Neal, T.M.S. & Grisso, T. (2014). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An International Snapshot. *Criminal Justice and Behavior, 41*, 1406-1421. doi:10.1177/00093854814548449

Assessment Practices and Expert Judgment Methods in Forensic Psychology and Psychiatry: An International Snapshot

Tess M.S. Neal & Thomas Grisso

University of Massachusetts Medical School

Author Note

Tess M.S. Neal and Thomas Grisso, Department of Psychiatry, University of Massachusetts Medical School.

Tess M.S. Neal is now at the University of Nebraska Public Policy Center.

The first author was supported in part by an NSF Interdisciplinary Postdoctoral Fellowship (SES1228559) during the writing of this manuscript. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect those of NSF. Portions of these results were presented at the 2014 annual conference of the American Psychology-Law Society (AP-LS) in New Orleans, LA.

Correspondence concerning this article should be addressed to Tess M.S. Neal, University of Nebraska Public Policy Center, 215 Centennial Mall South, Suite 401 (P.O. Box 880228), Lincoln NE 68588. E-mail: tess.neal@asu.edu

Abstract

We conducted an international survey in which forensic examiners who were members of professional associations described their two most recent forensic evaluations (*N*=434 experts, 868 cases), focusing on the use of structured assessment tools to aid expert judgment. This study describes: (a) the relative frequency of various forensic referrals, (b) what tools are used globally, (c) frequency and type of structured tools used, and (d) practitioners' rationales for using/not using tools. We provide general descriptive information for various referrals. We found most evaluations used tools (74.2%) and used several (on average 4). We noted the extreme variety in tools used (286 different tools). We discuss the implications of these findings and provide suggestions for improving the reliability and validity of forensic expert judgment methods. We conclude with a call for an assessment approach that seeks structured decision methods to advance greater efficiency in the use and integration of case-relevant information.

Keywords: judgment; decision; forensic; structure; actuarial

Assessment Practices and Expert Judgment Methods in Forensic Psychology and Psychiatry: An International Snapshot

Forensic psychologists and psychiatrists are expected to be experts in their subject areas and to make good use of the cumulative knowledge developed in their fields over time. How might experts use the body of knowledge in their fields to minimize decision errors? Systematic approaches have been developed to help experts harness field-based knowledge to remember everything one needs to know or do for a given task. The field of forensic mental health assessment has developed many structured assessment tools to aid forensic clinicians in making decisions related to forensic referral questions. Many of these tools are actuarial (i.e., mechanical, formula-based), while others are checklist-based methods frequently referred to as Structured Professional Judgment (SPJ) tools. In the SPJ approach, the expert is presented with *evidence-based factors* to consider with specific guidelines (Guy, Packer, & Warnken, 2012). This approach does not rely on fixed decision rules as there is no algorithm to combine the data to arrive at a decision, so this approach operates somewhere between actuarial and unaided clinical judgment methods (Douglas, Ogloff, Nicholls, & Grant, 1999).

The development of structured tools in the forensic mental health field has not been without controversy. Some argue that an unstructured intuitive approach can lead to better decisions at times, or that clinical judgment is more flexible and can take into account novel or powerful information that might not be included in existing formulas or checklists (e.g., Montgomery, 2005; Litwack, 2001; Skeem et al., 2005). However, the weight of evidence indicates that the structured approaches perform better than unaided clinical judgment when sound tools are available to assist decision tasks (e.g., Ægisdóttir et al., 2006; Dawes, Faust, & Meehl, 1989; Dolan & Doyle, 2000; Faust & Ziskin, 1988; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Guy, 2008; Haynes et al., 2009).

The Current Study

Despite the development of many structured tools to assist professional judgment in the past few decades, little is known about the degree to which these tools have become standard practice in the forensic mental health field. Neither is there much information available about the conditions under which they are used and with what perceived strengths and weaknesses. Our study explored forensic mental health professionals' self-reported use of structured tools in their forensic evaluations in civil and criminal contexts. We also wanted to know when forensic mental health professionals see the use of these tools as more or less justified.

