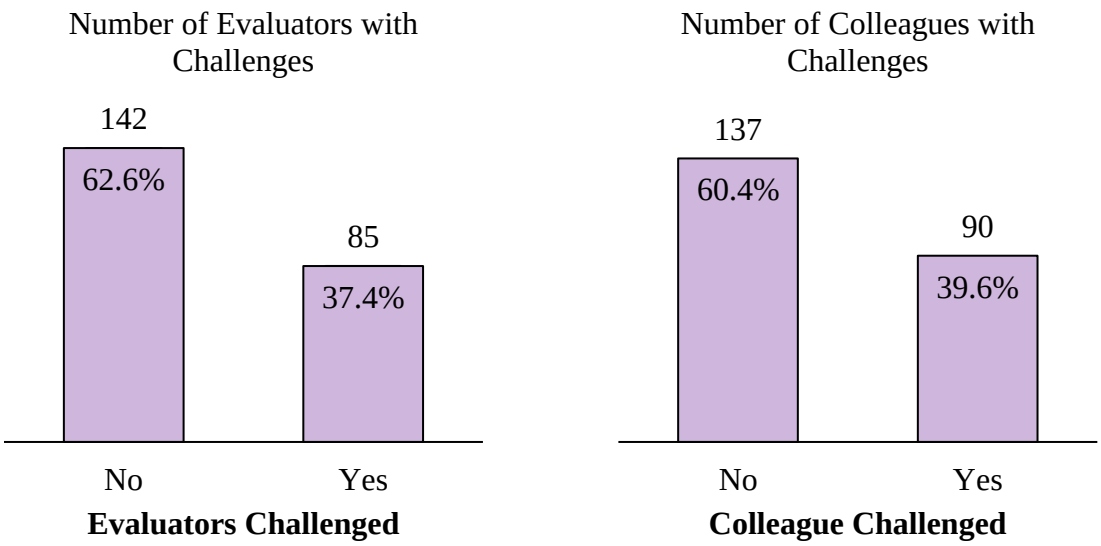
Responses to this item were classified as either neuropsychological or non-neuropsychological tools according to the same coding scheme used in the two studies described previously. The rates of challenges experienced by evaluators were compared to determine whether neuropsychological evaluators generally fare better or worse than non-neuropsychological evaluators. Differences in the preparation techniques implemented by psychologists and other assessment practices were also examined.

Results

The results described below are primarily descriptive. Psychologists were asked to report whether they had ever experienced a *Daubert*-related admissibility challenge to their assessment evidence in court. In the overall sample, about 37% of evaluators indicated that they had ever experienced an admissibility challenge, and about 49% reported that they knew of a colleague who had experienced an admissibility challenge (see Figure 4).

Figure 4

Admissibility Challenges to Evaluators and Their Colleagues

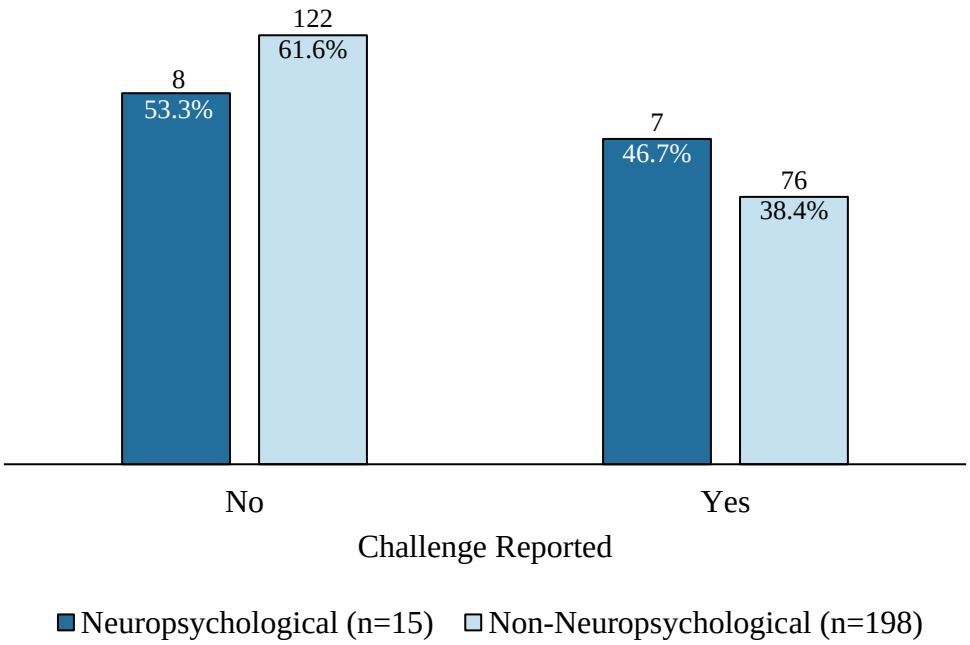


Note. This figure describes the number of challenges reported by mental health evaluators (left) and the number of evaluators who reported knowing a colleague who had experienced an admissibility challenge (right). Percentages and frequencies are displayed. $N = 227$.

We then considered whether the rate of admissibility challenges might differ between evaluators who used neuropsychological tools in their most recent assessment versus those who used non-neuropsychological tools. Fifteen responses were classified as neuropsychological tools and 198 as non-neuropsychological tools. Approximately 38% of those who used non-neuropsychological tools reported having ever experienced a legal admissibility challenge to their assessment. Of those who used neuropsychological tools, a higher proportion (47%) reported having ever experienced a challenge (see Figure 5).

Figure 5

Evaluators' Report of Experiencing Admissibility Challenges to Psychological Assessment Evidence

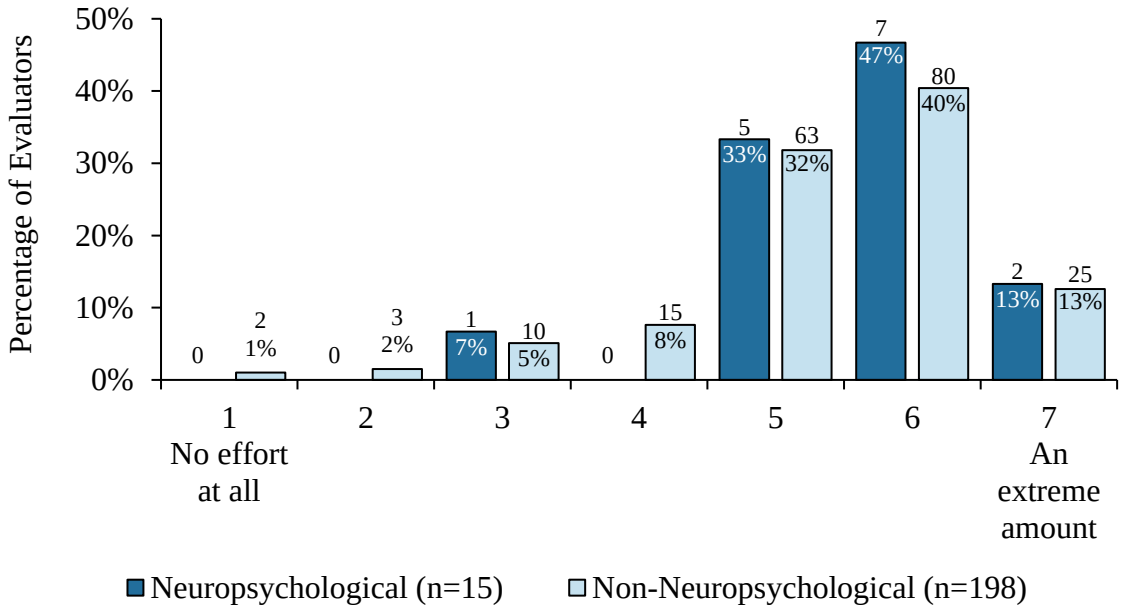


Note. This figure demonstrates the frequency of admissibility challenges evaluators reported having ever experienced between those who used a neuropsychological tool in their most recent evaluation (n = 15) versus those who last used a non-neuropsychological tool (n = 198). Relative percentages and frequencies are displayed.

Overall, participants described putting a moderate to high amount of effort into learning about a tool before using it in a forensic evaluation, reporting an average of 5.4 on a Likert scale ranging from 1 (no effort at all) to 7 (an extreme amount of effort). The distribution of effort that evaluators who used neuropsychological tools reported appeared similar to that of evaluators who used non-neuropsychological tools (see Figure 6).

Figure 6

Amount of Effort Preparing to Administer a Tool for the First Time

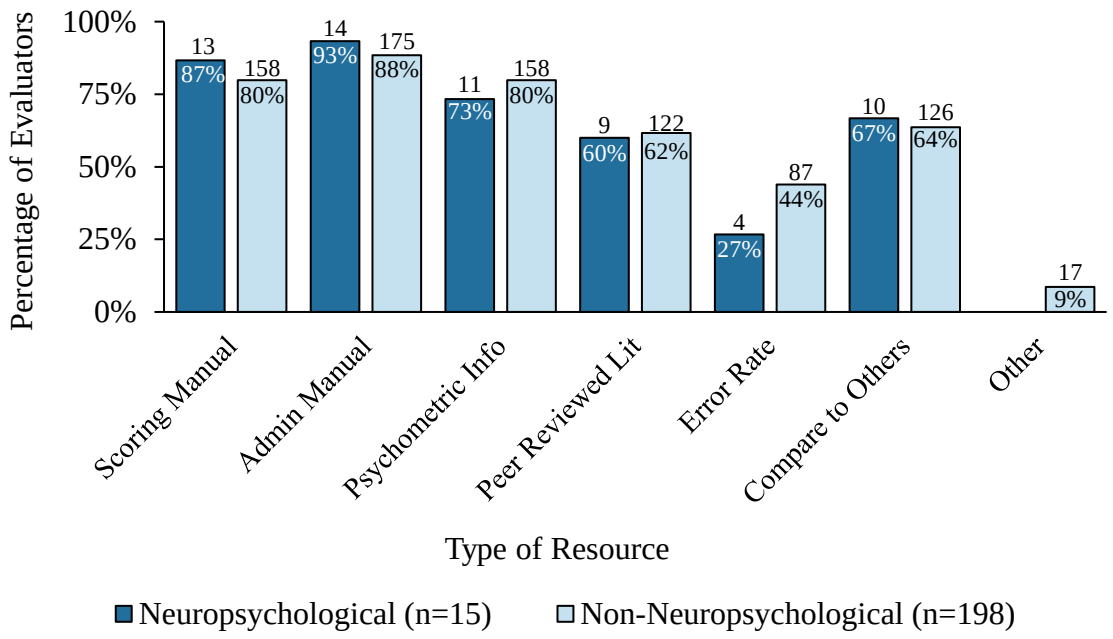


Note. This figure describes the amount of effort that evaluators who used a neuropsychological tool in their most recent evaluation (n = 15) reporting putting into preparing to use a tool for the first time in a forensic evaluation in comparison to those who used non-neuropsychological tools (n = 198). Relative percentages and frequencies are displayed.

Participants were asked to think back to what resources they used to prepare before administering the tool they were reporting on in this survey for the first time. These techniques included: reading the scoring manual, reading the administration manual, reviewing psychometric information about the tool, reading peer reviewed literature about the tool, considering the error rate of the tool, and comparing it to other available tools (see Figure 7). Participants could also choose an "other" option and provide additional information about their preparatory techniques. Overall, participants reported using an average of 4 types of resources (SD = 1.75) to prepare before administering a tool for the first time. Those who used neuropsychological tools in their evaluations appeared act similarly compared to those who used non-neuropsychological tools with regard to the number preparation techniques before administering an assessment tool (see Figure 8).

