



Critical review of the use of the Rorschach in European courts

Igor Areh^a , Fanny Verkampt^b  and Alfred Allan^c 

^aFaculty of Criminal Justice and Security, University of Maribor, Ljubljana, Slovenia; ^bCLLE, Université de Toulouse, CNRS, Toulouse, France; ^cSchool of Arts and Humanities, Edith Cowan University, Perth, Australia

In relation to the admissibility of evidence obtained using projective personality tests arose in *F v. Bevándorlási és Állampolgársági Hivatam* (2018). The Court of Justice of the European Union has held that an expert's report can only be accepted if it is based on the international scientific community's standards, but has refrained from stipulating what these standards are. It appears timely for European psychologists to decide what standards should be applied to determine whether or not a test is appropriate for psycholegal use. We propose standards and then apply them to the Rorschach because it was used in this case and is an exemplar of projective tests. We conclude that the Rorschach does not meet the proposed standards and that psychologists should abstain from using it in legal proceedings even in the absence of a clear judicial prohibition.

Keywords: psychology; law; forensic assessment; courts; professional standards; ethics; evidential value; acceptability; projective test; Rorschach.

The applicant in *F v. Bevándorlási és Állampolgársági Hivatal* ("*F v. Hungary*" 2018) was a Nigerian male who applied for asylum in Hungary claiming that he feared persecution in Nigeria because of his homosexuality. The Hungarian government (Government) asked a panel of psychologists to assess the veracity of the applicant's claim that he was homosexual before granting him asylum. The panel could neither confirm nor deny the applicant's sexual orientation and therefore the Government denied his application for asylum on the basis that it could not establish his general credibility. The applicant took the matter to the Hungarian Court of Appeal, which asked the Court of Justice of

the European Union (Court) to make a preliminary finding regarding the use of an expert report by psychologists who had used three projective tests (Draw-A-Person-In-The-Rain, Rorschach and Szondi). The Court in its decision made two comments and one ruling that are of importance in this article. The first comment is that in circumstances such as in this case 'it must be considered that consent of the person concerned [...] is not freely given' (*F v. Hungary*, 2018, para. 53) and that, under these circumstances, tests might be used 'only if they are necessary and genuinely meet the objectives of [...] the European Union or the need to protect the rights and freedoms of others' (*F v. Hungary*, 2018, para. 55). It

Correspondence: Igor Areh, Faculty of Criminal Justice and Security, University of Maribor, Kotnikova 8, 1000 Ljubljana, Slovenia, EU. Email: igor.arih@fvv.uni-mb.si

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

further commented that the use of projective tests (i.e. tests that use non-structured, unclear stimuli such as ink blots to induce responses; see Erickson et al., 2007) had been vigorously contested during the case. The Court made no decision regarding these points, saying that these matters would have to be decided by the relevant government and national court in each case. The Court continued to assert that an expert's report is only acceptable 'if it is based on sufficiently reliable methods and principles in the light of the standards recognised by the international scientific community' (*F v. Hungary*, 2018, para. 58). However, it did not state what these standards are or decide whether or not it was acceptable to use the three projective tests, saying that this was a decision that the national court had to consider on the facts.

Lawyers criticised the judges in *F v. Hungary* (2018) for failing to provide guidance as to how authorities can decide whether or not a psychological test is a method that meets international standards, and they made negative comments about psychological tests in general (e.g. Ferreira & Venturi, 2018; A. Gould, 2019). These negative comments were mostly aimed at projective tests, but they reflect on all the tests that psychologists use – and indeed on the profession itself. We therefore believe that the decision of the court and the lawyers' comments require European psychologists to publicly declare what standards they and lawyers should use to determine whether or not tests are 'sufficiently reliable methods [...] in the light of the standards recognised by the international scientific community' (*F v. Hungary*, 2018, para. 58).

We will first discuss the relevant international standards and propose a set of contemporary standards that psychologists and lawyers can use to determine whether or not a test is sufficiently reliable for use in psycholegal work. We will then use these standards to critically consider the use of the Rorschach in court because (a) it is frequently used in some European countries (see Areh, 2020) and (b) it

is an exemplar of projective tests, as well as being the best known, most thoroughly researched and arguably most frequently used of the three projective tests that were used in *F v. Hungary* (2018).

International standards

The legal question that the Court faced in *F v. Hungary* (2018) is whether or not the expert witnesses based their testimony on theories and techniques that fall into an *area of expertise*. Courts must generally answer this question when they decide whether or not to admit testimony based on emerging theories and techniques, and the different approaches used to do this are beyond the ambit of this article (for a discussion, see Freckelton, 2019). It is enough to say here that courts, in many countries, share the Court's reluctance to set out the specific standards that should be used to determine whether or not a specific theory or technique can be accepted as an area of expertise (Freckelton, 2019). The United States (US) Supreme Court is, however, an exception to this general rule. In *Daubert v. Merrell Dow Pharmaceuticals* (1993), it indeed set out four requirements (the *Daubert*-standards) that are very influential – even in jurisdictions where courts have not explicitly adopted them (Freckelton, 2019).

Daubert-case

The plaintiffs in this case alleged that a drug called Bendectin had caused the birth defects of their children. The defendant called an expert who, after reviewing the extensive published scientific literature, concluded 'that the maternal use of Bendectin has not been shown to be a risk factor for human birth defects' (*Daubert v. Merrell Dow*, 1993, p. 582). The plaintiff in response called eight expert witnesses who based on novel animal studies and chemical analyses of the drug concluded that Bendectin can cause birth defects. The Federal Court could not consider the opinion of these eight witnesses because it was bound by the

rule laid down in *Frye v. United States* (1923) ‘that expert opinion based on a scientific technique is inadmissible unless the technique is “generally accepted” as reliable in the relevant scientific community’ (*Daubert v. Merrell Dow*, 1993, p. 584). The Federal Court therefore ‘declared that expert opinion based on a methodology that diverges “significantly from the procedures accepted by recognized authorities in the field [. . .] cannot be shown to be” generally accepted as a reliable’ (*Daubert v. Merrell Dow*, 1993, p. 584). The Supreme Court, however, held that the rule in the *Frye*-case did not apply anymore and that judges must in each case assess the validity of scientific testimony (see Faigman, 1995). The Supreme Court, with reference to the work of Popper (1989), asserted:

Many considerations will bear on the inquiry, including whether the theory

or technique in question can be (and has been) tested, whether it

has been subjected to peer review and publication, its known or potential

error rate and the existence and maintenance of standards controlling

its operation, and whether it has attracted widespread acceptance

within a relevant scientific community. The inquiry is a flexible one,

and its focus must be solely on principles and methodology, not on the

conclusions that they generate. (*Daubert v. Merrell Dow*, 1993, p. 580)

In *Kumho Tire Company v. Carmichael* (1999), the Supreme Court clarified that this so-called Daubert test applies to all expert testimony (see Beecher-Monas, 1998), and the standards of this test have been applied to psychology in the US (see *United States v. Hall*, 1996). Both lawyers (e.g. Beecher-Monas, 1998) and psychologists

(e.g. Melton et al., 2018) have accepted that a psychological test – defined as ‘a set of items that has accepted levels of reliability and validity and allows measurement of some attribute of an individual’ (Australian Psychological Society, 2014, para. 7.1) – is a technique as envisaged in the *Daubert* case. The Daubert test is, however, very broad, and each discipline and profession should therefore consider how the test applies to it.

Testability

Psychologists generally interpret testability to mean that the theory upon which the psychological test is based must have been empirically examined and challenged (J. W. Gould et al., 2013). Lawyers such as Beecher-Monas (1998) have observed that this standard might create a dilemma for those parts of psychology that rely mainly on retrospective observational studies and not controlled experimentation. However, as with all the other Daubert standards, testability is a necessity and of relevance to psychological testimony (see Bow et al., 2006).

Peer review and publication

Psychologists have pointed out that the second standard, namely whether or not the test ‘has been subjected to peer review and publication’ (*Daubert v. Merrell Dow*, 1993, p. 593), should be interpreted with caution given publication biases (see Rothstein et al., 2005). Scholars have pointed out that authors prefer to submit – and editors prefer to publish – positive results, thereby favouring the alternative versus null hypothesis and causing a positive-results bias (Nosek et al., 2012). Consequently, the results of meta-analyses – which psychologists consider the most reliable method of assessing acceptance (Marlowe, 1995) – can be distorted, contributing to the survival of poor-quality theories (Ferguson & Heene, 2012).

Error rate

Lawyers (e.g. Beecher-Monas, 1998) and psychologists (e.g. J. W. Gould et al., 2013)

interpret the ‘known or potential rate of error’ (*Daubert v. Merrell Dow*, 1993, p. 594) as a reference to psychometric properties of tests like their validity and reliability. Validity can be in the form of construct (i.e. does it measure the construct of interest, such as psychopathy?) or predictive or discriminative (i.e. how accurate is it in the prediction of the behaviour of interest, such as that an offender will reoffend?) validity. Reliability refers to measurement consistency, precision and repeatability, and the relevant coefficients indicate the degree to which test scores are free from error (Groth-Marnat & Wright, 2016).

General acceptance

The general acceptance of a psychological test can be determined by examining either the frequency of its use or its general acceptance as a credible test by the scientific community, as recorded in notable scientific publications (e.g. McCann & Evans, 2008). The test must therefore meet the profession’s requirements for valid and reliable psychological tests, which we discuss next.

Professional requirements

Professional bodies such as the American Educational Research Association (2014), the International Test Commission (2013) and the United States National Council on Measurement in Education (2014) have developed standards for psychological tests in general. The uniqueness of assessments for courts (Melton et al., 2018), however, has prompted professional bodies representing psychologists who provide psycholegal services – like the American Psychological Association (APA, 2013) and the Australian Psychological Society (2013) – to stipulate the standards that their members should adhere to when conducting assessments for the courts. The APA stresses that tests should have an ‘adequate scientific foundation’ (APA, 2013, Guideline 2.05) and measure ‘response style, voluntariness, and situational stress associated with the

involvement in forensic or legal matters’ (APA, 2013, Guideline 10.02). Authors also contribute to the literature with papers exploring the importance of issues such as the standardisation of test administration (Lee et al., 2003).