Previous surveys of forensic mental health professionals have typically asked what clinical diagnostic tools are used in various kinds of forensic evaluations (such as multi-scale symptom inventories, clinical scales, cognitive and achievement tests, unstructured personality tests, and neuropsychological tests: Archer, Buffington-Vollum, Stredny, & Handel, 2006; Boccaccini & Brodsky, 1999; Keilin & Bloom, 1986; Lees-Haley, Smith, Williams, & Dunn, 1996; McLaughlin & Kan, 2014). Typically they have asked respondents to express how frequently they use such tools in their forensic evaluations (e.g., "never, sometimes, almost always," or percentage of time). In contrast, in this study we asked forensic clinicians to describe their use of tools in their two most recent forensic cases. Our intent was to obtain an estimate based on "sampling" of cases rather than relying on respondents to characterize the frequency of their use of tools. Moreover, this method allowed us to sample from the full range of forensic evaluations that forensic clinicians perform, whereas previous surveys typically asked about tools used in one or two particular kinds of forensic evaluations (and usually, by American psychologists).

None of the earlier studies inquired about the practicalities of using these instruments or the reasons that clinicians might not use them. In addition, the studies typically inquired about only one or

two kinds of forensic evaluations. It appears that only one study to date has examined the practicalities of routinely using structured tools in forensic assessments. Focusing on competence to stand trial evaluations, Pinals, Tillbrook, and Mumley's (2006) qualitative study suggested that there may be several reasons why structured tools might not be adopted in routine practice by forensic evaluators. The present study sought to address the potential gap between research and practice by exploring the degree to which forensic evaluators use tools to aid their clinical judgment as well as exploring reasons why they might not.

Method

Procedure and Materials

After obtaining institutional review board approval, we designed our survey online using REDCap software.[1] Professionals (described below) received an email inviting them to participate in the survey and were sent a reminder invitation after two weeks. In the survey, we asked participants to answer questions about the two most recent forensic evaluations they had completed. We defined forensic mental health evaluations as "*a psychological or psychiatric assessment of a person involved in a legal proceeding, conducted by the mental health professional in service to the legal system. Some examples include evaluations of civil and criminal competencies, criminal responsibility, mental disability, child custody and protection, violence and sexual offending risk assessments, and psychic injury, among others."* We requested that participants retrieve their reports (i.e., pull the hard-copy from their file cabinet or open an electronic version of the report) and refer to them as they answered the survey questions. We estimated that the survey required about 15 minutes.

[1] REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data collection for research studies, providing: 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources (Harris, Taylor, Thielke, Payne, Gonzales, & Conde, 2009).

Our questions inquired about the referral question, sources of information used, whether or not any standardized tools were used (which we defined as "*any tests, instruments, checklists, or rating systems*"), what tools were used if applicable, reasons tools were used (or not), length of the report (in pages), how long the evaluation took from the time of referral until completion (in days), and demographic questions about the evaluator. Responses were provided in menus when possible, usually with an "other" category that allowed for typed responses.

Participants

Psychologist and psychiatrist members of professional forensic mental health associations in the U.S.,[2] Canada,[3] Australia and New Zealand,[4] and Europe[5] were invited to complete the online survey. There were 434 respondents, reporting on 868 cases.[6] Most of the sample comprised doctoral-level clinicians (91%) and master's-level clinicians (7.4%). Regarding profession, more psychologists (51%) than psychiatrists (6%) responded.[7] This was an experienced sample, with an average of 16.56 years (*SD* = 12.01 years) of forensic evaluation experience. Overall, 16.4% of the sample was board-certified. Certifying boards included the American Board of Forensic Psychology (6.7%) and other specialties of

[2] American Psychology-Law Society (American Psychological Association Division 41), Psychologists in Independent Practice – Criminal Justice Section (American Psychological Association Division 18), American Board of Forensic Psychology, American Academy of Psychiatry and the Law

[3] Canadian Psychological Association – Criminal Justice Section, Canadian Academy of Psychiatry and the Law

[4] Australian and New Zealand Association of Psychiatry, Psychology and Law; Australian Psychological Society; Royal Australian and New Zealand College of Psychiatrists

[5] European Association of Psychology and Law, European Psychiatric Association-Forensic Section, Swiss Forensic Psychiatric Association

[6] We were unable to calculate the response rate given the fluid nature of some of the participant sources. For instance, some of the organizations sent our survey invitation to organized lists of members, whereas others posted to organizational listservs, and still others posted in organizational newsletters. When we attempted to track down how many people would have received the invitation from each organization, several of the organizations told us they were unsure how many people received their posts/mailings. Furthermore, there is likely some overlap between these organizations (e.g., members of the American Psychology-Law Society might also be members of APA Division 18 and/or be part of ABFP).