Figure 7

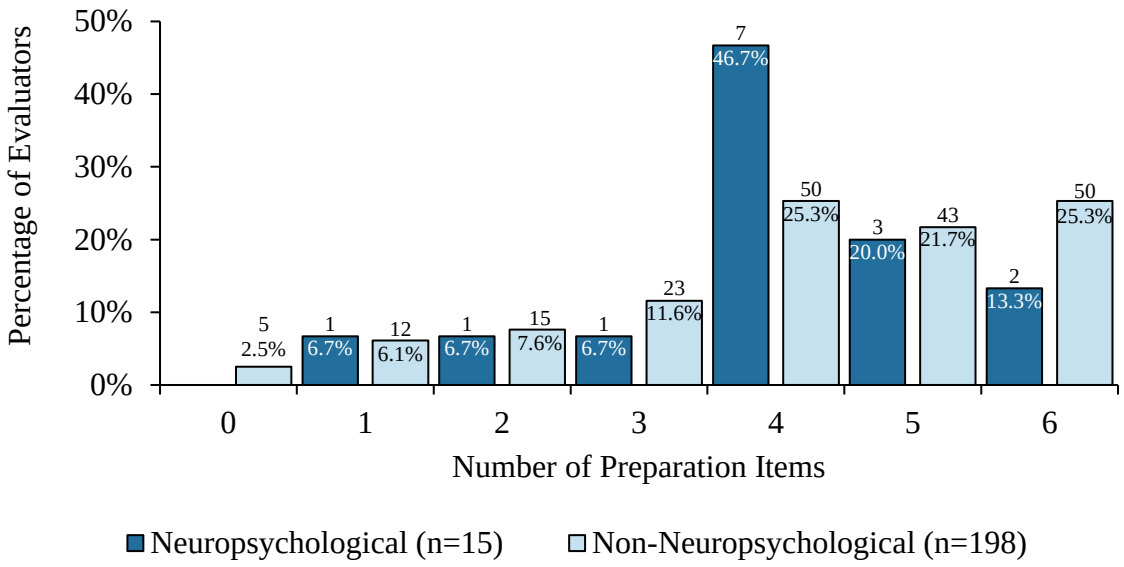
Types of Preparation Techniques Before Administering Neuropsychological versus Non-Neuropsychological Tools for the First Time in an Evaluation



Note. This figure displays the methods which evaluators reported using to prepare before administering a psychological assessment tool for the very first time between those who used a neuropsychological tool in their most recent evaluation (n = 15) versus those who last used a non-neuropsychological tool (n = 198). These techniques include: reading the scoring manual, reading the administration manual, reading psychometric information about the tool, reading peer reviewed literature about the tool, reviewing the tool's error rate, comparing the tool to other available options, or other methods not listed. Relative percentages and frequencies are displayed.

Figure 8

Number of Preparation Techniques Before Administering Neuropsychological versus Non-Neuropsychological Tools for the First Time in an Evaluation



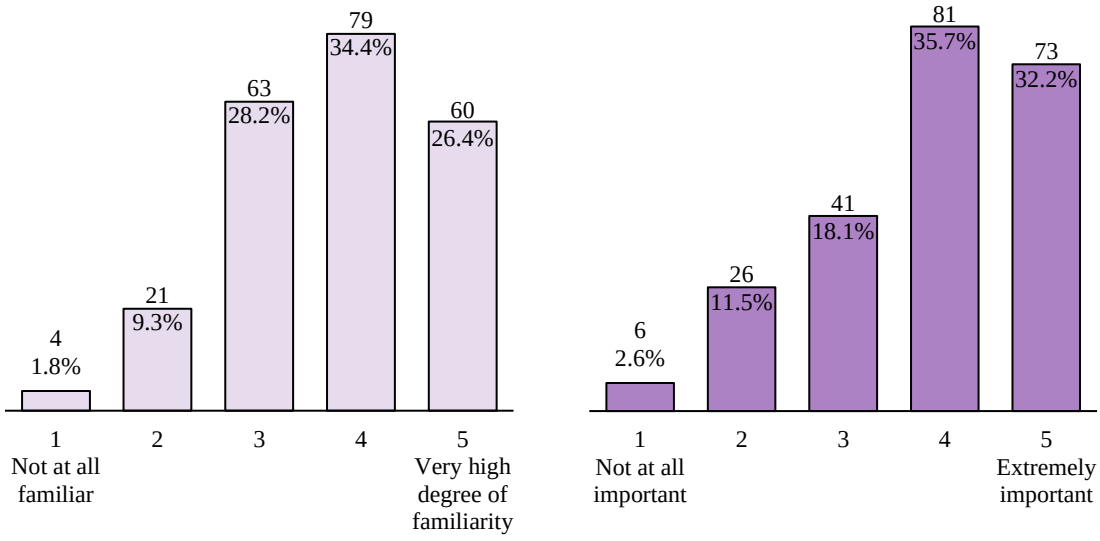
Note. This figure displays the number of techniques evaluators reported using to prepare before administering a psychological assessment tool for the very first time between those who used a neuropsychological tool in their most recent evaluation (n = 15) versus those who last used a non-neuropsychological tool (n = 198). These techniques include: reading the scoring manual, reading the administration manual, reading psychometric information about the tool, reading peer reviewed literature about the tool, reviewing the tool's error rate, comparing the tool to other available options, or other methods not listed. Relative percentages and frequencies are displayed.

Participants were asked about their familiarity with legal admissibility criteria and how important they believe these standards to be. Most participants had a high degree of

familiarity with legal admissibility criteria (M = 3.74, SD = 1.01) and believed that admissibility criteria were very important (M = 3.83, SD = 1.08). However, some participants reported having no familiarity with legal admissibility criteria, and some reported believing that admissibility criteria were not at all important (see Figure 9).

Figure 9

Ratings of Familiarity with and Importance of Legal Admissibility Criteria



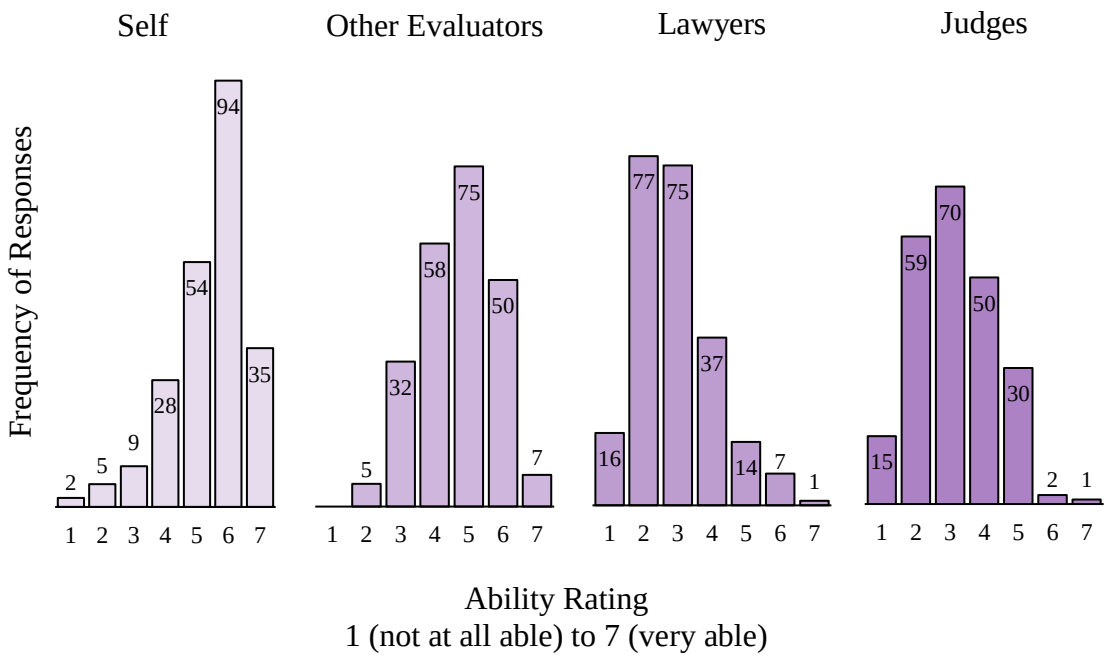
Note. This figure displays evaluators' reports of how familiar they are with legal admissibility criteria (left) and how important they believe admissibility criteria are (right). Relative percentages and frequencies are displayed. N = 227.

Participants were asked to describe their ability to detect which psychological assessment tools meet legal admissibility criteria. Ratings ranged from 1 (not at all able) to 7 (very able). Evaluators credited themselves with the highest ability to identify tools

which satisfy admissibility standards (*M* = 5.42, *SD* = 1.22), superseding their perceptions of their colleagues' abilities to do the same (*M* = 4.68, *SD* = 1.12). Evaluators were the most critical of lawyers' ability to detect tools which meet legal admissibility criteria (*M* = 2.92, *SD* = 1.17). Evaluators ascribed judges (*M* = 3.14, *SD* = 1.19) with a higher ability to identify tools which fulfill admissibility criteria than lawyers, but less than that of themselves and their colleagues (see Figure 10).

Figure 10

Ability to Detect Which Psychological Assessment Tools Meet Admissibility Criteria



Note. This figure demonstrates ratings of ability to identify psychological assessment tools which meet legal admissibility criteria on a Likert scale from 1 (not at all able) to 7

(very able). Evaluators rated their own ability (far left), the ability of other evaluators (left), lawyers (right), and judges (far right). Frequencies are displayed. $N = 227$.

Discussion

In light of the small sample size of evaluators who used neuropsychological tools, we focus here on the results of descriptive analyses. About 37% of the evaluators in our sample reported having ever experienced an admissibility challenge to their psychological assessment evidence in court, and about 40% reported knowing of a colleague who had experienced a challenge. These percentages were higher than we expected from the rates of challenges observed in case law by Neal and colleagues (5.1%; 2019) and from our case law analysis in Study 2 (0%). However, our sample had over two decades of experience on average, and we did not ask participants to specify how *often* they experienced legal admissibility challenges, which may explain the discrepancy.

A larger percentage of those who used neuropsychological tools in their evaluations reported having ever experienced a legal admissibility challenge than those who used non-neuropsychological tools. Nearly half (46.7%) of evaluators who used neuropsychological tools had experienced a challenge, while only 38.4% of those who used non-neuropsychological tools had. This contradicts our finding in Study 2 that neuropsychological tools were challenged significantly less frequently than non-neuropsychological tools. It is possible that, although we asked participants about *admissibility* challenges, they may have also disclosed challenges to the *weight* of their evidence. This technical distinction should be further explored in future research, along with the consequences and impact of both types of challenges.

The amount of effort, number of techniques, and types of resources that evaluators reported using while preparing to administer a psychological assessment tool for the first time did not appear to differ substantially between those who used neuropsychological tools versus those who used non-neuropsychological tools. We had expected that neuropsychologists might put forth more effort and prepare more rigorously due to more pervasive norms for board certification in the field of neuropsychology than in forensic psychology as a whole, but this did not appear to be the case. Evaluators using both types of tools appeared to use a similar number of preparation items. However, while nearly half of those who used neuropsychological tools reported using 4 techniques to prepare, those who used non-neuropsychological tools were more evenly distributed in the number of preparation items they reported using. The types of resources participants reported using were also very similar between those who used neuropsychological versus non-neuropsychological tools, but a smaller percentage of those who used neuropsychological tools reported that they examined the error rate of a tool prior to use.