Even before the *Daubert case*, psycholegal scholars considered what standards psychologists could use to justify their choice of tests for forensic assessment (e.g. Heilbrun, 1992). Many of their standards are similar to what the profession generally requires of psychological tests. Heilbrun (1992) for instance required that tests should be commercially available and have manuals that provide standardised administration procedures, scoring instructions, population norms to assist with interpretation and finally information which demonstrates that they have acceptable levels of validity and reliability. Psychologists like Melton et al. (2018) have, since the *Daubert case*, specifically considered its impact on psychologists’ choices of tests in psycholegal contexts.

Ethical requirements

Psychologists who do psycholegal work should further adhere to the ethical principles and standards of their profession (see Allan, 2013, 2018). The *Meta-Code of Ethics* of the European Federation of Psychologists’ Associations (EFPA, 2005) provides a useful guide in this regard. Paragraph 3.1.2 requires psychologists to collect only the information that they need for the professional tasks they are undertaking. Paragraph 3.2.3 reminds psychologists to consider the limits of the procedures that they use and the limits of the conclusions that can be drawn from data collected using such procedures, further requiring them ‘to practise within, and to be aware of the psychological community’s critical development of theories and methods’. Paragraph 3.3.1 reflects on psychologists’ obligation under the Responsibility principle not to bring the profession into disrepute (EFPA, 2005), which requires them to avoid behaviour that

might weaken courts', lawyers' and the public's perception of the profession as trustworthy (see Allan, 2013, 2018). Psychologists should also promote and maintain high standards of scientific and professional activity (EFPA, 2005, para. 3.3.2) while minimising reasonably foreseeable and preventable harm (EFPA, 2005, para. 3.3.3).

Possible standards

The four *Daubert* standards are so well established and recognised that it makes sense to use them as the basis of any proposed European standards for the use of psychological tests in psycholegal assessments. They do, however, need further explanation to make them useful to lawyers and psychologists – and they do not cover all the basic requirements found in the ethical codes and guidelines for psychologists. We believe that the standards identified can be consolidated under five headings as follows.

Theoretical basis with peer-reviewed support

Beecher-Monas (1998) stated that the most important *Daubert* standard is that the test must have a sound and tested theoretical basis, as well as supporting data accessible in peer-reviewed journals. The requirement that any supporting research should have been published in reputable peer-reviewed journals serves to verify that the research has been scrutinised by objective and competent peers and that the data has been made available to those who need it, such as forensic practitioners and lawyers (Allan, 2020). Psychologists should control for publication bias when they evaluate the level of support in the literature for the theoretical basis of an instrument.

Validity

Psychologists use tests in the psycholegal context to measure constructs that are legally

relevant in order to obtain information that can be used when preparing a psycholegal report and testimony (Grisso, 1986). Tests must therefore be valid in that they measure the construct that the psychologist, and ultimately the court, wants to know about. Psychologists should when possible use tests that are standardised for the population that the examinee is part of (Allan, 2013, 2018, 2020; Allan et al., 2019). The exact form of validity that is most important will depend on the legal issue in question (Heilbrun, 1992).

Construct validity, which subsumes all other types of validity (see Messick, 1995), is relevant if psychologists want to measure the presence of a specific construct, such as a mental disorder. Criterion (predictive) validity coefficients of psychological tests typically vary from .30 to .40, with values above .60 rare (Kaplan & Saccuzzo, 2017). This is poor when compared with medicine, where convergent validity values below .70 are generally considered as indications of validity problems (see Post, 2016). Given the impact that legal decisions can have on people's rights and interests (see Allan, 2018), psychologists working in the psycholegal context should therefore ideally use tests that have validity coefficients well above .60.

Discriminant and/or predictive validity is important when psychologists want to identify people who are likely to reoffend. Predictive validity is typically determined by calculating the area under the curve (AUC), which is a statistical index that represents the average difference in true positive and false positive rates across all possible cut-offs (Allan et al., 2006). The AUC can range from 0 to 1 (suggesting perfect performance) with an AUC of .50 suggesting a no better than random prediction. Sjöstedt and Grann (2002) considered AUC values of $\geq .60$ but $< .70$ as marginal, $\geq .70$ but $< .80$ as modest, $\geq .80$ but $< .90$ as moderate and $\geq .90$ as high. Forensic practitioners do not stipulate what they consider as an acceptable AUC but in the scientific literature .70 is generally considered minimally

acceptable (Rice & Harris, 2005; Steyerberg, 2009; Szmukler et al., 2012). It is not possible to recommend a generalised cut-off score for all kinds of validity measures; yet, psychologists should thoroughly consider which type of validity is relevant in each specific case and choose only tests that have the highest validity values.

Psychologists should only collect data that are relevant to the professional tasks that they are undertaking (EFPA, 2005, para. 3.1.2) and, therefore, the incremental validity of tests is important. This form of validity requires that a test should produce accurate information beyond what can be obtained more easily and cheaply with other methods (see Haynes & Lench, 2003; Hunsley & Meyer, 2003). Compared to other contexts, incremental validity is arguably more important in psycholegal contexts, wherein legal and practical restraints usually compel psychologists to collect as little data as possible in the most cost-efficient manner.

It is inevitable that many people who are tested in high-stakes situations will try to manipulate the outcome of an assessment (see Anglim et al., 2018). Examinees might therefore exaggerate or fabricate psychological symptoms for external gain, such as to evade criminal responsibility or attain financial compensation. The exact degree of malingering is unknown, but Mittenberg et al. (2002) have estimated that 19% of examinees in criminal cases exaggerate their symptoms. Cartwright and Roach (2016) furthermore found that 25.4% of their participants admitted having done so after a road traffic accident. Tests used in a psycholegal context should therefore ideally provide a method of checking response style to detect manipulation (see Paulson et al., 2019).

Reliability

Reliability in law generally refers to a legal threshold for the admissibility of evidence (see Edmond, 2012; Freckelton, 2019), but we follow the Daubert standards by using the word

in the scientific sense (see Popper, 1989). The indices of reliability that are relevant in a specific case will depend upon the test and the circumstances. In the psycholegal context, it will generally be test-retest reliability for measuring trait variables and interrater reliability for tests in which professional judgement plays a significant part in combining and interpreting data. Historically, psychologists have considered reliability coefficients above .60 as acceptable in clinical work (see Fleiss et al., 2003), but contemporary authors suggest that reliability should be 'around .90 for clinical decision-making and around .70 for research purposes' (Groth-Marnat & Wright, 2016, p. 10). Court decisions, however, have major consequences for litigants and therefore it is important to minimise the subjective biases of individual forensic assessors, which suggests a minimum interrater coefficient above .80 (see Heilbrun, 1992; Lilienfeld et al., 2000) and ideally above .90, with Nunnally and Bernstein (1994) recommending .95.

Availability of test and support documentation

Heilbrun (1992) requires that tests should be freely available to other psychologists and lawyers involved in the case. There are two reasons for this requirement. First, all expert witnesses must have access to the relevant test and supporting material to prepare to testify (see *Pownall v. Conlan Management*, 1995; *R v. Turner*, 1975). Second, lawyers need access to these materials in order to prepare to cross-examine and evaluate the testimony of expert witnesses. Today this could mean that tests and their supporting documentation should be available commercially or online in official repositories. The supporting information could be in a manual, book or secure website and must inform readers about the test's theoretical basis and development, references to published peer-reviewed research supporting it and its psychometric properties. The manual should also provide information regarding the standardised administration and scoring of the test,

and how to use the available information and population norms to assist with interpretation (Australian Psychological Society, 2014, para. 7.1).

Ethical considerations

The ethical issues that might arise will depend on the test and the circumstances in which it is used, and we therefore add other relevant ethical concerns to our standards. There are some requirements that will be relevant in almost all cases, and the first of these is that there must be evidence that contemporary psychologists accept the test as representing best practice (also see the general acceptance standard in the *Daubert case*). This requirement has two tiers in that the test must be accepted not only by the general profession for general practice but also by psychologists who carry out forensic work. It is important to consider the views of those who undertake psycholegal work because they have the best understanding of the demands of courts and the legal rules that regulate the admission of psycholegal evidence.

Psychologists who carry out psycholegal work should in priority be using tests that provide relevant and reliable data, and consider whether or not using the tests poses a risk of harm to litigants' rights and interests (see Allan, 2013, 2018). Psychologists should also consider the developers' and users' governance of the tests (EFPA, 2005, para. 3.2.2; *Mustac v. Medical Board of Western Australia*, 2007) and important data related to it – for instance, whether or not potential examinees can easily access information that will assist them in manipulating test results. This is becoming increasingly important as it becomes possible to place information on websites (Cartwright et al., 2019). Finally, psychologists should avoid any practice that might bring the profession into disrepute, a risk that is much higher when using tests in psycholegal work as this generally takes place within the public eye and can attract publicity.

Assessment of the Rorschach

Since Hermann Rorschach published the test named after him in his book *Psychodiagnostik* (Rorschach, 1921/1951), many authors have written about its strange origins and development (e.g. Krishnamurthy et al., 2011; Meyer, 2017; Schott, 2014). Many believe that this projective personality test is more effective than its structured self-report counterparts (e.g. Finn, 1996) because it goes beyond conscious and behavioural functioning (e.g. Stenius et al., 2018). Rorschach supporters believe that it circumvents examinees' conscious defences because they respond to ambiguous stimuli with a minimum of instructions (see Siipola & Taylor, 1952). They therefore believe that it provides them information about examinees' automatic processes and unconscious, structural and longitudinal functioning (see Weiner et al., 1996). Furthermore, they claim it can detect (see Cerney, 1990) and assess the authenticity of (see Leavitt & Labott, 1996) recovered repressed traumatic memories about abuse (however, for a critique see Otgaar et al., 2019).

If the assumption about the Rorschach's ability to circumvent examinees' conscious defences is correct then the Court's comment in the case of *F v. Hungary* (2018), as mentioned in the introduction, needs to be considered in every forensic assessment when projective tests are used. Namely, the Court stated, 'it must be considered that consent of the person concerned [...] is not freely given' (*F v. Hungary*, 2018, para. 53) and that tests might be used 'only if they are necessary and genuinely meet the objectives of [...] the European Union or the need to protect the rights and freedoms of others' (*F v. Hungary*, 2018, para. 55). Furthermore, the assumption also raises the question as to whether or not the use of the Rorschach leads to violations of the right to avoid self-incrimination and the right to remain silent when questioned, either prior to or during legal proceedings in a court of law.