[7] We did not require that the demographic questions be answered, whereas we did require answers for all the other questions on the survey. Because we did not require answers to the demographic questions, we are missing demographic data for several participants in this section. For example, regarding profession, 43% of respondents failed to indicate whether they were a psychologist or psychiatrist. And regarding where they practice, 41.5% failed to report in what country they work. Thus, the data reported in the participants section describes only those people who answered the demographic questions.

the American Board of Professional Psychology (4.8%), Royal College of Physicians and Surgeons (2.8%), and other boards (e.g., American Board of Psychiatry and Neurology, American Board of Sleep Medicine). Most of the participants reported they practiced in the U.S. (44.7%), followed by Canada (6.9%), Australia and New Zealand (4.2%), and Europe (2.8%). Within the United States, 39 states and Washington D.C. were represented in the sample.

Results

Figure 1 graphs the relative percentages of various referral questions in the overall sample of 868 cases. The most common referral question was Competence (Fitness) to Stand Trial (CST); followed by criminal risk, insanity, and sentencing aid assessments, then various types of evaluations in civil proceedings (adult and child; see Figure 1). We report descriptive statistics for each of the referral questions for which there were at least 25 reports, which corresponds to the "top ten" most common forensic referrals in this sample. Other kinds of evaluations that are not further detailed in this paper include False Confession Assessments, Immigration, Asylum, Board of Psychology Complaints, Fitness for Practice, and unspecified.

[Insert Figure 1 about here]

For each of the "top ten" types of forensic referrals identified in Figure 1, Table 1 shows the percent using structured tools, average number used per case, number of different types of tools, how long the evaluations took from referral to completion (in days), and report length (in pages). The referrals that were finished in the shortest amount of time were Workplace/Employment Disability evaluations, which took on average about 18 days. Child Custody evaluations took the longest, on average about 44 days. All the other referrals took between 25 and 36 days, with Violence Risk Assessments falling at the high end of that range at 36 days on average. The shortest reports were CST evaluations with an average of 13 pages. The longest were Child Custody evaluations at 32 pages. The

other referrals were between 15 and 24 pages on average, with Violence Risk Assessments in the middle with an average of 19 pages.

[Insert Table 1 about here]

With regard to use of structured tools, we found that most forensic mental health evaluations (74.2%) used one or more tools to aid clinical judgment. Of the evaluations that used at least one tool, most used several (4 on average, ranging up to 18). We noted the extreme number and variety of tools used; very few of the evaluations used the same sets of tools, and the sheer variety—286 tools—was surprising. Even within specific referral questions, there were many different tools used. Forensic assessments that were most likely to use tools were criminal risk assessments and child protection evaluations (89% or greater), while CST evaluations were the least likely to use tools (58.4%).

Other Sources of Information Used

In addition to use of structured tools, we asked what other sources of information forensic clinicians used for these various referrals (see Table 2). Almost every evaluation in every referral type relied on examinee interviews (at least 99.0% for each). Mental health or medical records were commonly used, with Insanity evaluations using them the most (97.4%) and Child Custody evaluations the least (72.1%). Use of justice system records, such as police reports, criminal record, depositions, witness statements, and justice facility disciplinary records, varied more widely in expected ways. For instance, more than 95% of the Violence Risk, Sex Offender Risk, Insanity, and Aid-in-Sentence evaluations used justice records, whereas only 17.2% of Disability evaluations did so.

[Insert Table 2 about here]

Use of collateral interviews (both professional and non-professional collaterals) was most common in Child Custody evaluations and least common in Disability evaluations. Non-professional collateral interview sources included relatives, acquaintances, and witnesses among others. Professional

collateral sources included attorneys, treating and educational professionals, and work supervisors. Child Custody evaluations were also most likely to include visiting the relevant environment (32.6%) and use of other records (11.6%), such as reviewing Facebook and text messages, telephone records, financial records, e-mail and written correspondence, photographs, and audio and visual recordings. Educational records were most frequently used for Aid-in-Sentence evaluations (40.3%) and least frequently used in Disability evaluations. Biological tests, such as medical examinations, blood tests, imaging techniques, polygraph examinations, and penile plethysmographs, were most commonly used for Insanity evaluations (26.9%) and least commonly used in Child Custody evaluations (11.6%). Work or employment records were common in Disability evaluations (24%) but rare for all the other referrals.