We asked participants about their familiarity with legal admissibility criteria and how important they believed these criteria to be. Most participants reported a having a high degree of familiarity and believed that these criteria were very important. However, some participants reported not being familiar with admissibility criteria, and some reported that they did not think these criteria were at all important. This is concerning, as legal admissibility criteria are set in place to protect the fairness and validity of the justice system. Weak and unreliable scientific techniques should not have a place in the courtroom, and admissibility guidelines are intended to protect this. The disparity between some psychologists' attitudes could be reflective of differences between this

idealized principle underlying admissibility criteria and the reality of how inconsistently standards are applied in cases, but their existence serves a valuable purpose in regulating the quality of evidence admitted in court.

We also found support for the presence of a bias blind spot in forensic mental health evaluators. With this cognitive bias, people perceive themselves to be less biased than their peers (Pronin et al., 2002). Evaluators rated their own ability to detect whether assessment tools met admissibility criteria higher than the ability of other evaluators, lawyers, and judges. Participants rated other evaluators moderately high on their ability to identify whether tools meet admissibility criteria (and higher than lawyers and judges), but not as high as their own ability to do so. Evaluators rated the abilities of judges and especially of lawyers to assess whether tools fulfill legal admissibility standards poorly overall, with few evaluators reporting that they believed judges and lawyers to be very proficient in completing this task.

General Discussion

The current project provides novel information empirically examining the quality of neuropsychological assessment evidence in court as well as its legal admissibility. Other lines of research have supported that judges and lawyers lack the necessary scientific expertise to accurately evaluate scientific evidence (e.g., DeMatteo et al., 2019; Gatowski et al., 2001; National Research Council, 2009), and ours supports a similar trend with neuropsychological assessment evidence specifically. Neuropsychological tools appeared to perform better than non-neuropsychological tools in some ways (e.g., higher rate of testing), but worse in others (e.g., more mixed reviews of psychometric

quality and fewer favorable reviews). As expected, we observed clear variation in the quality of neuropsychological tools that experts reported using in forensic contexts.

Encouragingly, we saw a low number of cases in which some of these questionable tools were introduced as evidence. For example, the National Adult Reading Test (NART) has no evidence of general acceptance and mixed reviews of psychometric quality, but was not cited in any legal cases in the last three calendar years. However, some tools with no evidence of general acceptance and no professional reviews (e.g., the Tactile Form Recognition Test [TFR], Test of Pre-Morbid Functioning [TOPF]) were included in forensic evaluations admitted in legal settings. It is unclear how the courts should treat tools which lack this information, but they should at the very least be scrutinized rather than accepted without question.

We observed a pattern consistent with the bias blind spot regarding psychologists' perceptions of their own ability to identify tools which meet admissibility criteria in comparison to other evaluators, attorneys, and judges. It *could* be true that psychologists are in fact better able to identify these tools. However, across the results of Studies 2 and 3, we find that, despite rating themselves higher in ability to determine which tools meet *Daubert*-related admissibility criteria, psychologists are still choosing to use assessments in their evaluations which do not.

The complete lack of admissibility challenges in our case law analysis in Study 2 suggests that the adversarial process is not sufficient to guard against the admission of questionable evidence. Neuropsychological tools were challenged less than non-neuropsychological tools, which might indirectly suggest that judges and attorneys are vulnerable to the phenomenon of neuroenchantment, or perhaps less critical of this

evidence based on its reputation. An intervention in the justice system to increase scientific literacy and bolster the strength of the courts could therefore be useful. This might be achieved by introducing a basic science and statistics courses in law schools or offering relevant continuing education credits to judges and attorneys. If this type of training renders attorneys more sensitive to the quality of neuropsychological evidence, this would be a reasonable direction to pursue to improve the future of the legal system. Indeed, some jurisdictions already hold court-appointed scientific advisors who aid judges in understanding scientific evidence (DeMatteo et al., 2019).

Playing the role of gatekeeper is a monumental responsibility. When this fails and poor science is admitted in court, there can be significant consequences. Foremost, forensic psychologists have an obligation to use strong scientific methods and select high-quality assessment tools for use in their evaluations. Not doing so undermines the credibility of psychology as a field and threatens the legitimacy of the justice system. The onus too lies on attorneys and judges to screen the evidence which enters the courtroom. In principle, weak evidence should be routinely challenged based on its reliability and validity, and ultimately barred from admission. It is clear that improvement in these domains is necessary.

Overall, our results aid in understanding how forensic psychologists' practices align with patterns of quality of neuropsychological versus non-neuropsychological tools admitted in court. The series of studies suggests that attorneys and judges may not be accurately evaluating neuropsychological evidence or effectively gatekeeping it from entering the courtroom. Given the significance of the decisions which are informed by these assessments, it is critical that psychologists, attorneys, and judges alike scrutinize

the quality of tools used in evaluations and adopt sound practices to ensure that this type of evidence is reliable, valid, and fair.

REFERENCES

Ali, S. S. (2012). *Neuroenchantment as a lens: A critical look at sense of control, authorship, and conscious will.* [Master's thesis, McGill University]. eScholarship@McGill.

Ali, S. S., Lifshitz, M., & Raz, A. (2014). Empirical neuroenchantment: From reading minds to thinking critically. *Frontiers in Human Neuroscience, 8*(357), 1-4. https://doi.org/10.3389/fnhum.2014.00357

American Psychological Association. (2008). *Clinical neuropsychology.* https://www.apa.org/ed/graduate/specialize/neuropsychology

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87*, 84–94. https://doi.org/10.1207/s15327752jpa8701_07

Bechara, A. (2007). *Iowa gambling task professional manual*. Psychological Assessment Resources, Inc.

Beery K. E., Buktenica N. A., & Beery N. A. (2010). *The Beery–Buktenica developmental test of visual–motor integration: Administration, scoring, and teaching manual* (6th ed.). Pearson.

Bender, L. (2003). *Bender visual-motor gestalt test* (2nd ed., G. Brannigan & S. Decker, Eds.). Western Psychological Services.

Benedict, R. H. B. (1997). *Brief visuospatial memory test - revised: Professional manual.* Psychological Assessment Resources, Inc.

Benton, A. L., Hamsher, K., & Sivan, A. B. (1994). *Multilingual aphasia examination* (3rd ed.). Psychological Assessment Resources, Inc.

Benton-Sivan, A. (1991). *Benton visual retention test* (5th ed.). Pearson.

Brodsky, S. L., Neal, T. M. S., Cramer, R. J., Ziemke, M. H. (2009). Credibility in the courtroom: How likeable should an expert witness be? *Journal of the American Academy of Psychiatry and the Law, 37*(4), 525-532.

Carlson, J. F., Geisinger, K. F., & Jonson, J. L. (Eds.). (2017). *The twentieth mental measurements yearbook.* Buros Center for Testing.

Chorn, J. A. & Kovera, M. B. (2019). Variations in reliability and validity do not influence judge, attorney, and mock juror decisions about psychological expert evidence. *Law and Human Behavior, 43*(6), 542-557. https://doi.org/10.1037/lhb0000345

Cizek, G. J., Koons, H. K., & Rosenberg, S. L. (2012). Finding validity evidence: An analysis using *Mental Measurements Yearbook*. In J. A. Bovaird, K. F. Geisinger, & C. W. Buckendahl (Eds.), *High-stakes testing in education: Science and practice in K-12 settings* (pp 119–138).

Cohen, D., Sepehry, A. A. & Schultz, I. Z. (2019). Forensic neuropsychological aspects of competency evaluations: Financial and legal competency in older adults. *Psychological Injury and Law, 13*, 19-32. https://doi.org/10.1007/s12207-019-09364-8

Damon-Moore, S. L. (2017). Trial judges and the forensic science problem. *New York University Law Review, 92*(5), 1532-1570.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 43 F.3d 1311 (9th Cir. 1995).

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

De Swert, K. (2012). Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha. *Center for Politics and Communication, 15*, 1-15.

de Vogel, V., De Beuf, T., Shepherd, S., & Schneider, R. D. (2022). Violence risk assessment with the HCR-20^{V3} in legal contexts: A critical reflection, *Journal of Personality Assessment,* 104(2), 252-264, https://doi.org/10.1080/00223891.2021.2021925

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California verbal learning test* (2nd ed.). Pearson.

DeMatteo, D., Fishel, S., & Tansey, A. (2019). Expert evidence: The (unfulfilled) promise of *Daubert*. *Psychological Science in the Public Interest, 20*(3), 129–134. https://doi.org/10.1177/1529100619894336

Douglas, K. S., Ogloff, J. R. P., Nicholls, T. L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening Version. *Journal of Consulting and Clinical Psychology, 67*(6), 917–930. https://doi.org/10.1037/0022-006X.67.6.917

Dunn, L. M., & Dunn, D. M. (2007). *The Peabody picture vocabulary test* (4th ed.). Pearson.

Edens, J. F., & Boccaccini, M. T. (Eds.). (2017). Field reliability and validity of forensic psychological assessment instruments and procedures [Special issue]. *Psychological Assessment, 29*(6). https://doi.org/10.1037/pas0000475

Elhai, J. D., Gray, M. J., Kashdan, T. B., & Franklin, L. C. (2005). Which instruments are most commonly used to assess traumatic event exposure and posttraumatic effects? A survey of traumatic stress professionals. *Journal of Traumatic Stress, 18,* 541–545. https://doi.org/10.1002/jts.20062

Ewing, C. P., & McCann, J. T. (2006). *Minds on trial: Great cases in law and psychology.* Oxford University Press.

Faigman, D. L., Saks, M. J., Sanders, J., & Cheng, E. K. (2008). *Modern scientific evidence: The law and science of expert testimony* (2007-2008 ed.). Thomson-West.

Faust, D., & Ziskin, J. (1988). The expert witness in psychology and psychiatry. *Science, 241*(4861), 31–35. https://doi.org/10.1126/science.3291114

Fed. R. Evid. 702. (2020). *Testimony by expert witnesses.* Retrieved from: https://www.rulesofevidence.org/articleiv/rule-702/

Fisher, S. (1988). In search of the virtuous prosecutor: A conceptual framework. *American Journal of Criminal Law, 15*(197), 197-261.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state:" A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research, 12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences.* Macmillan and Co. https://doi.org/10.1037/13474-000

Galton, F. (1879). Psychometric experiments. *Brain, 2*(2), 149-162. https://doi.org/10.1093/brain/2.2.149

Gatowski, S. I., Dobbin, S. A., Richardson, J. T., Ginsburg, G. P., Merlino, M. L., & Dahir, V. (2001). Asking the gatekeepers: A national survey of judges on judging expert evidence in a post-*Daubert* world. *Law and Human Behavior, 25*(5), 433–458. https://doi.org/10.1023/A:1012899030937

General Electric Co. v. Joiner, 522 U.S. 136 (1997).

Gioia, G. A., Isquith, P. K., Guy, S. C., Kenworthy, L. (2015). *Behavior rating inventory of executive function* (2nd ed.). Psychological Assessment Resources, Inc.