The Rorschach is not a unitary test (Exner, 1969) because there are at least eight well-known systems that differ notably due to being developed by people with different theoretical and professional backgrounds (Groth-Marnat & Wright, 2016). Examiners can therefore administer, code and interpret the Rorschach in different ways (Bornstein & Masling, 2005). Exner (1974, 2000, 2008) tried to overcome this confusion by developing the Rorschach Comprehensive System (hereafter Comprehensive System). This is the most frequently used version of the Rorschach and it provides normative information for non-patient adults and children, coupled with statistical tables for some clinical groups (Exner, 2008). Meyer et al. (2007) further published international norms for the Rorschach known as the Composite International Reference Values (CIRV). The Rorschach has always been problematic (e.g. S. J. Beck, 1937), controversial (Dawes, 1996) and the subject of many professional debates (e.g. Hibbard, 2003; Mihura et al., 2015; Wood et al., 2015). The development of the Rorschach continues (e.g. Gurley, 2017), but despite a century's extensive research, development and use, it remains divisive (Groth-Marnat & Wright, 2016).

Test must have a theoretical basis and supporting empirical data that have been peer reviewed and published in scientific journals

The projective hypothesis was the original theoretical assumption behind the Rorschach (Rorschach, 1921/1951). At first, it was based on Freud's (1911) theory that people unconsciously assign their characteristics and impulses to others as defence mechanisms. Nowadays the theoretical underpinning for projective tests such as the Rorschach is the assimilative projective assumption (see Sundberg, 1977). The premise here is that people's understanding, interpretation and explanation of vague and unstructured test stimuli reflect their internal constructs, such as their

feelings, needs, personal experiences, thought processes, conflicts and impulses. However, empirical evidence shows that implicit and explicit processes might interact to determine task performance on implicit tasks (Dunn & Kirsner, 1989). People's expectations (see Sherman et al., 2016) and the social context within which the testing takes place (see Otten et al., 2017) might therefore influence examinees' responses.

There is also little theoretical support for the inferences that examiners draw from various aspects of cards, such as colour and personality. For instance, psychologists associate attention to red stimuli with anger, impaired performance on cognitive tasks, sexualised behaviour, increase aggressiveness, dominance, caution and avoidance behaviour (see Barchard et al., 2017; Tham et al., 2020). It is, however, possible that examinees' responses might reflect learned associations rather than their personalities because it is known that emotion-colour associations vary across cultures (e.g. Hupka et al., 1997).

The Rorschach meets the peer-review requirement because it has attracted much scholarly interest since its publication. A search for the term 'Rorschach' on the PsycINFO© database reveals 6679 peer-reviewed publications between 1921 and 2019, with about 52% of them published from 1980, and an average of 86.02 peer-reviewed manuscripts per year from 1980 to 2019. This amount of research has led to several meta-analyses, which had mostly confirmed the effectiveness of the Rorschach (e.g. Parker et al., 1988). However, more recently the results of these meta-analyses appear to have been affected by publication bias and serious methodological flaws, leading to incorrect and inflated effect sizes (see Erickson et al., 2007; Hunsley & Bailey, 1999; Lilienfeld et al., 2000; Wood et al., 2015). Researchers who want to examine the published studies complain that they can frequently not obtain the relevant quantitative material except by purchasing raw data from the original researchers

(Wood et al., 1996). It is also the case that a researcher's request to examine data was refused even though payment was offered (see Garb et al., 2020). Other researchers have criticised the methodology of several key studies and pointed out that half of the studies with positive findings were unpublished and, therefore, had not been peer reviewed (e.g. Lilienfeld et al., 2000). Lilienfeld et al. (2000) have concluded that despite the volume of research there is a lack of peer-reviewed research that examines the empirical foundation of the Rorschach in general and the Comprehensive System in particular.

Validity of test must be appropriate for the legal question

The organic development of the Rorschach means that it does not have a specific standardisation population, and even supporters of the Rorschach such as Meyer et al. (2015) acknowledge that developers of norms have not given enough attention to the possible effects of the age, gender, culture, education, intelligence and linguistic and socioeconomic background of examinees, and that some of the findings are in conflict. For example, Giromini et al., (2017) found that demographic variables do not influence Rorschach scores, whereas others found significant effects (Delavari et al., 2013; Meyer et al., 2007).

Various features of the Rorschach could influence its validity in cultural and language groups. Emotion-colour associations, for instance, vary across cultures, with envy associated with black, purple and yellow in Russia and black, red and green in the United States (US; see Hupka et al., 1997). French examinees often see a chameleon in card VIII, Scandinavian examinees often see Christmas elves in card II and Japanese examinees often provide a musical instrument-related answer to card VI. All of these answers are unusual responses according to the Comprehensive System (Weiner, 2014). The attempt by Meyer et al. (2007) to overcome the problems of using the Rorschach is welcome but the

norms have been criticised on several grounds (see Gurley, 2017; Meyer et al., 2017).

Several meta-analyses have compared the Rorschach's construct validity with that of the Minnesota Multiphasic Personality Inventory (MMPI; e.g. Garb et al., 1998; Hiller et al., 1999). These studies have typically shown that the Rorschach's variables or indexes have a low average validity of about .30, and further that some variables do not correlate with corresponding MMPI variables (see Krishnamurthy et al., 1996; Lilienfeld et al., 2000). However, due to the publication bias and methodological artefacts observed in these meta-analyses, the .30 value may be overestimated (see Lilienfeld et al., 2000). Researchers have failed to find evidence that the Rorschach can identify a specific psychological problem (e.g. Bartell & Solanto, 1995; Kaplan & Saccuzzo, 2017).

When Comprehensive System variables were correlated with externally (e.g. psychiatric diagnosis) and introspectively (e.g. self-report questionnaires) assessed criteria, the mean validities were .27 and .08, respectively (Mihura et al., 2013). Mihura et al. (2013) found minor support for the validity of Comprehensive System indices such as Severe Perceptual-Thinking, Suicide Constellation and Cognitive Mediation (at least .33), but little or no support for indices like Aggressive Movement, Egocentricity and Coping Style. Wood et al. (2015) replicated the Mihura et al. (2013) study but included the data from unpublished studies to counter the impact of publication bias. They found that several Comprehensive System scores based on complexity/synthesis and productivity differentiate between normal and affected populations. They, however, found no evidence supporting a relationship between Comprehensive System indexes and non-cognitive characteristics such as negative affect and emotionality. Lilienfeld et al. (2000) have further reported that the correlations between the Comprehensive System scores and most psychiatric diagnoses were not replicated in later studies.

The validity of the Rorschach overall is not good (Kaplan & Saccuzzo, 2017). It appears to be useful in diagnosing bipolar disorder, schizophrenia and schizotypal personality disorder (Wood et al., 2000). It is, however, less useful in diagnosing post-traumatic stress disorder (PTSD) and other anxiety disorders, major depressive disorder, suicide attempts, dissociative identity disorder, psychopathy, antisocial personality disorder and dependent, narcissistic (see Exner, 1995) or conduct disorders (see Carlson et al., 1997; Hunsley et al., 2015; Wood et al., 2015). Even examiners who use the Comprehensive System's norms can over-pathologise people (Cooke & Norris, 2011; Costello, 1998), especially those from lower socio-economic groups (Frank, 1994) and children (Hamel et al., 2000). Research also suggests that the Rorschach is not a valid instrument for assessing impulsiveness, criminal behaviour or tendency toward violence, or for detecting child sexual abuse (Lilienfeld et al., 2001). Suggestions that references by examinees to buttocks, anuses, feminine clothing and sex organs in their responses predict homosexuality (e.g. Seitz et al., 1974) have never been supported by any scientific evidence (see APA, 2009).

The incremental validity of projective techniques is generally poor (Garb et al., 2003; Miller & Nickerson, 2006; Ruscio, 2000). This is also true for the Rorschach, with little evidence that most Comprehensive System scores provide information beyond the data gathered by other psychological instruments (Wood et al., 2015).

The Rorschach's lack of a method for checking response style is not a concern for most of its supporters because they believe it is immune to attempts to fake good or bad responses (for references, see Acklin, 2007). For others, the Rorschach is at best relatively resistant (Grossman et al., 2002) or at worst vulnerable to faking attempts that might not be detected through the Rorschach indices (e.g. Elhai et al., 2004; Hartmann & Hartmann,

2014). The Rorschach performance of malingerers were compared to those of patients with schizophrenia (see Albert et al., 1980), psychosis (Albert et al., 1980), depression (Meisner, 1984) and PTSD (Frueh & Kinder, 1994), for example. Overall, the results show that only a few Rorschach variables differ significantly between the participants who malingered and those in the control groups (see Frueh & Kinder, 1994). In addition, the differences found between role-informed malingerer versus control and patient groups in some studies (e.g. Frueh & Kinder, 1994; Overton, 1984) could be explained by prior training, knowledge about the pathology and information about how to behave during administration of the test (Albert et al., 1980; Overton, 1984). Albert et al. (1980) furthermore found that fellows of the Society for Personality Assessment, who could be considered Rorschach experts, could only accurately classify 9% of informed fakers in the malingerer group (see also Sewell & Helle, 2018).

A major problem with the Rorschach is that interpreters put weight on details such as the area or location of the inkblot that examinees focus on when they respond. There are, however, many unknown internal and contextual factors that could influence examinees' perceptions and therefore their responses. For instance, mood (see Matlin, 2012) and affect (e.g. Anderson et al., 2012; Kleckner et al., 2017) influence people's perception of stimuli and judgements of the actual content of perceptions. This phenomenon of selective and context-dependent focus is well known because it is also evident in psychotherapy (Matt et al., 1992), eyewitness memory recall (Loftus, 2004) and everyday situations (see Loeffler et al., 2013). More specifically, Kingery (2003) found that there is a positive correlation between dispositional negative affect and the selection of specific parts of the inkblots. Any deduction made based on the locations that examinees focus on therefore carries little weight without research that

excludes the possibility that factors such as situational-induced mood rather than some underlying personality factor influenced the relevant response.

Reliability must be appropriate for the purpose for which test will be used

A problem with the Rorschach as alluded to earlier is that there are several methods of administering it and of coding and interpreting the results (see Bornstein & Masling, 2005). The administration is in many approaches deliberately vague (e.g. Siipola & Taylor, 1952) in order to allow examinees to respond freely to the ambiguous stimuli. This is problematic because researchers have found that differences in administration can significantly influence examinees' responses (e.g. Lis et al., 2007) and that it can involve something simple, such as whether the cards are presented vertically or horizontally (e.g. D. M. Beck et al., 2005). The outcomes of the different scoring systems can also lead to significantly different interpretations (Kaplan & Saccuzzo, 2017).

Rorschach supporters have traditionally relied on experience and/or training, but neither can substitute for clear, structured and empirically supported instructions (e.g. President's Council of Advisors on Science and Technology, 2016). The best method for minimising error is the scientific method and not clinical experience (Garb et al., 2016), as it is examiners' subjectivity that influences the interpretation of data sets (e.g. Dimsdale, 2015). Harrower (1976) claimed to have shown this by requesting that 10 examiners analyse and identify 17 anonymised Rorschach records of Nazi war criminals that were mixed up with non-Nazis ones. The outcome was that the examiners performed no better than chance and were thus not able to discriminate Nazi war criminals' Rorschach protocols from non-Nazi ones. A limitation of this study is that the examiners did not have the background information that they would generally have in a clinical or psychological

context (Groth-Marnat & Wright, 2016). While trying to reduce the impact of examiners' subjectivity in the coding process by providing working tables, Exner (2008) conceded that it would be naïve to expect objective coding (see also Acklin et al., 2000).

Research conducted in the 1980s and 1990s therefore predictably yielded contradicting and confusing results, but Parker's (1983) meta-analysis of 39 research papers nevertheless found an overall internal reliability coefficient of .83. Garb et al. (1998) criticised Parker's method, but methodologically sound research has made compatible findings with median test-retest coefficients of .80 reported (e.g. Acklin et al., 2000). The interrater reliability of the various Comprehensive System variables is even slightly better, with Exner (2003) reporting that the interrater scoring reliability of all variables was above .85 and Meyer's (1997) meta-analysis of 16 studies finding interrater reliability coefficients ranging from .72 to .96. Other researchers, however, have found lower coefficients (e.g. W. Perry et al., 1995; Wood et al., 1996) and criticised Meyer's study as flawed. Acklin et al. (2000) found that half of the Comprehensive System variables attained reliability coefficients of .85 or above whilst some of the others, such as the Schizophrenia Index (SCZI), are as low as .45 with a median retest coefficient of .80. A later study by Sultan et al. (2006), who retested 75 French non-patient adults after 3 months using the 47 Comprehensive System variables, found that 9 coefficients were above .70 and 21 yielded moderate reliability of above .50, while the median correlation coefficient was .53.

Availability of test and supporting documentation

The Rorschach is commercially available to examiners, and each of the systems give administering, scoring and coding instructions (see Bornstein & Masling, 2005). It serves little purpose to examine the supporting documentation of each system, but Exner's (2008)

workbook for the extensively used Comprehensive System is arguably the most informative. It provides a standardised method of administration, scoring and interpretation, as well as information about the test's psychometric properties and population norms. Several researchers have nevertheless criticised the composition of the normative sample (Hibbard, 2003; Viglione & Giromini, 2016).

Ethical requirements

The first ethical question regarding the Rorschach is whether or not it is generally accepted by psychologists as representing best practice and as being appropriate for psycholegal practice. It is difficult to give a definite answer because it is one of the most divisive tests used in psychology, with many supporters and detractors (Groth-Marnat & Wright, 2016). In making an award for distinguished professional contributions to John E. Exner, the APA (1998) described the Rorschach as probably the most powerful psychometric tool ever invented, and many others share the belief that it is the equivalent of a psychological X-ray test (e.g. Brown, 1992; Piotrowski, 1980). Others criticise it as a pseudoscience (e.g. Fontan & Andronikof, 2017; Wood et al., 2008) and criticise its use in clinical (e.g. Huberty, 2019; Hunsley & Bailey, 2001; Lilienfeld, 2015) and forensic (e.g. Garb, 1999; Wood et al., 2001) settings due to its poor psychometric characteristics. Surveys of forensic practitioners in North America have nevertheless revealed that the Rorschach was the most frequently used unstructured projective test, with up to 36% of research participants reporting that they had used it (see Archer et al., 2006). More recent North American research, however, shows that the frequency of Rorschach use has dropped down to 20% (Viljoen et al., 2010) or even to 3% (Neal & Grisso, 2014). The acceptance of the Rorschach amongst forensic psychologists in other countries might never have been good. Martin et al. (2001), for instance, found in

their survey that Australian psychologists doing court work used the Rorschach less frequently than other tests; it had a *weighted use score* of only 23 compared to the 141 score of the Wechsler Intelligence Scale and the 98 of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2).

The limitations of the Rorschach discussed above further raise questions about how confident psychologists who use it can be that the data they provide to courts are objective, relevant and reliable (see Erickson et al., 2007; Iudici et al., 2015; Neal et al., 2019). They might therefore find it difficult to demonstrate that they were taking reasonable steps to prevent reasonably foreseeable and preventable harm such as was present in *F v. Hungary* (2018), where the applicant ran the risk of being returned to a country where he could have been persecuted.

The governance of the Rorschach is also in question because there is so much information available about it in publications and the media. There are over 20 million references associated with the term 'Rorschach' on the Internet and it has its own English Wikipedia page (Wikipedia, 2021), with descriptions of the most common responses extracted from Weiner (2003) and Weiner and Greene (2007). Schultz and Loving (2012) concluded that 19% of the Rorschach-related information they found on the Internet posed a direct threat to the security of the test. Schultz and Brabender (2013) assessed the influence of the Rorschach Wikipedia page's content on the protocols of non-patients who were instructed to act as parents who want to keep their children, and thus, fake-good. They showed that, relative to an uninformed (control) group, well-educated non-patients benefited from exposure to Wikipedia content. Researchers were, however, not able to replicate these findings with psychiatric outpatients (Hartmann & Hartmann, 2014) or incarcerated violent offenders (Nørbech et al., 2016). Patients in the informed group failed to present as mentally well adjusted on the Rorschach but

they were able to inhibit provocative, aggressive and dramatic responses relative to other patient groups (Hartmann & Hartmann, 2014).

The use of the Rorschach in court further weakens the credibility of the profession and undermines courts', lawyers' and the public's perception of the profession as being trustworthy. This is demonstrated by remarks made by the judges in *F v. Hungary* (2018), lawyers who commented on the case (e.g. Ferreira & Venturi, 2018) and journalists from reputable outlets such as the *New York Times* (Goode, 2001), *New Scientist* (Wilson, 2020) and BBC News ('What's behind the Rorschach inkblot test?', 2012).

A feature of the administration of the Rorschach which requires consideration is that examiners often repeat requests for extensive answers, and may make negative inferences if examinees fail to provide enough information (Exner, 2008). There is a risk that examinees in forensic contexts will therefore feel coerced by examiners, whom they perceive as authority figures. It is notable that the judges in *F v. Hungary* (2018) found that consent for psycholegal assessment in that specific type of case can never be voluntary.

Discussion

F v. Hungary (2018) is a reminder to psychologists who do psycholegal work of the impact their reports can have on the rights and interests of people and that lawyers and courts will therefore approach their reports and testimony critically. Psychologists should consequently ensure that they use tests which can withstand the scrutiny of both their peers and lawyers. Some might therefore regret that the judges in *F v. Hungary* (2018) did not set out clear standards that can be used to consider the credibility of a test – but it does nevertheless provide the profession with an opportunity to formulate and publish a set of generally accepted standards that they and lawyers can refer to when assessing the acceptability of a test. We believe that most psychologists who

undertake forensic work might consciously or non-consciously apply the standards that we have identified when deciding whether or not to use a specific test to collect data. We deliberately did not add details such as what would be an acceptable test-retest correlation coefficient for two reasons. First, it would be more appropriate to do so after a debate by psycholegal professionals. Second, it is always at a court's discretion whether or not it allows testimony, and courts will therefore not be bound by our standards or by any details that the profession adds. Courts and lawyers will nevertheless most likely consider standards and guidelines that represent psychologists' general views.

The proposed standards are primarily aimed at assessing the appropriateness of emerging theories, techniques and tests in court, which makes their application to a well-established and well-researched test such as the Rorschach more difficult but nevertheless not impossible. A major concern regarding the Rorschach is that it lacks a theoretical basis supported by empirical data (e.g. Krishnamurthy et al., 2011) and that it is rather based on subjective common-sense thinking than on scientific research. The strategies of Exner (2008) and others to develop norms and train examiners is no substitute for an empirically tested theory (President's Council of Advisors on Science and Technology, 2016) and is contrary to the contemporary emphasis on evidence-based methods in assessments (see Stewart & Chambless, 2007).

The Comprehensive System's construct validity is another major problem because a test does not have a place in psycholegal work if it does not measure what it is supposed to measure. Groth-Marnat and Wright (2016) concluded that the overall validity of the Rorschach is moderate (between .30 and .50). It is difficult to draw a conclusive finding about the validity of many Comprehensive System indices because some have not yet been studied well, and the accuracy of the

studies that have examined the validity of other indices is contested (Garb et al., 2005). An expert witness will therefore find it difficult to justify using even the Comprehensive System, especially given that it lacks an acceptable method of measuring response style and that research data show that the Rorschach is susceptible to faking (e.g. Perry & Kinder, 1990).

The test-retest reliability coefficients of the Rorschach indices that range from .27 to .94 (see Meyer, 1997) limit its application in forensic work where the stakes for people are high, and this could explain why Rorschach experts disagree when they give forensic opinions (see Guarnera et al., 2017). The Rorschach meets the availability standard and Exner's (2008) book provides most of the information required regarding the Comprehensive Systems.