Golden, C. J., Purishch, A. D., & Hammeke, T. (1985). *Manual for the Luria-Nebraska neuropsychological battery.* Western Psychological Services.

Gowensmith, W. N., Pinals, D. A., & Karas, A. C. (2015). States' standards for training and certifying evaluators of Competency to Stand Trial. *Journal of Forensic Psychology Practice, 15*(4), 295–317. https://doi.org/10.1080/15228932.2015.1046798

Greenberg L. M., Holder C., Kindschi C. L., Dupuy T. R. (2018). *Test of variables of attention (T.O.V.A.): Clinical manual.* The TOVA Company.

Grisso, T. (2003). *Evaluating competencies: Forensic assessments and instruments* (2nd ed.). Kluwer Academic.

Gronwall, D. M. A. (1977). Paced auditory serial-addition task: A measure of recovery from concussion. *Perceptual and Motor Skills, 44*, 367-373.

Guarnera, L. A., Murrie, D. C., & Boccaccini, M. T. (2017). Why do forensic experts disagree? Sources of unreliability and bias in forensic psychology evaluations. *Translational Issues in Psychological Science, 3*(2), 143-152. https://doi.org/10.1037/tps0000114

Halstead, W. C., & Wepman, J. M. (1949). The Halstead-Wepman aphasia screening test. *Journal of Speech & Hearing Disorders, 14*, 9–15. https://doi.org/10.1044/jshd.1401.09

Harvey, P. D. (2012). Clinical applications of neuropsychological assessment. *Dialogues in Clinical Neuroscience, 14*(1), 91–99. https://doi.org/10.31887/DCNS.2012.14.1/pharvey

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*, 77-89.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtis, G. (1993). *Wisconsin card sorting test manual: Revised and expanded.* Psychological Assessment Resources, Inc.

Hebben, & Milberg, W. (2009). *Essentials of neuropsychological assessment* (2nd ed.). John Wiley & Sons.

Hill, J. L. (2020, May). *The states of Daubert after Florida*. LexVisio. https://www.lexvisio.com/article/2019/07/09/the-states-of-daubert-after-florida

Holdnack, J. A., & Drozdick, L. W. (2009). *Advanced clinical solutions for WAIS-IV and WMS-IV: Clinical and interpretive manual*. Pearson.

Hooper, H. E. (1983). *Hooper visual organization test manual.* Western Psychological Services.

Institute for the Advancement of the American Legal System. (2015, August). *FAQs: Judges in the United States*. University of Denver. https://iaals.du.edu/sites/default/files/documents/publications/judge_faq.pdf

King, C., Wade, N., & Tillson, J. (2017, March). *Case law references as a big-picture snapshot of psychological test use in forensic mental health assessment* [Paper presentation]. American Psychology-Law Society, Seattle, WA.

Krippendorff, K. (2017). Intercoder reliability techniques: Krippendorff's alpha. In *Encyclopedia of communication research methods* (pp. 743–751). Sage Research Methods. https://doi.org/10.4135/9781483381411

Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).

LaDuke, C., Barr, W., Brodale, D. L., & Rabin, L. A. (2018). Toward generally accepted forensic assessment practices among clinical neuropsychologists: A survey of professional practice and common test use. *The Clinical Neuropsychologist, 32,* 145–164. https://doi.org/10.1080/13854046.2017.1346711

LaDuke, C., DeMatteo, D., Heilbrun, K., & Swirsky-Sacchetti, T. (2012). Clinical neuropsychology in forensic contexts: Practitioners' experience, training, and practice. *Professional Psychology: Research and Practice, 43*(5), 503-509. https://doi.org/10.1037/a0028161

Lally, S. J. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology: Research and Practice, 34*, 491–498. https://doi.org/10.1037/0735-7028.34.5.491

Lilienfeld, S. O., Aslinger, E., Marshall, J., & Satel, S. (2018). Neurohype: A field guide to exaggerated brain-based claims. In L. S. M. Johnson & K. S. Rommelfanger (Eds.), *The Routledge handbook of neuroethics* (pp. 241–261). Routledge/Taylor & Francis Group.

Lin, L., Christidis, P., & Stamm, K. (2017). A look at psychologists' specialty areas. *American Psychological Association*. https://www.apa.org/monitor/2017/09/datapoint

Line, E. C. N. (2020). *Experts screening experts: Are courts effectively gatekeeping psychological assessment evidence?* (Publication No. 28023363) [Master's thesis, Arizona State University]. ProQuest Dissertations Publishing.

Mart, E. G. (2016). Neuropsychological assessment of testamentary capacity and undue influence. *Archives of Clinical Neuropsychology, 31*(6), 554-561. https://doi.org/10.1093/arclin/acw048

McAuliff, B. D., Kovera, M. B., & Nunez, G. (2009). Can jurors recognize missing control groups, confounds, and experimenter bias in psychological science*? Law and Human Behavior, 33*(3), 247–257. https://doi.org/10.1007/s10979-008-9133-0

McLaughlin, J. L., & Kan, L. Y. (2014). Test usage in four common types of forensic mental health assessment. *Professional Psychology: Research and Practice, 45,* 128–135. https://doi.org/10.1037/a0036318

Melton, G. B., Petrila, J., Poythress, N. G., Slobogin, C., Otto, R. K., Mossman, D., & Condie, L. O. (2017). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (4th ed.). Guilford Press.

Miller, P. R., Dasher, R., Collins, R., Griffiths, P., Brown, F. (2001). Inpatient diagnostic assessments: 1. Accuracy of structured vs. unstructured interviews. *Psychiatry Research, 105*(2001), 255-264.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

National Research Council. (2009). *Strengthening forensic science in the United States: A path forward.* The National Academies Press. https://doi.org/10.17226/12589

Neal, T. M. S. & Grisso, T. (2014). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice and Behavior, 41,* 1406–1421. https://doi.org/10.1177/0093854814548449

Neal, T. M. S. (2020, March), *Calibration in court: Jurors' use of scientific information.* American Psychology-Law Society, New Orleans, LA.

Neal, T. M. S., Slobogin, C., Saks, M. J., Faigman, D. L., & Geisinger, K. F. (2019). Psychological assessments in legal contexts: Are courts keeping "junk science" out of the courtroom? *Psychological Science in the Public Interest, 20*(3), 135-164. https://doi.org/10.1177/1529100619888860

Nelson, H. E., & Willison, J. (1991). *National adult reading test (NART).* NFER-Nelson.

Neufeld, P. J. (2005). The (near) irrelevance of Daubert to criminal justice and some suggestions for reform. *American Journal of Public Health, 95*(1), S107–S113. https://doi.org/10.2105/AJPH.2004.056333

Nir, E., & Liu, S. (2021). What do the gatekeepers see? Perceptions and evaluations of scientific evidence among state court judges. *Criminology, Criminal Justice, Law & Society, 22*(1), 20-35.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

O'Brien, T. I. (2018). Beyond reliable: Challenging and deciding expert admissibility in U.S. civil courts. *Law, Probability, and Risk, 17*, 29-45. https://doi.org/10.1093/lpr/mgx010
pcast_forensic_science_report_final.pdf

President's Council of Advisors on Science and Technology. (2016). *Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods*. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin, 28*(3), 369–381. https://doi.org/10.1177/0146167202286008

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology, 20*, 33–65. https://doi.org/10.1016/j.acn.2004.02.005

Rachlinski, J. J., & Wistrich, A. J. (2017). Judging the judiciary by the numbers: Empirical research on judges. *Annual Review of Law and Social Science, 13*, 203–229. https://doi.org/10.1146/annurev-lawsocsci-110615-085032

Randolph, C. (2012). *Repeatable battery for the assessment of neuropsychological status update (RBANS Update).* Pearson.

Redding, R. E., Floyd, M. Y., & Hawk, G. L. (2001). What judges and lawyers think about the testimony of mental health experts: A survey of the courts and bar. *Behavioral Sciences & the Law, 19*(4), 583-594. https://doi.org/10.1002/bsl.455

Reitan, R. M., & Wolfson, D. (1994). The Halstead-Reitan neuropsychology test battery: Theory and clinical interpretation (2nd ed.). *Archives of Clinical Neuropsychology, 9*(3), 289–290.

Rothke, S. E., Demakis, G. J., & Amsbaugh, H. M. (2019). State statutes regarding the role of psychologists in performing capacity evaluations for guardianship determinations. *Professional Psychology: Research and Practice, 50*(4), 228-239. https://doi.org/10.1037/pro0000240

Ryba, N. L., Cooper, V. G., & Zapf, P. A. (2003). Juvenile competence to stand trial evaluations: A survey of current practices and test usage among psychologists. *Professional Psychology: Research and Practice, 34*, 499–507. https://doi.org/10.1037/0735-7028.34.5.499

Satel, S., & Lilienfeld, S. O. (2013). *Brainwashed: The seductive appeal of mindless neuroscience*. Basic Books.

Sbordone, R. J. & Long, C. J. (Eds.). (1996). *Ecological validity of neuropsychological testing*. GR Press.

Schweitzer, N. J., & Saks, M. J. (2009). The gatekeeper effect: The impact of judges' admissibility decisions on the persuasiveness of expert testimony. *Psychology, Public Policy, and Law, 15*(1), 1–18. https://doi.org/10.1037/a0015290

Segal, D. L., & Hersen, M. (Eds.). (2010). *Diagnostic interviewing* (4th ed.). Springer.

Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science, 20*(1), 38–42. https://doi.org/10.1177/0963721410397271

Slick, D. J., Tan, J. E., Strauss, E. H., & Hultsch, D. F. (2004). Detecting malingering: A survey of experts' practices. *Archives of Clinical Neuropsychology, 19*, 465–473. https://doi.org/10.1016/j.acn.2003.04.001

Smith, A. (1973). *Symbol digit modalities test.* Western Psychological Services.

Strauss, E., Sherman, E., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). Oxford University Press.

Wasserman, J. D., & Bracken, B. A. (2013). Fundamental psychometric considerations in assessment. In I. B. Weiner, J. R. Graham, & J. A. Naglieri (Eds.), *Handbook of psychology. Assessment psychology (*2nd ed., Vol. 10, pp. 50-81). John Wiley & Sons.

Wechsler, D. (2009). *Wechsler memory scale* (4th ed.). Pearson.