A major hurdle for anybody who plans to use the Rorschach in court is, however, whether or not it would be ethical to do so. Our assessment is that there is no general acceptance amongst forensic psychologists internationally that it is a test that should be used in court (e.g. Archer et al., 2016). Psychologists who want to use the Rorschach test should tell courts that it is highly controversial and not widely accepted (Lilienfeld et al., 2000), but this is unlikely to happen.

The evidence we have reviewed further suggests that the Rorschach does not provide objective, relevant and reliable data and therefore that psychologists who use it cannot exclude the possibility that their reports and testimony will cause reasonably foreseeable and avoidable harm. The governance of the test is also poor, and this is particularly concerning given that it lacks a method of measuring response style. Psychologists' use of the Rorschach in court further appears to impact on the credibility of the whole profession and might constitute non-proportional violation of examinees' autonomy. Our overall conclusion is therefore that – despite its popularity in some European countries (see Areh, 2020) – the Rorschach does not meet the standards

which we believe the international scientific community requires of tests used in psycholegal work, and consequently that psychologists should not use it in court (but for a contrary view, see Weiner et al., 1996).

In conclusion, psychologists should only use tests whose reliability and validity have been demonstrated rigorously. We expect that some other commonly used techniques and tests might also be found wanting if the standards set out above are applied to them. We do not think that these standards are unnecessarily arduous because if psychologists expect the public and more specifically legal professionals to trust them and their testimony, they must ensure that they use tests and techniques that are trustworthy and that do not violate the human rights of the persons assessed. In the case of *F v. Hungary* (2018), the Court of Justice of the European Union ruled that an expert's report is only acceptable if it is based on sufficiently reliable methods which meet standards recognised by the international scientific community. The Court did not convey what these standards are or decide whether or not projective tests are acceptable. This might be due to the Court's trust in psychology as a profession, and other sciences, to fulfil its expectations. Should the profession fail to do so, a court could set standards for psychologists or it might decide that psychological expert opinions are not reliable enough to be received as scientific evidence.

Acknowledgements

The authors thank Dr Maria Allan for her comments on drafts of the paper.

Ethical standards

Declaration of conflicts of interest

Igor Areh has declared no conflict of interests. Fanny Verkampt has declared no conflict of interests.

Alfred Allan has declared no conflict of interests.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

ORCID

Igor Areh  <http://orcid.org/0000-0001-5410-9907>

Fanny Verkampt  <http://orcid.org/0000-0003-0123-6464>

Alfred Allan  <http://orcid.org/0000-0001-7039-797X>

References

- Acklin, M. W. (2007). The Rorschach test and forensic psychological evaluations: Psychosis and the insanity defense. In C. B. Gacono & F. B. Evans (Eds.), *The handbook of forensic Rorschach assessment* (pp. 157–174). Routledge.
- Acklin, M. W., McDowell, C. J., Verschell, M. S., & Chan, D. (2000). Interobserver agreement, intraobserver reliability, and the Rorschach comprehensive system. *Journal of Personality Assessment*, 74(1), 15–47. <https://doi.org/10.1207/s15327752jpa740103>
- Albert, S., Fox, H. M., & Kahn, M. W. (1980). Faking psychosis on the Rorschach: Can expert judges detect malingering? *Journal of Personality Assessment*, 44(2), 115–119. https://doi.org/10.1207/s15327752jpa4402_1
- Allan, A. (2013). Ethics in correctional and forensic psychology: Getting the balance right. *Australian Psychologist*, 48(1), 47–56. <https://doi.org/10.1111/j.1742-9544.2012.00079.x>
- Allan, A. (2018). Moral challenges for psychologists working in psychology and law. *Psychiatry, Psychology, and Law*, 25(3), 485–499. <https://doi.org/10.1080/13218719.2018.1473173>
- Allan, A. (2020). Structuring the debate about research ethics in the psychology and law field: An international perspective. *Psychiatry, Psychology and Law*, 27(3), 397–411. <https://doi.org/10.1080/13218719.2020.1742243>
- Allan, A., Dawson, D., & Allan, M. M. (2006). Prediction of the risk of male sexual reoffending in Australia. *Australian Psychologist*, 41(1), 60–68. <https://doi.org/10.1080/00050060500391886>
- Allan, A., Parry, C. L., Ferrante, A., Gillies, C., Griffiths, C. S., Morgan, F., Smallbone, S. W., Tubex, T., Spiranic, C., & Wong, S. C. P. (2019). Assessing the risk of Australian Indigenous sexual offenders reoffending: A review of the research literature and court decisions. *Psychiatry, Psychology, and Law*, 26(2), 274–294. <https://doi.org/10.1080/13218719.2018.1504242>
- American Educational Research Association. (2014). *Standards for educational and psychological testing*. AERA.
- American Psychological Association. (1998). Awards for distinguished professional contributions: John E. Exner, Jr. *American Psychologist*, 53(4), 391–392. <https://doi.org/10.1037/0003-066x.53.4.391>
- American Psychological Association. (2009). *Report of the American Psychological Association Task Force on appropriate therapeutic responses to sexual orientation*. <http://www.apa.org/pi/lgbcc/publications/therapeutic-resp.html>
- American Psychological Association. (2013). Specialty guidelines for forensic psychology. *American Psychologist*, 68(1), 7–19. <https://doi.org/10.1037/a0029889>
- Anderson, E., Siegel, E., White, D., & Barrett, L. F. (2012). Out of sight but not out of mind: Unseen affective faces influence evaluations and social impressions. *Emotion (Washington, D.C.)*, 12(6), 1210–1221. <https://doi.org/10.1037/a0027514>
- Anglim, J., Bozic, S., Little, J., & Lievens, F. (2018). Response distortion on personality tests in applicants: Comparing high-stakes to low-stakes medical settings. *Advances in Health Sciences Education: Theory and Practice*, 23(2), 311–321. <https://doi.org/10.31234/osf.io/dpcx6>
- Archer, R. P., Buffington-Vollum, J. K., Vauter Stredny, R., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment*, 87(1), 84–94. https://doi.org/10.1207/s15327752jpa8701_07
- Archer, R. P., Wheeler, E. M. A., & Vauter, R. A. (2016). Empirically supported forensic assessment. *Clinical Psychology: Science and Practice*, 23(4), 348–364. <https://doi.org/10.1111/cpsp.12171>
- Areh, I. (2020). Forensic assessment may be based on common sense assumptions rather than science. *International Journal of Law and Psychiatry*, 71, 101607. <https://doi.org/10.1016/j.ijlp.2020.101607>

- Australian Psychological Society. (2013). *Ethical guidelines for psychological practice in forensic contexts*. APS. <https://www.psychology.org.au/for-members/resource-finder/resources/ethics/Ethical-guidelines-psychological-practice-forensic>.
- Australian Psychological Society. (2014). *Ethical guidelines for psychological assessment and the use of psychological tests*. APS. <https://www.psychology.org.au/for-members/resource-finder/resources/ethics/>
- Barchard, K. A., Grob, K. E., & Roe, M. J. (2017). Is sadness blue? The problem of using figurative language for emotions on psychological tests. *Behavior Research Methods*, 49(2), 443–456. <https://doi.org/10.3758/s13428-016-0713-5>
- Bartell, S. S., & Solanto, M. V. (1995). Usefulness of the Rorschach inkblot test in assessment of attention deficit hyperactivity disorder. *Perceptual and Motor Skills*, 80(2), 531–541. <https://doi.org/10.2466/pms.1995.80.2.531>
- Beck, D. M., Pinsk, M. A., & Kastner, S. (2005). Symmetry perception in humans and macaques. *Trends in Cognitive Sciences*, 9(9), 405–406. <https://doi.org/10.1016/j.tics.2005.07.002>
- Beck, S. J. (1937). Some present Rorschach problems. *Rorschach Research Exchange*, 2(1), 15–22. <https://doi.org/10.1080/08934037.1937.10381217>
- Beecher-Monas, E. (1998). Blinded by science: How judges avoid the science in scientific evidence. *Temple Law Review*, 71, 55–102.
- Bornstein, R. F. & Masling, J. M. (Eds.) (2005). *Scoring the Rorschach: Seven validated systems*. Lawrence Erlbaum Associates.
- Bow, J. N., Gould, J. W., Flens, J. R., & Greenhut, D. (2006). Testing in child custody evaluations –selection, usage, and Daubert admissibility: A survey of psychologists. *Journal of Forensic Psychology Practice*, 6(2), 17–38. https://doi.org/10.1300/J158v06n02_02
- Brown, J. (1992). *The definition of a profession: The authority of metaphor in the history of intelligence testing, 1890–1930*. Princeton University Press.
- Carlson, C. R., Kula, M. L., & St. Laurent, C. M. (1997). Rorschach revised DEPI and GDI with inpatient major depressives and borderline personality disorder with major depression: Validity issues. *Journal of Clinical Psychology*, 53(1), 51–58. [https://doi.org/10.1002/\(sici\)1097-4679\(199701\)53:1%3C51::aid-jclp7%3E3.0.co;2-y](https://doi.org/10.1002/(sici)1097-4679(199701)53:1%3C51::aid-jclp7%3E3.0.co;2-y) [https://doi.org/10.1002/\(SICI\)1097-4679\(199701\)53:1<51::AID-JCLP7>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-4679(199701)53:1<51::AID-JCLP7>3.0.CO;2-Y)
- Cartwright, A., & Roach, J. (2016). Fraudulently claiming following a road traffic accident: A pilot study of UK residents' attitudes. *Psychiatry, Psychology and Law*, 23(3), 446–461. <https://doi.org/10.1080/13218719.2015.1080148>
- Cartwright, A., Roach, J., & Armitage, R. (2019). Mission impossible? Assessing the veracity of a mental health problem as result of a road traffic accident: A preliminary review of UK experts' practices. *Journal of Forensic Psychiatry & Psychology*, 30(1), 89–111. <https://doi.org/10.1080/14789949.2018.1502338>
- Cerney, M. S. (1990). The Rorschach and traumatic loss: Can the presence of traumatic loss be detected from the Rorschach? *Journal of Personality Assessment*, 55(3–4), 781–789. <https://doi.org/10.1080/00223891.1990.9674112>
- Cooke, G., & Norris, D. M. (2011). Child custody and parental fitness. In E. Y. Drogin, F. M. Dattilio, R. L. Sadoff, & T. G. Gutheil (Eds.), *Handbook of forensic assessment: Psychological and psychiatric perspectives* (pp. 433–458). John Wiley & Sons. <https://doi.org/10.1002/9781118093399>
- Costello, R. M. (1998). Psychometric definition of Rorschach determinant component structure. *European Journal of Psychological Assessment*, 14(2), 116–123. <https://doi.org/10.1027/1015-5759.14.2.116>
- Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579. (1993). <https://supreme.justia.com/cases/federal/us/509/579/>
- Dawes, R. M. (1996). *House of cards: Psychology and psychotherapy built on myth*. The Free Press.
- Delavari, M., Shairi, M., & Asghari-Moghadam, M. (2013). Role of culture and gender in Rorschach findings in 9 year old Iranian children. *Procedia – Social and Behavioral Sciences*, 84, 1565–1570. <https://doi.org/10.1016/j.sbspro.2013.06.789>
- Dimsdale, J. E. (2015). Use of Rorschach tests at the Nuremberg war crimes trial: A forgotten chapter in history of medicine. *Journal of Psychosomatic Research*, 78(6), 515–518. <https://doi.org/10.1016/j.jpsychores.2015.04.001>
- Dunn, J. C., & Kirsner, K. (1989). Implicit memory: Task or process? In S. Lewandowsky, J. C. C. Dunn, & K. Kirsner (Eds.), *Implicit memory: Theoretical issues* (pp. 17–31). Lawrence Erlbaum Associates.