Zapf, P. A., Hubbard, K. L., Cooper, V. G., Wheeles, M. C., & Ronan, K. A. (2004). Have the courts abdicated their responsibility for determination of competency to stand trial to clinicians? *Journal of Forensic Psychology Practice, 4*(1), 27-44, https://doi.org/10.1300/J158v04n01_02

APPENDIX A

NEUROPSYCHOLOGICAL AND NON-NEUROPSYCHOLOGICAL TOOL CLASSIFICATIONS

Neuropsychological Tools:

ANIMALS/Animal Naming [verbal fluency test/task]
Austin Maze
BEERY VMI (Beery-Buktenica Developmental Test of Visual-Motor Integration)
Bender-Gestalt (Bender Visual-Motor Gestalt Test)
BNT (Boston Naming Test)
BRIEF (Behavior Rating Inventory of Executive Function)
BVMT (Brief Visuospatial Memory Test)
BVRT (Benton Visual Retention Test)
CAT/HCT/BCT (Category Test/Halstead Category Test/Booklet Category Test)
CAVLT (Children's Auditory Verbal Learning Test)
CCT (Children's Category Test)
Cognistat; NCSE (Neurobehavioral Cognitive Status Examination)
CogScreen
COWA/COWAT (Controlled Oral Word Association Test)
CVLT (California Verbal Learning Test)
CVLT-C (California Verbal Learning Test-Children's Version)
CVMT (Continuous Visual Memory Test)
D-KEFS (Delis-Kaplan Executive Function System)
FAS [FAS verbal fluency test/task]
Finger Tap/ping/Oscillation Test
GNDS (General Neuropsychological Deficit Scale) [for the Halstead-Reitan Neuropsychological Test Battery (HRNTB)]
Grip Strength (Hand Dynamometer)
Grooved Peg (Grooved Pegboard Test)
Halstead–Wepman Aphasia Screening Test
HRNTB (Halstead-Reitan Neuropsychological Test Battery)
IGT (Iowa Gambling Task)
KABC (Kaufman Assessment Battery for Children)
Lateral Dominance Examination
LNNB (Luria-Nebraska Neuropsychological Battery)
LNNB-C (Luria-Nebraska Neuropsychological Battery for Children)
MFD (Memory-for-Designs Test)
MMSE (Mini-Mental State Examination)
MoCA (Montreal Cognitive Assessment)
NAB / NAB-SM / S-NAB (Neuropsychological Assessment Battery)
NART (National Adult Reading Test)
PAL (Paired Associate/s Learning)
PASAT (Paced Auditory Serial Addition Test)
PPC (Psychological Processing Checklist)
PPVT (Peabody Picture Vocabulary Test)
RAVLT (Rey Auditory Verbal Learning Test)
RBANS (Repeatable Battery for the Assessment of Neuropsychological Status)
Reitan-Indiana Aphasia Screening Test
ROCF/RCFT (Rey–Osterrieth Complex Figure/Rey Complex Figure Test)
SCT (Short Category Test)
SDMT (Symbol Digit Modalities Test)
SPE (Reitan–Klove Sensory Perceptual Examination)
SRT (Seashore Rhythm Test)
SSPT (Speech Sounds Perception Test)
Stroop [e.g., Stroop Color and Word Test]
TFR (Tactile Form Recognition Test)
TOPF (Test of Pre-Morbid Functioning)
TOVA (Test of Variables of Attention)
TPT (Tactual Performance Test)

Trails A & B / TMT / CTMT (Trail Making Test / Comprehensive TMT) VOT (Hooper Visual Organization Test)	WCST / WCST-64 (Wisconsin Card Sorting Task) WMS (Wechsler Memory Scale)

Non-Neuropsychological Tools:

16pf (Sixteen Personality Factor Questionnaire) 21-Item Test AAMD ABS (American Association on Mental Deficiency Adaptive Behavior Scales) AAPI (Adult Adolescent Parenting Inventory) AASI (Abel Assessment for sexual interest) ABAS (Adaptive Behavior Assessment System) ABCS (Abel and Becker Cognition Scale) Abel and Becker Card Sort ACCESS (A Comprehensive Custody Evaluation Standard System) ACS (Advanced Clinical Solutions for the WAIS-IV and WMS-IV) ACUTE 2007 ADHD Rating Scale (ADHD-RS) ADS (Anger Disorders Scale) AMCAT (Audit Tool for Mental Capacity Assessments) AQ (Aggression Questionnaire) ASBI (Aggressive Sexual Behavior Inventory)	ASIC (Adolescent Sexual Interest Card Sort) ASPECT (Ackerman-Schoendorf Scales for Parent Evaluation of Custody) ASQ (IPAT Anxiety Scale/IPAT Anxiety Scale Questionnaire) ASRS (Autism Spectrum Rating Scale) ASTM (Amsterdam Short-Term Memory Test) AUDIT (Alcohol Use Disorders Identification Test) b Test B-SAFER (Brief Spousal Assault Form for the Evaluation of Risk) BAI (Beck Anxiety Inventory) BASC (Behavior Assessment System for Children) BDHI (Buss-Durkee Hostility inventory) BDI (Beck Depression Inventory) BHI (Battery for Health Improvement) BHS (Beck Hopelessness Scale) BIDR (Balanced Inventory of Desirable Responding)	BIS (Barratt Impulsiveness Scale) BPAQ (Buss-Perry Aggression Questionnaire) BPI (Basic Personality Inventory) BPRS (Brief Psychiatric Rating Scale) BPS (Bricklin Perceptual Scales) BrownADDScales (Brown Attention-Deficit Disorder Scales) BSI (Brief Symptom Inventory) BSS (Beck Scale for Suicide Ideation) Bumby MOLEST Scale Bumby RAPE Scale BYI (Beck Youth Inventories) CADCOMP (Computer-Assisted Determination of Competency to Proceed) CAI (Career Assessment Inventory) CAI (Child Attachment Interview) CAI (Competency Assessment Instrument/Competency to Stand Trial Assessment Instrument)

CAIR (Clinical Assessment of Interpersonal Relationships)
CAPI (Child Abuse Potential Inventory)
CAPS (Clinician-Administered PTSD Scale)
CAQ (Clinical Analysis Questionnaire)
CARB (Computerized Assessment of Response Bias)
CASS (Coolidge Austistic Symptoms Survey)
CAST*MR (Competence Assessment for Standing Trial for Defendants with Mental Retardation)
CAT (Children's Apperception Test)
CATI (Coolidge Axis II Inventory)
CBCL ([ASEBA] [Achenbach System of Empirically Based Assessment] Child Behavior Checklist)
CCEI (Crown-Crisp Experiential Index)
CCI (Coolidge Correctional Inventory)
CDI (Children's Depression Inventory)
CES (Combat Exposure Scale)
Coercive Sexual Fantasies Questionnaire
Conners/CRS/CPRS (Conners' Rating Scales/Conner's Parent Rating Scales)
COPS (Career Occupational Preference System Interest Inventory)
CPC (Children's Problem Checklist)
CPI (California Psychological Inventory/California Personality Inventory)
CPNI (Coolidge Personality and Neuropsychological Inventory for Children)
CPS (Carlson Psychological Survey)
CPT (Conners Continuous Performance Test)
CQ (Custody Quotient)
CS (Credibility Scale)
CSBI (Child Sexual Behavior Inventory)
CSS (Criminal Sentiments Scale)
CST (Competence/y Screening Test)
CTONI (Comprehensive Test of Nonverbal Intelligence)
CTS (Conflict Tactics Scale)
DAPS (Detailed Assessment of Posttraumatic Stress)
DASS (Depression Anxiety Stress Scales)
DAST (Drug Abuse Screen[ing] Test)
DBQ (Disability Benefits Questionnaire)
DCT (Dot Counting Test)
DHS (Depression, Hopelessness and Suicide Screening Form)
DIVA (Diagnostic Interview for ADHD in adults/Diagnostic Interview for Adult ADHD)
DMT (Digit Memory Test)
DRS (Dementia Rating Scale)
DTS (Davidson PTSD scale)
ECST-R (Evaluation of Competence to Stand Trial–Revised)
EMP (Early Memories Procedure)
ERASOR (Estimate of Risk of Adolescent Sexual Offense Recidivism)
EXIT25 (Executive Interview)
Family Relations Test
FIT (Fitness Interview Test)
FIT (Rey 15-Item Test)
FOTRES (Forensic Operationalized Therapy/Risk Evaluation-System)
FRAT-Y (Firesetting Risk Assessment Tool for Youth)
GCCT (Georgia Court Competency Test)
GCS (Gudjonsson Compliance Scale)
GDS (Geriatric Depression Scale)
GSS (Gudjonsson Suggestibility Scales)
H-T-P (House-Tree-Person)
HAI (Health Anxiety Inventory)
HAM-A (Hamilton Anxiety Rating Scale)
HBI-R (Hilson Background Investigation Inventory-Revised)
HCR-20 (Historical Clinical Risk Management-20)

HCSI (Hilson Career Satisfaction Index)
HCTI (Horney-Coolidge Tridimensional Inventory)
HFD (Draw-A-Person Test/Human Figure Drawing)
HIT (Holtzman Inkblot Technique)
HLAP (Hilson Life Adjustment Profile)
HMI (Hypermasculinity Index)
HRNB-OC (Halstead-Reitan Neuropsychological Test Battery for Older Children)
HRSD/HDRS/HAM-D (Hamilton Rating Scale for Depression/Hamilton Depression Rating Scale)
ICU (Inventory of Callous-Unemotional Traits)
IES (Impact of Event Scale)
ILK (Inventory of Legal Knowledge)
Interdisciplinary Fitness Review Interview (IFI) golding
IORNS (Inventory of Offender Risk, Needs, and Strengths)
IPAT-D (IPAT Depression Scale)
IPI (Inwald Personality Inventory)
IRMAS (Illinois Rape Myth Acceptance Scale)
ISC (Wechsler Intelligence Scale for Children)
J-SOAP (Juvenile Sex Offender Assessment Protocol)
JACI (Juvenile Adjudicative Competence Interview)
JI (Jesness Inventory)
Juvenile Sentence Completion
K-SADS/Kiddie-Sads (Kiddie Schedule for Affective Disorders and Schizophrenia)
KBIT (Kaufman Brief Intelligence Test)
KFD (Kinetic Family Drawing)
KTEA (Kaufman Test of Educational Achievement)
LMT (Letter Memory Test)
LSI/LSI-R/LSI-OR/LS/CMI/LS/RNR (Level of Service Inventory, Level of Service Inventory-Revised, Level of Service Ontario Revision, Level of Services/Case Management Inventory, Level of Service/Risk, Need, Responsivity]
M-FAST (Miller Forensic Assessment of Symptoms Test)
M-PTSD (Mississippi Scale for Combat-Related PTSD)
M-Test
M.I.N.I. (Mini International Neuropsychiatric Interview)
M.I.N.I. Kid (Mini International Neuropsychiatric Interview for Children and Adolescents)
MAB (Multidimensional Aptitude Battery)
MacCAT-CA (MacArthur Competence Assessment Tool–Criminal Adjudication)
MacCAT-T (MacArthur Competence Assessment Tool for Treatment)
MACI (Millon Adolescent Clinical Inventory)
MAST (Michigan Alcoholism Screening Test)
MATS (Mehrabian Achieving Tendency Scale)
MAYSI (Massachusetts Youth Screening Instrument)
MCAA (Measures of Criminal Attitudes and Associates)
MCMI (Millon Clinical Multiaxial Inventory)
MCPS (Missouri Children's Picture Series)
MDI (Major Depression Inventory)
MDI (Multiscale Dissociation Inventory)
MENT (Morel Emotional Numbing Test for PTSD)
Millon Pre-Adolescent Clinical Inventory (M-PACI)
MMPI (Minnesota Multiphasic Personality Inventory)
MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2-Restructured Form)
MMPI-A (Minnesota Multiphasic Personality Inventory-Adolescent)
MMPI-A-RF (Minnesota Multiphasic Personality

Inventory-Adolescent-Restructured Form)
MnSOST (Minnesota Sex Offender Screening Tool)
MPS (Malingering Probability Scale)
MRCI / IAU (Miranda Rights Comprehension Instruments/ Instruments for Assessing Understanding and Appreciation of Miranda Rights)
MSCA (McCarthy Scales of Children's Abilities/McCarthy Scales of Cognitive Abilities)
MSCL (Medical Symptom Checklist)
MSE (Mental State at the Time of the Offense Examination/Mental State at the Time of the Offense Screening Examination)
MSI - Adolescent Male/Female Form (Multiphasic Sex Inventory - Adolescent Male/Female Form)
MSI - Adult Male/Female Form (Multiphasic Sex Inventory - Adult Male/Female Form)
MSI (Multiphasic Sex Inventory)
NAS-PI (Novaco Anger Scale and Provocation Inventory)
NEO-FFI (NEO Five-Factor Inventory)
NEO-PI (NEO Personality Inventory)
NOC (Nims Observation Checklist)
NODS (National Opinion Research Center DSM Screen for Gambling Problems)
ODARA (Ontario Domestic Assault Risk Assessment)
OMNI Personality Inventory/OMNI-IV
Personality Disorder Inventory
ORT (Object Relations Technique)
ORT (Opioid Risk Tool)
P-3 (Pain Patient Profile)
PAI (Personality Assessment Inventory)
PAI-A (Personality Assessment Inventory-Adolescent)
PANSS (Positive and Negative Syndrome Scale)
PASS (Parent Awareness Skills Survey)
PCIT (Parent-Child Interaction Test)
PCL-C/PCL-M (PTSD Checklist, -Civilian, -Military)
PCL-R / PCL-SV (Psychopathy Checklist–Revised / -Screening Version)
PCL-YV (Hare Psychopathy Checklist: Youth Version)
PCL:SV (Hare Psychopathy Checklist: Screening Version)
PCRI (Parent-Child Relationship Inventory)
PCS (Pain Catastrophizing Scale)
PCS (Post-Concussion Scale)
PDI (Pain Disability Index)
PDI-R (Psychiatric Diagnostic Interview-Revised)
PDRT (Portland Digit Recognition Test)
PDS (Paulhus Deception Scales)
PDS (Posttraumatic Diagnostic Scale)
PDS (Posttraumatic Distress Scale)
PENN PTSD scale (Penn Inventory)
PEPQ (PsychEval Personality Questionnaire)
PFS (Rosenzweig Picture-Frustration Study)
PHQ (Patient Health Questionnaire)
PHQ-A (Patient Health Questionnaire for Adolescents)
PHS (Parenting History Survey)
PIAT (Peabody Individual Achievement Test)
PIC (Personality Inventory for Children)
PICTS (Psychological Inventory of Criminal Thinking Styles)
PID / PIDS (Pride in Delinquency Scale)
PORT (Perception-of-Relationships Test)
PPC (Personal Problems Checklist for Adolescents)
PPC (Personal Problems Checklist for Adults)
PPCP (Parent Perception of Child Profile)
PPI (Psychopathic Personality Inventory)

PRF (Personality Research Form)
PRQ (Parenting Relationship Questionnaire)
PSI (Parenting Stress Index)
PSI (Psychological Screening Inventory)
PSQ (Pain and Sleep Questionnaire)
PSQ (Patient Satisfaction Questionnaire)
PSQ (Perceived Stigmatization Questionnaire)
PSQ (Perceived Stress Questionnaire)
PSQ (Personality Structure Questionnaire)
PSS (Parenting Satisfaction Scale)
QT (Ammons Quick Test)
R-CRAS (Rogers Criminal Responsibility Assessment Scales)
R-PAS (Rorschach Performance Assessment System [excluding the Rorschach and -CS])
RADS (Reynolds Adolescent Depression Scale)
RCMAS (Revised Children's Manifest Anxiety Scale)
RIAS/RIST (Reynolds Intellectual Screening Test/Reynolds Intellectual Assessment Scales)
RIASI (Research Institute on Addictions Self-Inventory)
RINTB (Reitan-Indiana Neuropsychological Test Battery)
RISB (Rotter Incomplete Sentences Blank)
RMAS (Burt Rape Myth Acceptance Scale)
RMT (Recognition Memory Test)
RMT (Rey Memory Test)
Roberts (Roberts Apperception Test for Children)
Rorschach (Rorschach Inkblot Test, including the CS [Comprehensive System], excluding the R-PAS [Rorschach Performance Assessment System])
RPM (Raven's Progressive Matrices)
RRASOR (Rapid Risk Assessment for Sex Offense Recidivism)
RSTI (Risk-Sophistication-Treatment Inventory)
RSVP (Risk for Sexual Violence Protocol)
SADS (Schedule for Affective Disorders and Schizophrenia)
SAQ (Self-Appraisal Questionnaire)
SAQ (Substance Abuse Questionnaire)
SARA (Spousal Assault Risk Assessment Guide)
SAS (Zung Self-Rating Anxiety Scale)
SASSI (Substance Abuse Subtle Screening Inventory)
SAST (Sexual Addiction Screening Test)
SAVRY (Structured Assessment of Violence Risk in Youth)
SB (Stanford-Binet Intelligence Scales)
SCID (Structured Clinical Interview for DSM) Axis 1 / Axis 2
SCL-90 (Symptom Checklist-90)
SCT (sentence completions test)
SDS (Zung Self-Rating Depression Scale)
SFQ/WSFQ (Wilson Sex Fantasy Questionnaire)
Shipley / SILS (Shipley Institute of Living Scale)
SHQ (Clarke Sex History Questionnaire for Males[–Revised])
SII (Strong Interest Inventory/Strong-Campbell Interest Inventory)
SIMS (Structured Inventory of Malingered Symptomatology)
SIPA (Stress Index for Parents of Adolescents)
SIRS (Structured Interview of Reported Symptoms)
SIT and S-FRIT (Slosson Intelligence Test and Slosson Full-Range Intelligence Test)
SIVRA-35 (Structured Interview for Violence Risk Assessment)
SORAG (Sex Offender Risk Appraisal Guide)
SOS (Self-Improvement Orientation Scheme)
SPIN (Social Phobia Inventory)

SPS (Suicide Probability Scale)
SRA-FV (Structured Risk Assessment-Forensic Version)
SRP (Hare Self-Report Psychopathy Scale)
SRP (Stalking Risk Profile) Stable 2007
STAI (State-Trait Anxiety Inventory)
START (Short-Term Assessment of Risk and Treatability)
START:AV (Short-Term Assessment of Risk and Treatability: Adolescent Version)
Static-99/R/2002/2002R/2007 (Static-99, 99-R, 2002, 2002R, 2007)
STAXI (State-Trait Anger Expression Inventory)
SVR-20 (Sexual Violence Risk–20)
TAT (Thematic Apperception Test)
TED (Tasks of Emotional Development Test)
TOMM (Test of Memory Malingering)
TONI (Test of Nonverbal Intelligence)
TOPF-UK (Test of Pre-Morbid Functioning - standalone test in the UK)
TRF (Teacher's Report Form)
TSCC and TSCYC (Trauma Symptom Checklist for Young Children and Trauma Symptom Checklist for Children)
TSCS (Tennessee Self-Concept Scale)
TSCYC (Trauma Symptom Checklist for Young Children)
TSI (Trauma Symptom Inventory)
UCCES (Uniform Child Custody Evaluation System)
VAS/VAS-E/VAS-R (Vocabulary Assessment Scales/–Expressive/–Receptive
VASOR (Vermont Assessment of Sex Offender Risk)
Vineland/VABS (Vineland Adaptive Behavior Scales)
VIP (Validity Indicator Profile)
VRAG (Violence Risk Appraisal Guide)
VRS (Violence Risk Scale)
VRS-SO (Violence Risk Scale: Sexual Offender Version)
VRS-YV (Violence Risk Scale: Youth Version)
VSMS (Vineland Social Maturity Scale)
VSVT (Victoria Symptom Validity Test)
WAIS (Wechsler Adult Intelligence Scale)
WASI (Wechsler Abbreviated Scale of Intelligence)
WAVR-21 (Workplace Assessment of Violence Risk)
WCT (Word Choice Test [part of ACS Advanced Clinical Solutions for the WAIS-IV and WMS-IV])
WIAT (Wechsler Individual Achievement Test)
WJ (Woodcock-Johnson Tests of Achievement)
WJ (Woodcock-Johnson Tests of Cognitive Abilities)
WMT (Word Memory Test)
WPPSI (Wechsler Preschool and Primary Scale of Intelligence)
WPSI (Wahler Physical Symptoms Inventory)
WRAT (Wide Range Achievement Test)
WURS (Wender Utah Rating Scale)
WVI (Work Values Inventory)
WVRS (World-View Rating Scale)
WZT/WDCT (Wartegg Zeichen Test/Wartegg Drawing Completion Test)
YOLSI/YLSI/YLS/CMI (Youth Offender Level of Service Inventory, Youth Level of Service Inventory, Youth Level of Service/Case Management Inventory)
YSR/YSQ ([Achenbach System of Empirically Based Assessment] Youth Self-Report)

APPENDIX B

LIST OF NEUROPSYCHOLOGICAL TOOL NAMES AND ACRONYMS

Test/tool name	Acronym(s)
Animal Naming [verbal fluency test/task]	ANIMALS
Austin Maze	
Beery-Buktenica Developmental Test of Visual-Motor Integration	Beery VMI
Bender Visual-Motor Gestalt Test	Bender-Gestalt
Boston Naming Test	BNT
Behavior Rating Inventory of Executive Function	BRIEF
Brief Visuospatial Memory Test	BVMT
Benton Visual Retention Test	BVRT
Category Test/Halstead Category Test/Booklet Category Test	CAT/HCT/BCT
Children's Auditory Verbal Learning Test	CAVLT
Children's Category Test	CCT
Cognistat/Neurobehavioral Cognitive Status Examination	Cognistat/NCSE
CogScreen	
Controlled Oral Word Association Test	COWA/COWAT
California Verbal Learning Test	CVLT
California Verbal Learning Test-Children's Version	CVLT-C
Continuous Visual Memory Test	CVMT
Delis-Kaplan Executive Function System	D-KEFS
FAS verbal fluency test/task	FAS
Finger Tap/ping/Oscillation Test	FTT
General Neuropsychological Deficit Scale	GNDS
Grip Strength (Hand Dynamometer)	
Halstead-Wepman Aphasia Screening Test	
Iowa Gambling Task	IGT
Kaufman Assessment Battery for Children	KABC
Lateral Dominance Examination	
Luria-Nebraska Neuropsychological Battery	LNNB
Luria-Nebraska Neuropsychological Battery for Children	LNNB-C
Memory-for-Designs Test	MFD
Mini-Mental State Examination	MMSE
Montreal Cognitive Examination	MoCA
Neuropsychological Assessment Battery	NAB/NAB-SM/S-NAB
National Adult Reading Test	NART
Paired Associate/s Learning	PAL
Paced Auditory Serial Addition Test	PASAT
Psychological Processing Checklist	PPC
Peabody Picture Vocabulary Test	PPVT
Rey Auditory Verbal Learning Test	RAVLT
Repeatable Battery for the Assessment of Neuropsychological Status	RBANS
Reitan-Indiana Aphasia Screening Test	
Rey-Osterrieth Complex Figure/Rey Complex Figure Test	ROCF/RCFT

Test/tool name	Acronym(s)
Short Category Test	SCT
Symbol Digit Modalities Test	SDMT
Reitan-Klove Sensory Perceptual Examination	SPE
Seashore Rhythm Test	SRT
Speech Sounds Perception Test	SSPT
Stroop [e.g., Stroop Color and Word Test]	
Tactile Form Recognition Test	TFR
Test of Pre-Morbid Functioning	TOPF
Test of Variables of Attention	TOVA
Tactual Performance Test	TPT
Trails A & B/Trail Making Test/Comprehensive TMT	TMT/CTMT
Hooper Visual Organization Test	VOT
Wisconsin Card Sorting Task	WCST/WCST-64
Wechsler Memory Scale	WMS

APPENDIX C

CODING GUIDE FOR LEGAL CASE LAW ANALYSIS

Coding Guide

In 2019, Tess Neal and colleagues published a systematic review of the quality and admissibility of psychological assessment tools that psychologists include in their evaluations in legal settings. Out of the 364 assessment tools in the sample, only 67% were generally accepted by psychologists, and only 40% had favorable reviews of their psychometric properties (i.e., reliability and validity). Legal challenges were rarely raised to this evidence, occurring in only about 5% of cases, and challenges that were raised were successful in excluding the evidence only about a third of the time. We are now interested in the admissibility of neuropsychological tools in court and might expect to see some differences in the rate of challenges raised due to potential variation in psychometric quality and its general reputation.