- Edmond, G. (2012). Is reliability sufficient? The Law Commission and expert evidence in international and interdisciplinary perspective (Part 1). *International Journal of Evidence & Proof*, 16(1), 30–65. <https://doi.org/10.1350/ijep.2012.16.1.391>
- Elhai, J. D., Kinder, B. N., & Frueh, B. C. (2004). Projective assessment of malingering. In M. J. Hilsenroth, D. L. Segal, & M. Hersen (Eds.), *Comprehensive handbook of psychological assessment*, Volume 2: *Personality Assessment* (pp. 553–561). John Wiley & Sons.
- Erickson, S. K., Lilienfeld, S. O., & Vitacco, M. J. (2007). Failing the burden of proof: The science and ethics of projective tests in custody evaluations. *Family Court Review*, 45(2), 185–192. <https://doi.org/10.1111/j.1744-1617.2007.00138.x>
- European Federation of Psychologists' Associations. (2005). *Meta-code of ethics*. <http://www.efpa.eu/ethics/ethical-codes>
- Exner, J. E. (1969). *The Rorschach systems*. Grune and Statton.
- Exner, J. E. (1974). *The Rorschach: A comprehensive system*. John Wiley & Sons.
- Exner, J. E. (1995). *Issues and methods in Rorschach research*. Lawrence Erlbaum Associates.
- Exner, J. E. (2000). *A primer for Rorschach interpretation*. Rorschach Workshops.
- Exner, J. E. (2003). *The Rorschach: A comprehensive system* (Vol. 1) (4th ed.). John Wiley & Sons.
- Exner, J. E. (2008). *A Rorschach workbook for the Comprehensive System* (5th ed.). Rorschach Workshops.
- Faigman, D. L. (1995). The evidentiary status of social science under Daubert: Is it 'scientific,' 'technical,' or 'other' knowledge? *Psychology, Public Policy, and Law*, 1(4), 960–979. <https://doi.org/10.1037/1076-8971.1.4.960>
- Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, 7(6), 555–561. <https://doi.org/10.1177/1745691612459059>
- Ferreira, N., & Venturi, D. (2018). *Testing the untestable: The CJEU's decision in Case C-473/16, F v. Bevándorlási és Állampolgársági Hivatal (June 28, 2018)*. European Database of Asylum Law. <https://ssrn.com/abstract=3204321>.
- Finn, S. E. (1996). Assessment feedback integrating MMPI-2 and Rorschach findings. *Journal of Personality Assessment*, 67(3), 543–557. https://doi.org/10.1207/s15327752jpa6703_10
- Fleiss, J. L., Levin, B., & Cho Paik, M. (2003). *Statistical methods for rates & proportions* (3rd ed.). John Wiley & Sons.
- Fontan, P., & Andronikof, A. (2017). Complexity, pseudoscience and Rorschach performance assessment system. *Clinical and Experimental Psychology*, 3(1), 1–5. <https://doi.org/10.4172/2471-2701.1000145>
- Frank, G. (1994). Socioeconomic status and the Rorschach. *Psychological Reports*, 74(1), 95–98. <https://doi.org/10.2466/pr074.1.95>
- Freckelton, I. (2019). *Expert evidence: Law, practice, procedure and advocacy*. Thomson Reuters.
- Freud, S. (1911). Psycho-analytic notes on an autobiographical account of a case of paranoia (dementia paranoides). In *Standard edition of the complete works of Sigmund Freud* (Vol. 12, J. Strachey, Trans., pp. 9–79). Hogarth.
- Frueh, B. C., & Kinder, B. N. (1994). The susceptibility of the Rorschach inkblot test to malingering of combat-related PTSD. *Journal of Personality Assessment*, 62(2), 280–298. https://doi.org/10.1207/s15327752jpa6202_9
- Frye v. United States 293 F 1013 (DC Cir) (1923).
- F v. Bevándorlási és Állampolgársági Hivatal [F v. Hungary]. (2018). *CJEU Case C-473/416*. <http://curia.europa.eu/juris/document/document.jsf?text=&docid=198766&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=865294>
- Garb, H. N. (1999). Call for a moratorium on the use of the Rorschach inkblot test in clinical and forensic settings. *Assessment*, 6(4), 313–318. <https://doi.org/10.1177/107319119900600402>
- Garb, H. N., Florio, C. M., & Grove, W. M. (1998). The validity of the Rorschach and the Minnesota Multiphasic Personality Inventory: Results from meta-analyses. *Psychological Science*, 9(5), 402–404. <https://doi.org/10.1111/1467-9280.00075>
- Garb, H. N., Lilienfeld, S. O., & Fowler, K. A. (2016). Psychological assessment and clinical judgment. In J. E. Maddux & B. A. Winstead (Eds.), *Psychopathology: Foundations for a contemporary understanding* (4th ed., pp. 111–126). Routledge.
- Garb, H. N., Lilienfeld, S. O., & Fowler, K. A. (2020). Psychological assessment and clinical judgment. In J. E. Maddux & B. A. Winstead (Eds.), *Psychopathology:*

- Foundations for a contemporary understanding* (5th ed., pp. 125–140). Routledge.
- Garb, H. N., Wood, J. M., Lilienfeld, S. O., & Nezworski, M. T. (2005). Roots of the Rorschach controversy. *Clinical Psychology Review, 25*(1), 97–118. <https://doi.org/10.1016/j.cpr.2004.09.002>
- Garb, H. N., Wood, J. M., Nezworski, M. T., & Lilienfeld, S. O. (2003). The Rorschach inkblot test, fortune tellers, and cold reading. *Skeptical Inquirer, 27*(4), 29–34. <https://skepticalinquirer.org/2003/07/the-rorschach-inkblot-test-fortune-tellers-and-cold-reading/>
- Giromini, L., Viglione, D. J., Zennaro, A., & Cauda, F. (2017). Neural activity during production of Rorschach responses: An fMRI study. *Psychiatry Research: Neuroimaging, 262*, 25–31. <https://doi.org/10.1016/j.psychres.2017.02.001>
- Goode, E. (2001, February 20). What's in an inkblot? Some say, not much. *New York Times*. <https://www.nytimes.com/2001/02/20/science/what-s-in-an-inkblot-some-say-not-much.html>
- Gould, A. (2019, 4 June). The gay test: Medicalisation and queer asylum seekers. *EMA Human Rights Blog*. <http://www.ema-humanrights.org/2019/06/04/the-gay-test-medicalisation-and-queer-asylum-seekers/>
- Gould, J. W., Martindale, D. A., & Flens, J. R. (2013). Responsible use of psychological tests: Ethical and professional practice concerns. In D. H. Saklofske, C. R. Reynolds, & V. Schwane (Eds.), *The Oxford Handbook of Child Psychological Assessment* (pp. 222–237). Oxford Handbooks Online. <https://doi.org/10.1093/oxfordhb/9780199796304.013.0011>
- Grisso, T. (1986). *Evaluating competencies: Forensic assessments and instruments*. Plenum Press.
- Grossman, L., Wasyliv, O., Benn, A., & Gyoerkoe, K. (2002). Can sex offenders who minimize on the MMPI conceal psychopathology on the Rorschach? *Journal of Personality Assessment, 78*(3), 484–501. https://doi.org/10.1207/S15327752JPA7803_07
- Groth-Marnat, G., & Wright, A. J. (2016). *Handbook of psychological assessment* (6th ed.). John Wiley & Sons.
- Guamera, 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>
- Gurley, J. R. (2017). *Essentials of Rorschach assessment: Comprehensive system and R-PAS*. John Wiley & Sons.
- Hamel, M., Shaffer, T. W., & Erdberg, P. (2000). A study of nonpatient preadolescent Rorschach protocols. *Journal of Personality Assessment, 75*(2), 280–294. https://doi.org/10.1207/S15327752JPA7502_8
- Harrower, M. (1976). Rorschach records of the Nazi war criminals: An experimental study after thirty years. *Journal of Personality Assessment, 40*(4), 341–351. https://doi.org/10.1207/s15327752jpa4004_1
- Hartmann, E., & Hartmann, T. (2014). The impact of exposure to Internet-based information about the Rorschach and the MMPI-2 on psychiatric outpatients' ability to simulate mentally healthy test performance. *Journal of Personality Assessment, 96*(4), 432–444. <https://doi.org/10.1080/00223891.2014.882342>
- Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment, 15*(4), 456–466. <https://doi.org/10.1037/1040-3590.15.4.456>
- Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior, 16*(3), 257–272. <https://doi.org/10.1007/BF01044769>
- Hibbard, S. (2003). A critique of Lilienfeld et al.'s (2000) 'The scientific status of projective techniques'. *Journal of Personality Assessment, 80*(3), 260–271. https://doi.org/10.1207/S15327752JPA8003_05
- Hiller, J. B., Rosenthal, R., Bornstein, R. F., Berry, D. T. R., & Brunell-Neuleib, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment, 11*(3), 278–296. <https://doi.org/10.1037/1040-3590.11.3.278>
- Huberty, T. (2019). Depression. In S. Hupp (Ed.), *Pseudoscience in child and adolescent psychotherapy: A skeptical field guide* (pp. 124–139). Cambridge University Press. <https://doi.org/10.1017/9781316798096.011>
- Hunsley, J., & Bailey, J. M. (1999). The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. *Psychological Assessment, 11*(3), 266–277. <https://doi.org/10.1037/1040-3590.11.3.266>
- Hunsley, J., & Bailey, J. M. (2001). Whither the Rorschach? An analysis of the evidence. *Psychological Assessment, 13*(4), 472–485. <https://doi.org/10.1037/1040-3590.13.4.472>
- Hunsley, J., Lee, C. M., Wood, J. M., & Taylor, W. (2015). Controversial and questionable assessment techniques. In S. O. Lilienfeld,

- S. Y. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 42–82). Guilford Press.
- Hunsley, J., & Meyer, G. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment, 15*(4), 446–455. <https://doi.org/10.1037/1040-3590.15.4.446>
- Hupka, R. B., Zaleski, Z., Otto, J., Reidl, L., & Tarabrina, N. V. (1997). The colours of anger, envy, fear, and jealousy: A cross-cultural study. *Journal of Cross-Cultural Psychology, 28*(2), 156–171. <https://doi.org/10.1177/0022022197282002>
- International Test Commission. (2013). *International guidelines for test use*. International Test Commission. https://www.intestcom.org/files/guideline_test_use.pdf
- Iudici, A., Salvini, A., Faccio, E., & Castelnuovo, G. (2015). The clinical assessment in the legal field: An empirical study of bias and limitations in forensic expertise. *Frontiers in Psychology, 6*, 1831. <https://doi.org/10.3389/fpsyg.2015.01831>
- Kaplan, R. M., & Saccuzzo, D. P. (2017). *Psychological testing: Principles, applications, and issues* (9th ed.). Cengage Learning.
- Kingery, L. R. (2003). *The psychological correlates of asymmetric cerebral activation* [Doctoral dissertation, University of Maine]. Electronic Theses and Dissertations, 54. University of Maine, Department of Psychology. <https://digitalcommons.library.umaine.edu/etd/54/>
- Kleckner, I. R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W. K., Quigley, K. S., Dickerson, B. C., & Feldman Barrett, L. (2017). Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nature Human Behaviour, 1*(5), 0069. <https://doi.org/10.1038/s41562-017-0069>
- Krishnamurthy, R., Archer, R. P., & Groth-Marnat, G. (2011). The Rorschach and performance-based assessment. In T. M. Harwood, L. E. Beutler, & G. Groth-Marnat (Eds.), *Integrative assessment of adult personality* (3rd ed., pp. 276–328). The Guilford Press.
- Krishnamurthy, R., Archer, R. P., & House, J. J. (1996). The MMPI-A and Rorschach: A failure to establish convergent validity. *Assessment, 3*(2), 179–191. <https://doi.org/10.1177/107319119600300210>
- Kumho Tire Company v. Carmichael 566 US 137. (1999). <https://www.law.cornell.edu/supct/pdf/97-1709P.ZS>
- Leavitt, F., & Labott, S. M. (1996). Authenticity of recovered sexual abuse memories: A Rorschach study. *Journal of Traumatic Stress, 9*(3), 483–496. <https://doi.org/10.1002/jts.2490090307>
- Lee, D., Reynolds, C. R., & Willson, V. L. (2003). Standardized test administration. *Journal of Forensic Neuropsychology, 3*(3), 55–81. https://doi.org/10.1300/J151v03n03_04
- Lilienfeld, S. O. (2015). Introduction to Special Section on Pseudoscience in Psychiatry. *Canadian Journal of Psychiatry, 60*(12), 531–533. <https://doi.org/10.1177/070674371506001202>
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1*(2), 27–66. <https://doi.org/10.1111/1529-1006.002>
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2001). What's wrong with this picture? *Scientific American, 284*(5), 80–87. <https://doi.org/10.1038/scientificamerican0501-80>
- Lis, A., Parolin, L., Calvo, V., Zennaro, A., & Meyer, G. (2007). The impact of administration and inquiry on Rorschach Comprehensive System protocols in a national reference sample. *Journal of Personality Assessment, 89*(sup1), S193–S200. <https://doi.org/10.1080/00223890701583614>
- Loeffler, S. N., Myrtek, M., & Peper, M. (2013). Mood-congruent memory in daily life: Evidence from interactive ambulatory monitoring. *Biological Psychology, 93*(2), 308–315. <https://doi.org/10.1016/j.biopsycho.2013.03.002>
- Loftus, E. F. (2004). Memories of things unseen. *Current Directions in Psychological Science, 13*(4), 145–147. <https://doi.org/10.1111/j.0963-7214.2004.00294.x>
- Marlowe, D. B. (1995). A hybrid decision framework for evaluating psychometric evidence. *Behavioral Sciences & the Law, 13*(2), 207–228. <https://doi.org/10.1002/bsl.2370130205>
- Martin, M., Allan, A., & Allan, M. M. (2001). The use of psychological tests by Australian psychologists who do assessments for the courts. *Australian Journal of Psychology, 53*(2), 77–82. <https://doi.org/10.1080/13218710009524981> <https://doi.org/10.1080/00049530108255127>

- Matlin, M. W. (2012). Pollyanna principle. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* (pp. 255–272). Psychology Press.
- Matt, G. E., Vázquez, C., & Campbell, W. K. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review, 12*(2), 227–255. [https://doi.org/10.1016/0272-7358\(92\)90116-P](https://doi.org/10.1016/0272-7358(92)90116-P)
- McCann, J. T., & Evans, F. B. (2008). Admissibility of the Rorschach. In C. B. Gacono, F. B. Evans, N. Kaser-Boyd, & L.A. Gacono (Eds.), *The handbook of forensic Rorschach assessment* (pp. 55–78). Routledge.
- Meisner, J. S. (1984). *Susceptibility of Rorschach depression correlates to malingering* [Doctoral dissertation, California School of Professional Psychology]. Dissertation Abstracts International, 45, 3951B. Alliant International University, The California School of Professional Psychology.
- Melton, G. B., Pertrila, J., Poythress, N. G., Slobogin, C., Otto, R. K., Mossman, D., & Condie, L. O. (2018). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (4th ed.). Guilford Press.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Meyer, G. J. (1997). Assessing reliability: Critical corrections for a critical examination of the Rorschach Comprehensive System. *Psychological Assessment, 9*(4), 480–489. <https://doi.org/10.1037/1040-3590.9.4.480>
- Meyer, G. J. (2017). What Rorschach performance can add to assessing and understanding personality. *International Journal of Personality Psychology, 3*(1), 36–49. <https://ijpp.rug.nl/article/view/29881>
- Meyer, G. J., Erdberg, P., & Shaffer, T. W. (2007). Towards international normative reference data for Comprehensive System. *Journal of Personality Assessment, 89*(sup1), S201–S216. <https://doi.org/10.1080/00223890701629342>
- Meyer, G. J., Giromini, L., Viglione, D. J., Reese, J. B., & Mihura, J. L. (2015). The association of gender, ethnicity, age, and education with Rorschach scores. *Assessment, 22*(1), 46–64. <https://doi.org/10.1177/1073191114544358>
- Meyer, G. J., Viglione, D. J., & Mihura, J. L. (2017). Psychometric foundations of the Rorschach Performance Assessment System (R-PAS). In R. E. Erard & F. B. Evans (Eds.), *The Rorschach in multimethod forensic assessment: Conceptual foundations and practical applications* (pp. 23–91). Routledge.
- Mihura, J. L., Meyer, G. J., Bombel, G., & Dumitrascu, N. (2015). Standards, accuracy, and questions of bias in Rorschach meta-analyses: Reply to Wood, Garb, Nezworski, Lilienfeld, and Duke (2015). *Psychological Bulletin, 141*(1), 250–260. <https://doi.org/10.1037/a0038445>
- Mihura, J. L., Meyer, G. J., Dumitrascu, N., & Bombel, G. (2013). The validity of individual Rorschach variables: Systematic reviews and meta-analyses of the comprehensive system. *Psychological Bulletin, 139*(3), 548–605. <https://doi.org/10.1037/a0029406> <https://doi.org/10.1037/a0029406>
- Miller, D. N., & Nickerson, A. B. (2006). Projective assessment and school psychology: Contemporary validity issues and implications for practice. *The California School Psychologist, 11*(1), 73–84. <https://doi.org/10.1007/BF03341117>
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24*(8), 1094–1102. <https://doi.org/10.1076/jcen.24.8.1094.8379>
- Mustac v. Medical Board of Western Australia. (2007). WASCA 156. <http://decisions.justice.wa.gov.au/supreme/supdcsn.nsf/judgment.xsp?documentId=35E998F6467030A248256EDF000D50B7&action=openDocument>
- National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. <https://www.ncme.org/publications/testing-standards>
- 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*(12), 1406–1421. <https://doi.org/10.1177/0093854814548449>
- 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>
- Nørbech, P. C. B., Fodstad, L., Kuisma, I., Lunde, K. B., & Hartmann, E. (2016). Incarcerated violent offenders' ability to avoid revealing their potential for violence on the Rorschach and the MMPI-2. *Journal of Personality Assessment*, 98(4), 419–429. <https://doi.org/10.1080/00223891.2015.1129613>
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7(6), 615–631. <https://doi.org/10.1177/1745691612459058>
- Nunnally, J. C., & Bernstein, I. C. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Otgaar, H., Howe, M. L., Patihis, L., Merckelbach, H., Lynn, S. J., Lilienfeld, S. O., & Loftus, E. F. (2019). The return of the repressed: The persistent and problematic claims of long-forgotten trauma. *Perspectives on Psychological Science*, 14(6), 1072–1095. <https://doi.org/10.1177/1745691619862306>
- Otten, M., Seth, A. K., & Pinto, Y. (2017). A social Bayesian brain: How social knowledge can shape visual perception. *Brain and Cognition*, 112, 69–77. <https://doi.org/10.1016/j.bandc.2016.05.002>
- Overton, M. N. (1984). *An examination of naïve and role informed faking in Rorschach performance* [Doctoral dissertation, University of Tennessee]. Dissertation Abstracts International, 45683B.
- Parker, K. (1983). A meta-analysis of the reliability and validity of the Rorschach. *Journal of Personality Assessment*, 47(3), 227–231. https://doi.org/10.1207/s15327752jpa4703_1
- Parker, K. C., Hanson, R. K., & Hunsley, J. (1988). MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability, and validity. *Psychological Bulletin*, 103(3), 367–373. <https://doi.org/10.1037/0033-2909.103.3.367>
- Paulson, K. L., Straus, E., Bull, D. M., MacArthur, S. K., DeLorme, J., & Dalenberg, C. J. (2019). Knowledge and views of psychological tests among psychiatrists and psychologists. *Journal of Forensic Psychology Research and Practice*, 19(2), 112–127. <https://doi.org/10.1080/24732850.2018.1546071>
- Perry, G. G., & Kinder, B. N. (1990). The susceptibility of the Rorschach to malingering: A critical review. *Journal of Personality Assessment*, 54(1–2), 47–57. https://doi.org/10.1207/s15327752jpa5401&2_5 <https://doi.org/10.1080/00223891.1990.9673972>
- Perry, W., McDougall, A., & Viglione, D. (1995). A five-year follow-up on the temporal stability of the Ego Impairment Index. *Journal of Personality Assessment*, 64(1), 112–118. https://doi.org/10.1207/s15327752jpa6401_7
- Piotrowski, Z. A. (1980). CPR: The psychological X-ray in mental disorders. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 85–108). Ablex.
- Popper, K. R. (1989). *Conjectures and refutations: The growth of scientific knowledge* (5th ed.). Harper & Row.
- Post, M. W. (2016). What to do with 'moderate' reliability and validity coefficients? *Archives of Physical Medicine and Rehabilitation*, 97(7), 1051–1052. <https://doi.org/10.1016/j.apmr.2016.04.001>
- Pownall & Ors v Conlan Management Pty Ltd as Trustee for the Kalbarri Trust. (1995). *WASC 117*. <https://www.austlii.edu.au/au/cases/wa/WASC/1995/117.pdf>
- President's Council of Advisors on Science and Technology. (2016). *Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods*. <http://scipol.duke.edu/content/forensic-science-criminal-courts-ensuring-scientific-validity-feature-comparison-methods>
- Rice, M. R., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law and Human Behavior*, 29(5), 615–620. <https://doi.org/10.1007/s10979-005-6832-7>
- Rorschach, H. (1921/1951). *Psychodiagnostics* (5th ed.). Hans Huber.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 1–7). John Wiley & Sons.
- Ruscio, J. (2000). The role of complex thought in clinical prediction: Social accountability and the need for cognition. *Journal of Consulting and Clinical Psychology*, 68(1), 145–154. <https://doi.org/10.1037/0022-006X.68.1.145>
- R v Turner. (1975). 1 All ER 70.
- Schott, G. D. (2014). Revisiting the Rorschach ink-blot: From iconography and psychology to neuroscience. *Journal of Neurology, Neurosurgery, and Psychiatry*, 85(6), 699–706. <https://doi.org/10.1136/jnnp-2013-305672>