The original paper is accessible here. Part II of the paper (starting on page 15) details this systematic case law analysis looking at admissibility challenges. Reading this section and viewing the tables and coding information will be helpful moving forward with this current coding project as we are aiming to replicate this process.

A few other clarifying points: Throughout this guide, in general, and likely in the cases you will read, several phrases may be used to refer to the same concept. In this project, I am focusing on the admissibility of neuropsychological assessment evidence. Specifically, I am referring to the assessment tools that psychological experts include in the testimony they give in court. You might see discussion about "psychological testimony," "expert testimony," "expert evidence," the neuropsychological tools in particular (e.g., the ANIMALS test, etc.), or other related terms. These all roughly refer to the same concept of interest.

Coding assignments can be viewed on the first tab in the Excel document. The full names of each tool and their acronyms can be viewed on the second tab. Each tool has its own tab in the document with cases which included the tool.

> *Note:* I have preliminarily screened the cases to ensure that they do include the neuropsychological tool (i.e., for cases that use the Trail Making tests, excluding those that discuss hiking trails), but please let me know if you come across any cases which do not mention the tools.

PDFs of all of the cases can be accessed in Dropbox using the following link. They are organized in folders under each of the neuropsychological tools.
https://www.dropbox.com/sh/7sp5wh1px3hkvrm/AAB7pj9LwYxYSXt8j_x5Pj6Va?dl=0

Summary of the flow of coding:

> If the tool was not challenged: only answer "Tool challenged?"
> If the tool was challenged: "Tool challenged?"
> "What was the basis of the challenge?"
> "Summary of challenge"
> "Where did you find info about the challenge?"

"Was the evidence ruled inadmissible?"
"FS"
"FM"
"DM"
"DI"

Tool challenged?
0 = No, 1 = Yes
Was a legal admissibility challenged raised in regard to the neuropsychological tool? If a challenge was raised, please also identify the basis of the challenge and whether the evidence was ruled inadmissible.

If Yes, what was the basis of the challenge?
If the tool was challenged, identify the basis of the admissibility challenge. You may select multiple options if more than one admissibility issue was considered.
1 = "Fit" or Relevance to Legal Issue. This refers to the notion that the evidence must answer the legal question presented.

Example: For *United States v. Jones* (2018), the court rejected the government's argument that the **expert's testimony** (i.e., that the defendant had significant deficits in cognition, according to Rorschach results) **was irrelevant** to whether the defendant had specific intent to commit fraud).

2 = Reliability / Validity. This broadly refers to the tool's psychometric quality, such as its reliability, validity, testability, error, etc. Does the tool measure what it claims to? Is it accurate? Has it been tested? Does it have a high error rate?

Example: For *Reaes v. City of Bridgeport* (2017), the court rejected plaintiff's claim that the PAI was "**culturally biased**," thus resulting in disparate impact in an employment decision

3 = General / Peer Acceptance. This refers to whether the tool is generally accepted in its relevant field (i.e., by psychologists). Is the tool used or endorsed by other experts? Is the tool commonly used in similar matters?

Example: For *In the Matter of Garcia,* (2018) the court concluded that, although the defense expert "did not provide any **evidence that [the Rorschach is] routinely used for sexually violent predator determinations**" (p. 1), and although the state had a plausible argument that the evidence should be inadmissible under Daubert or Frye, the admission of the testimony was harmless, because the expert described it and other evidence as merely "pieces of the puzzle" (p.9), and there was no indication that the trial court considered the test in making its decision.

4 = Qualifications of Expert. This refers to whether the expert has the specialized knowledge or skills necessary to offer the opinion proffered.

Example: The court in *J.K.J. v. Polk County* (2017) considered at length the plaintiff's claims that the **experts were not qualified** to administer the MMPI.

5 = Helpfulness. This refers to whether the evidence adds to what a jury could readily discern for itself.

6 = Prejudicial Impact. This refers to the potential to induce bias, rouse emotions, or otherwise interfere with the ability to reach an impartial decision.
7 = *Other (please detail)*. If the admissibility challenge does not fall into one of the above categories, please explain.

Summary of challenge

Please provide a brief summary of the reasoning underlying the admissibility challenge (maybe a sentence or two).

Where did you find info about the challenge?

Indicate where you located this information about the admissibility challenge. This can be the page number, or a quote from the case containing the relevant information.

Was the evidence ruled inadmissible?

Indicate whether the evidence related to the assessment tool was ultimately excluded from the case.

The following categories rely on the distinction between framework evidence and diagnostic evidence. *Framework testimony* is testimony about general scientific concepts that can be applicable to more than one case. *Diagnostic testimony* seeks to draw conclusions about a particular case by applying the general scientific/framework knowledge to the case at hand.

If the tool was challenged, please describe the focus of the admissibility challenge:

FS

The focus of the admissibility analysis was on the substance of framework evidence.

FM

The focus of the admissibility analysis was on the methodology of the framework evidence.

DM

The focus of the admissibility analysis was on the methodology used to obtain diagnostic evidence.

DI

The focus of the admissibility analysis was on the credibility/rationale of testimony about how the diagnostic methodology in the case at hand was used and the results that the expert reached.

APPENDIX D

FORENSIC MENTAL HEALTH EVALUATOR STUDY 3 IRB APPROVAL AND APPROVED DOCUMENTS

ASU Knowledge Enterprise Development

EXEMPTION GRANTED

Tess Neal
NCIAS: Social and Behavioral Sciences, School of (SSBS)
602/543-5680
Tess.Neal@asu.edu

Dear Tess Neal:

On 9/6/2022 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Neuropsychological Assessment Evidence in Court: Quality and Challenges to Admissibility
Investigator:	Tess Neal
IRB ID:	STUDY00016438
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• Consent Form_Elizabeth Mathers_IRB_2022.8.24, Category: Consent Form; • Debrief_ElizabethMathers_IRB_20220901.pdf, Category: Other; • IRB Protocol_Elizabeth Mathers Neuropsych Thesis 2022.9.3.docx, Category: IRB Protocol; • Participant recruitment letter for mailing.pdf, Category: Recruitment Materials; • Survey Questionnaire_Elizabeth Mathers_IRB.2022.8.24.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions);

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (2)(i) Tests, surveys, interviews, or observation (non-identifiable), (2)(ii) Tests, surveys, interviews, or observation (low risk) on 9/6/2022.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

If any changes are made to the study, the IRB must be notified at research.integrity@asu.edu to determine if additional reviews/approvals are required. Changes may include but not limited to revisions to data collection, survey and/or interview questions, and vulnerable populations, etc.

REMINDER - - Effective January 12, 2022, in-person interactions with human subjects require adherence to all current policies for ASU faculty, staff, students and visitors. Up-to-date information regarding ASU's COVID-19 Management Strategy can be found here. IRB approval is related to the research activity involving human subjects, all other protocols related to COVID-19 management including face coverings, health checks, facility access, etc. are governed by current ASU policy.

Sincerely,

IRB Administrator

cc: Elizabeth Mathers

ASU School of Social and Behavioral Sciences

Appendix A.

Informed Consent Form

Title of research study: Neuropsychological Assessment Evidence in Court: Quality and Challenges to Admissibility

Investigators:

Elizabeth Mathers, BS, and Tess M.S. Neal, PhD

Why am I being invited to take part in a research study?

We invite you to take part in this research study because we are interested in how psychological assessment evidence is received in court.

You are qualified to participate in this research because you are a mental health expert with interest or experience conducting forensic mental health evaluations in the United States.

Why is this research being done?

A limited body of research exists exploring how psychological assessment evidence is received, screened, excluded, and relied upon by the courts. We are interested in adding to that body of research.

How long will the research last?

We expect that individuals will spend approximately 10-15 minutes participating in the proposed activities.

How many people will be studied?

We expect about 120 people will participate in this research study.

What happens if I say yes, I want to be in this research?

You will be asked to fill out a survey with questions about your experiences in regard to how psychological assessment evidence is received in court, your career demographics, and basic demographics.

What happens if I say yes, but I change my mind later?

You can leave the research at any time with no penalty.

Is there any way being in this study could be bad for me?

The risks of this research are expected to be minimal. The survey questions are similar to the types of questions you may encounter on a daily basis regarding your work.

Will being in this study help me in any way?

There are no direct personal benefits to you for participating; however, your participation may contribute valuable information to increase knowledge in the field. In regard to the importance of the research to the legal and psychological professions, benefits may include contributing knowledge about how psychological assessment evidence is received, screened, excluded, and relied upon by the court.

What happens to the information collected for the research?

Organizations that may inspect and copy your information include the University board that reviews research who want to make sure the researchers are doing their jobs correctly and protecting your information and rights.

We will not collect any identifying information in the survey. Basic demographic information will be collected and aggregated with other participant data, and then may be published in scientific journals, presented at scientific meetings, and included in training workshops. All anonymous data will be archived, encrypted, and stored indefinitely in secure drives monitored by the ASU technology services. After publication of the data, we will also make the anonymous database available through the Open Science Framework.

Who can I talk to?

If you have questions, concerns, or complaints, talk to the research team at tneal6@asu.edu or 602-543-5680.

This research has been reviewed and approved by the Social Behavioral IRB. You may talk to them at (480) 965-6788 or by email at research.integrity@asu.edu if:

- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You have questions about your rights as a research participant.
- You want to get information or provide input about this research.

ASU School of Social and Behavioral Sciences

September, 2022

Dear Colleague:

Greetings from the Law and Behavioral Science group at Arizona State University (lawpsych.asu.edu)! We are writing in the hopes that you will join us in gathering information about our profession. We would appreciate it if you would give 10 to 15 minutes of your time to respond to a brief survey, ***"Neuropsychological Assessment Evidence in Court: Quality and Challenges to Admissibility"*** about legal admissibility challeges to psychological assessment evidence.

Limited information is available about how psychological assessment evidence is received, screened, excluded, and relied upon by the courts. We are interested in adding to the body of data by gathering information on the frequency and circimstances of challenges to psychological assessment evidence in court. The 1993 *Daubert v. Merrell Dow Pharmaceuticals* case established criteria for admitting scientific evidence in court. It holds that scientific evidence must be valid, reliable, and relevant, and judges are required to be "gatekeepers" of evidence by screening out unreliable science. Jurisdictions throughout the U.S. have adopted *Daubert*-related standards for the admissibility of scientific evidence into legal proceedings. Yet, little is known about whether psychological assessment tools are subjected to legal scrutiny.