- Schultz, D. S., & Brabender, V. M. (2013). More challenges since Wikipedia: The effects of exposure to Internet information about the Rorschach on selected Comprehensive System variables. *Journal of Personality Assessment*, 95(2), 149–158. <https://doi.org/10.1080/00223891.2012.725438>
- Schultz, D. S., & Loving, J. L. (2012). Challenges since Wikipedia: The availability of Rorschach information online and Internet users' reactions to online media coverage of the Rorschach-Wikipedia debate. *Journal of Personality Assessment*, 94(1), 73–81. <https://doi.org/10.1080/00223891.2011.627963>
- Seitz, F. C., Andersen, D. O., & Braucht, G. N. (1974). A comparative analysis of Rorschach signs of homosexuality. *Psychological Reports*, 35(3), 1163–1169. <https://doi.org/10.2466/pr0.1974.35.3.1163>
- Sewell, K. W., & Helle, A. C. (2018). Dissimulation on projective measures: An updated appraisal of a very old question. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (pp. 301–313). Guilford Press.
- Sherman, M. T., Seth, A. K., & Kanai, R. (2016). Predictions shape confidence in right inferior frontal gyrus. *Journal of Neuroscience*, 36(40), 10323–10336. <https://doi.org/10.1523/jneurosci.1092-16.2016>
- Siipola, E., & Taylor, V. (1952). Reactions to ink blots under free and pressure conditions. *Journal of Personality*, 21(1), 22–47. <https://doi.org/10.1111/j.1467-6494.1952.tb01857.x>
- Sjöstedt, G., & Grann, M. (2002). Risk assessment: What is being predicted by actuarial prediction instruments? *International Journal of Forensic Mental Health*, 1(2), 179–183. <https://doi.org/10.1080/14999013.2002.10471172>
- Stenius, J., Lindfors, O., Antikainen, R., Wahlström, J., Sares-Jäske, L., & Knekt, P. (2018). Associations between the Rorschach Ego Impairment Index and measures on intrapsychic and interpersonal functioning. *Open Journal of Psychiatry*, 8(1), 1–19. <https://doi.org/10.4236/ojpsych.2018.81001>
- Stewart, R. E., & Chambless, D. L. (2007). Does psychotherapy research inform treatment decisions in private practice? *Journal of Clinical Psychology*, 63(3), 267–281. <https://doi.org/10.1002/jclp.20347>
- Steyerberg, E. W. (2009). *Clinical prediction models: A practical approach to development, validation, and updating*. (Vol. 1). Springer.
- Sultan, S., Andronikof, A., Réveillère, C., & Lemmel, G. (2006). A Rorschach stability study in a nonpatient adult sample. *Journal of Personality Assessment*, 87(3), 330–348. https://doi.org/10.1207/s15327752jpa8703_13
- Sundberg, N. D. (1977). *Assessment of persons*. Prentice Hill.
- Szmukler, G., Everitt, B., & Leese, M. (2012). Editorial: Risk assessment and receiver operating characteristic curves. *Psychological Medicine*, 42(5), 895–898. <https://doi.org/10.1017/S003329171100208X>
- Tham, D. S. Y., Sowden, P. T., Grandison, A., Franklin, A., Lee, A. K. W., Ng, M., Park, J., Pang, W., & Zhao, J. (2020). A systematic investigation of conceptual color associations. *Journal of Experimental Psychology: General*, 149(7), 1311–1332. <https://doi.org/10.1037/xge0000703>
- United States v. Hall. (1996). 93 F.3d 1337, 1342-43, 7th Cir. <https://www.casemine.com/judgement/us/591482b4add7b0493449ab87#>
- Viglione, D. J., & Giromini, L. (2016). The effects of using the international versus Comprehensive System Rorschach norms for children, adolescents, and adults. *Journal of Personality Assessment*, 98(4), 391–397. <https://doi.org/10.1080/00223891.2015.1136313>
- Viljoen, J. L., McLachlan, K., & Vincent, G. M. (2010). Assessing violence risk and psychopathy in juvenile and adult offenders: A survey of clinical practices. *Assessment*, 17(3), 377–395. <https://doi.org/10.1177/1073191109359587>
- Weiner, I. B. (2003). *Principles of Rorschach interpretation*. Lawrence Erlbaum Associates.
- Weiner, I. B. (2014). *Principles of Rorschach interpretation* (2nd ed.). Routledge.
- Weiner, I. B., Exner, J. E., & Sciarra, A. (1996). Is the Rorschach welcome in the courtroom? *Journal of Personality Assessment*, 67(2), 422–424. https://doi.org/10.1207/s15327752jpa6702_15
- Weiner, I. B., & Greene, R. L. (2007). *Handbook of personality assessment*. John Wiley & Sons.
- What's behind the Rorschach inkblot test? (2012, July 25). *BBC News*. <https://www.bbc.com/news/magazine-18952667>
- Wikipedia. (2021, April 16). Rorschach test. Retrieved January 28, 2021. https://en.wikipedia.org/wiki/Rorschach_test
- Wilson, C. (2020, February). Controversial psychology tests are often still used in US courts. *New Scientist*, 3270. <https://www.newscientist.com/article/2233956->

- controversial-psychology-tests-are-often-still-used-in-us-courts/
- Wood, J. M., Garb, H. N., Nezworski, M. T., Lilienfeld, S. O., & Duke, M. C. (2015). A second look at the validity of widely used Rorschach indices: Comment on Mihura, Meyer, Dumitrascu, and Bombel (2013). *Psychological Bulletin*, *141*(1), 236–249. <https://doi.org/10.1037/a0036005>
- Wood, J. M., Lilienfeld, S. O., Garb, H. N., & Nezworski, M. T. (2000). The Rorschach test in clinical diagnosis: A critical review, with a backward look at Garfield (1947). *Journal of Clinical Psychology*, *56*(3), 395–430. [https://doi.org/10.1002/\(SICI\)1097-4679\(200003\)56:3<395::AID-JCLP15>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-4679(200003)56:3<395::AID-JCLP15>3.0.CO;2-O)
- Wood, J. M., Nezworski, M. T., Lilienfeld, S. O., & Garb, H. N. (2008). The Rorschach inkblot test, fortune tellers, and cold reading. In S. O. Lilienfeld, J. Ruscio, & S. J. Lynn (Eds.), *Navigating the mindfield: A user's guide to distinguishing science from pseudoscience in mental health* (pp. 145–157). Prometheus Books/Rowman & Littlefield.
- Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science*, *7*(1), 3–10. <https://doi.org/10.1111/j.1467-9280.1996.tb00658.x>
- Wood, J. M., Nezworski, M. T., Stejskal, W. J., & McKinzey, R. K. (2001). Problems of the Comprehensive System for the Rorschach in forensic settings: Recent developments. *Journal of Forensic Psychology Practice*, *1*(3), 93–103. https://doi.org/10.1300/J158v01n03_06