This project is one of the first studies to examine these questions. You are invited to participate because you are a licensed psychologist in the 50 United States with forensically-related interests. To participate, please go to the following web address. We opted to use an electronic survey platform rather than paper-and-pencil to reduce researcher error. The survey is available at this address. If you have any trouble, I can email you a direct link that will work.

(link)

We are happy to share the results. If you need clarification, or have questions, or would like a summary of the results, please contact either of us at the email address listed below. Thank you for your time and your attention to this request. We hope you join us in this project.

Warm regards,

Elizabeth Mathers, B.S.
Graduate Student in Psychology
School of Social and Behavioral Sciences
Arizona State University
emather2@asu.edu
psych-law.lab.asu.edu

Tess M.S. Neal, Ph.D.
Assistant Professor of Psychology
School of Social and Behavioral Sciences
Arizona State University
tess.neal@asu.edu
psych-law.lab.asu.edu
602.543.5680

ASU School of Social and Behavioral Sciences

Appendix C.

Debriefing Information

Thank you for participating in this study! The purpose of this study was to gather information on *Daubert*-related challenges to psychological assessment evidence in court. Currently, there is no known research on the frequency and circumstances of challenges to psychological assessment evidence. The information you have provided will be helpful in obtaining initial data on this topic. This data will reveal if challenges are happening, when, and if they are occurring at appropriate times.

The 1993 *Daubert v. Merrell Dow Pharmaceuticals* case established criteria for admitting scientific evidence in court (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*). It holds that scientific evidence must be valid, reliable, and relevant, and judges are required to be "gatekeepers" of evidence by screening out junk science. Yet, little is known about whether psychological assessment tools are subjected to scrutiny through the standards courts are supposed to apply. Not only are we unaware of the frequencies of challenges in court, we are just now realizing the sheer number and variety of psychological assessment tools used by forensic mental health evaluators (Neal & Grisso, 2014). With so many different assessment tools in use, this data will provide some insight into how courts are receiving the large variety of psychological assessment evidence.

If you have any questions or are interested in knowing more about this study, we'd be happy to send you the results. Please contact us via email at emather2@asu.edu if you'd like to be added to the list of people to whom we will send the results when data collection is complete. Thank you once again for your participation.

Sincerely,

Elizabeth Mathers & Tess Neal
Arizona State University
emather2@asu.edu

APPENDIX E

FORENSIC MENTAL HEALTH EVALUATOR SURVEY QUESTIONS

Forensic Mental Health Evaluator Questionnaire

The 1993 *Daubert v. Merrell Dow Pharmaceuticals* case established criteria for admitting scientific evidence in court.1 It holds that scientific evidence must be valid, reliable, and relevant, and judges are required to be "gatekeepers" of evidence by screening out unreliable science. Jurisdictions throughout the U.S. have adopted *Daubert*-related standards for the admissibility of scientific evidence into legal proceedings. Yet, little is known about whether psychological assessment tools are subjected to legal scrutiny. Not only are we unaware of the frequencies of challenges in court, we are just now realizing the sheer number and variety of psychological assessment tools used by forensic mental health evaluators.2-3 How often are psychological assessment tools challenged by attorneys? How often do judges exclude psychological assessments from being admitted as evidence? Are mental health experts experiencing *Daubert*-related challenges to their testimony? This research project provide some answers to these questions from the perspective of mental health experts.

(1) *Daubert v. Merrell Dow Pharmaceuticals, Inc*., 509 U.S. 579, 113 S. Ct. 2786 (1993). (2) Neal, T.M.S., & Grisso, T. (2014). Assessment practices & expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice & Behavior*, 41, 1406–21. (3) Neal, T., Philipp, C, & Goddard, H. (2018, March). Daubert & psychological tests in forensic evaluations. Paper presented at American Psychology-Law Society, Memphis, TN.

Have you ever been involved in a case in which a lawyer raised a legal admissibility challenge to psychological assessment evidence you provided? By psychological assessment evidence, we mean any case in which you were hired to conduct an assessment of someone's mental health or mental status related to a legally-related question.

- ☐ Yes
- ☐ No

Please tell us the details of what happened in the most recent case in which you experienced a legal admissibility related challenge to your psychological assessment evidence. We are most interested in what type of issue you were hired to assess, what psychological assessment tool(s) you relied upon in your evaluation, the grounds on which the evidence was challenged, and the outcome of the challenge. Note that we are going to ask some additional close-ended questions below about the case, but we'd like you to start with an open-ended description of your experience.

How justified do you think the attorney was in raising the challenge(s) against your psychological assessment evidence?

- ☐ Very Unjustified
- ☐ Somewhat Unjustified
- ☐ Unsure
- ☐ Somewhat Justified
- ☐ Very Justified

What was the referral question for the most recent case in which you experienced a legal admissibility challenge to your psychological assessment evidence?

____ Competence (fitness) to Stand Trial
____ Criminal Responsibility/NGRI/MSO
____ Violence/Recidivism Risk
____ Sex Offender Risk
____ Capacity to Waive Miranda Rights
____ Transfer (waiver) of a Juvenile to/from Adult Court
____ Sentencing/Disposition
____ Child Custody
____ Child Protection (e.g., child abuse, termination of parental rights)
____ Competence to Consent to Treatment
____ Guardianship/Conservatorship
____ Civil Commitment
____ Workplace Disability
____ Psychiatric or Psychological Disability in Civil Suits
____ Other (Please Specify_____)

If you answered 'other,' please specify the referral question.

What psychological assessment tool(s) did you use in the most recent case in which you experienced a legal admissibility challenge (e.g., tests, instruments, checklists, rating systems)? Please list each tool that was challenged:

Please give us your best estimate as to how many times have you experienced a legal admissibility challenge to psychological assessment evidence you intended to offer in a legal case.

_____ times

Do you know of any colleagues who have experienced a legal admissibility challenge to psychological assessment evidence?

- ☐ Yes
- ☐ No

What were the conditions of their legal admissibility challenge (if you are aware of them)? Please tell us the details of the most recent challenge you know of that one of your colleagues experienced. Again, we are most interested in what type of issue your colleague was hired to assess, what psychological assessment tool(s) s/he relied upon in the evaluation, the grounds on which the evidence was challenged, and the outcome of the challenge. Even if you don't know all the details, please tell us as much as you know.

How familiar are you with the legal admissibility criteria relevant to psychological assessments in the jurisdictions you provide evidence?

- ☐ Not at all familiar
- ☐ Small degree
- ☐ Moderate degree
- ☐ High degree
- ☐ Very high degree of familiarity

How important are your jurisdiction's legal admissibility criteria to you when you are performing a forensic evaluation?

- ☐ Not at all important
- ☐ Somewhat important
- ☐ Moderately important
- ☐ Highly important
- ☐ Extremely important

Do you regularly use any specific resources to investigate the scientific integrity of psychological assessment tools?

- ☐ Yes
- ☐ No

Please name and/or describe the resource(s) you use to investigate the scientific integrity of psychological assessment tools here.

How able are **<u>you</u>** to identify psychological assessment tools that meet legal admissibility criteria?

1 2 3 4 5 6 7
Not at all able Very able

How able do you think **<u>other evaluators</u>** are to identify psychological assessment tools that meet legal admissibility criteria?

1 2 3 4 5 6 7
Not at all able Very able

How able do you think **lawyers** are to identify psychological assessment tools that meet legal admissibility criteria?

1 2 3 4 5 6 7

Not at all able Very able

How able do you think **judges** are to identify psychological assessment tools that meet legal admissibility criteria?

1 2 3 4 5 6 7

Not at all able Very able

Please choose the statement that best applies to you. For the purpose of this question, a tool is defined as any standardized test, instrument, checklist, or rating system intended to assist your evaluation.

- ☐ I regularly use psychological assessment tools in forensic evaluations I conduct.
- ☐ I occasionally use psychological assessment tools in forensic evaluations I conduct.
- ☐ I never use psychological assessment tools in forensic evaluations I conduct.
- ☐ I have never conducted a forensic evaluation.

How much effort do you put into learning about a psychological assessment tool before using it in a forensic evaluation?

1 2 3 4 5 6 7

No effort at all An extreme amount

Think back to the last tool you used for a forensic evaluation. What was it? If the evaluation included multiple tools, please enter only the last tool administered for that evaluation.

Now, instead of thinking about the most recent time you used that tool, think back to the very first time you ever used that tool. What did you do to prepare before you administered the tool for the first time? Check all that apply.

- ☐ Read the scoring manual
- ☐ Read the administration manual
- ☐ Read psychometric information about the development, reliability, and validity of the tool (either in manual or primary peer-reviewed literature)
- ☐ Read the primary peer-reviewed literature directly relevant to the tool
- ☐ Looked for evidence of an error rate specific to the case demands (e.g. race/ethnicity of evaluee)
- ☐ Compared it to other available tools
- ☐ Other

What else did you do to prepare before you administered the tool for the first time?

Highest degree you have obtained

- ☐ Ph.D.
- ☐ Psy.D.
- ☐ Joint J.D. / Ph.D.
- ☐ Joint J.D. / Psy.D.
- ☐ Ed.D.
- ☐ Other (Please specify)

In what year did you obtain your highest degree?

- ☐ Dropdown box of years 1940-2022

What is your primary place of employment?

- ☐ Institution or agency (e.g., hospital, prison, court clinic, etc.) - Specify type of institution
- ☐ Private - Specify type of practice
- ☐ University
- ☐ Other or more than one - Specify

In a typical month, what percentage of your work (for a total of 100%) do you spend in each of the following?

- ☐ Clinical practice (non-forensic)
- ☐ Forensic practice
- ☐ Administrative duties
- ☐ Consultation
- ☐ Research
- ☐ Teaching
- ☐ Total percentage (will add up as participants fill in above boxes)

Years of experience conducting forensic mental health evaluations

_____ Years

Did you complete a formal postdoctoral fellowship in forensic psychology?

- ☐ No
- ☐ Yes (Please specify where) _____

Are you board certified (ABPP) in forensic psychology?

- ☐ Yes
- ☐ No

Are you board certified (ABCN) in neuropsychology?

- ☐ Yes
- ☐ No

Are you licensed to practice psychology?

- ☐ Yes
- ☐ No

What is the primary country in which you practice?

- ☐ United States
- ☐ Other, please specify: _____

Which is the primary state in which you practice?

- ☐ Dropdown menu of U.S. states and territories

What is your gender?

- ☐ Male
- ☐ Female
- ☐ Other

What is your age?

Racial/Ethnic Identity (check all that apply)

- ☐ African-American/Black
- ☐ American Indian/Native American
- ☐ Asian/Asian American
- ☐ White
- ☐ Hispanic/Latino(a)
- ☐ Other (please specify) _____

Finally, please be honest when answering the following question. The study you have just participated in is a psychological study aimed at examining mental health experts' experiences with legal admissibility challenges to psychological assessment evidence. Psychological research depends on people willing to participate in research. Your responses to surveys like this one are an incredibly valuable source of data for researchers. It is therefore crucial for research that participants pay attention, avoid distractions, and take all study tasks seriously (even when they might seem silly). **Do you feel that you paid attention, avoided distractions, and took this survey seriously?**

- ☐ No, I was distracted.
- ☐ No, I had trouble paying attention.
- ☐ No, I did not take this survey seriously.
- ☐ No, something else affected my participation negatively.
- ☐ Yes.

Do you have any feedback or comments related to this survey?