Specific Tools Used

Table 3 provides information about the ten most frequently used tools per referral type. Overall, personality assessment instruments were the most common kind of tool used, with any version of the *Minnesota Multiphasic Personality Inventory* (MMPI; Butcher, Graham, Ben-Porath, Tellegen, & Dahlstrom, 2003) used in 15.2% of the evaluations and the *Personality Assessment Inventory* (PAI; Morey & Boggs, 1991) in 9.6% of the evaluations. These two tools were in the "top ten" list for every one of the ten most common referral types. The other "top ten" tools varied by referral question.

[Insert Table 3 about here]

Several tools that were designed for specific kinds of referrals appeared most often or exclusively for those referrals. For instance, the Historical Clinical Risk Management 20 (HCR-20), a violence risk assessment tool (Webster & Eaves, 1995) was tied for the number-one tool for violence risk assessments (35.6%), but did not appear in the top ten list for any of the other referrals. Another violence risk assessment tool, the *Violence Risk Appraisal Guide* (VRAG; Quinsey, Harris, Rice, &

Cormier, 1998) was used in 17.8% of the violence risk assessments and otherwise only appeared in the Sex Offender Risk Assessment list.

Five of the 10 most common tools used in Sex Offender Risk Assessments were designed specifically for these kinds of evaluations (e.g., the *Static-99-R* and related versions, Hanson & Thornton, 1999; *Sexual Violence Risk-20* [SVR-20; Boer, Hart, Kropp, & Webster, 1997], *Risk for Sexual Violence Protocol* [RSVP; Hart, Kropp, Laws, Klaver, Logan, & Watt, 2003], Stable 2007 [Harris, Scott, & Helmus, 2007], and *Sex Offender Risk Appraisal Guide* [SORAG; Quinsey et al., 1998]). Tools designed to help assess trial competence related abilities were 3 of the 10 most frequently used tools for CST evaluations, including the *Evaluation of Competence to Stand Trial – Revised* (ECST-R; Rogers, Tillbrook, & Sewell, 2004), *MacArthur Competence Assessment Tool-Criminal Adjudication* (MacCAT-CA; Hoge, Bonnie, Poythress, & Monahan, 1999) and *Juveniles' Adjudicative Competence Interview* (JACI; Grisso, 2005).

The *Psychopathy Checklist-Revised* (PCL-R; Hare, 2003) was the other tool tied for the number-one most frequently used tool in Violence Risk Assessments (35.6%) and was also common in Sex Offender Risk Assessments, Civil Commitments, Aid-in-Sentence, and Insanity evaluations. Response style or malingering tools comprised 5 of the 10 most common tools used in Insanity evaluations and 2 of the 10 for both CST and Disability evaluations. For instance, the *Test of Memory Malingering* (TOMM; Tombaugh, 1996) was used in about 10-15% of Insanity, Disability, and CST evaluations.

Symptom inventories were specific to Disability and Civil Tort cases. Seven of the 10 most common tools used in Disability evaluations were symptom inventories, as were 4 of the Civil Tort tools. Measures of trauma symptoms, such as the *Trauma Symptom Inventory* (TSI; Briere, 1996) and *Posttraumatic Stress Disorder Checklist-Military Version* (PCL-M; Weathers, Huska, & Keane, 1991)

among others, were used in about one-quarter of Disability and Civil Tort evaluations. Various depression, anxiety, and pain inventories were also common in Disability and Civil Tort evaluations.

Forensic Clinicians' Reasons for Using or Not Using Structured Tools

Respondents were asked why they used (or did not use) tools in the cases in question (see Figure 2). They were provided with a list of options (including "other") and asked to select any of the factors that affected their choice. Overall, the most common reason for using structured tools was "to use an evidence-based method," followed closely by "to improve the credibility of my assessment" and "to standardize the assessment." Less-common reasons included "I was required to (by policy, by supervisor, etc.)," "to learn or test out a new instrument that I had not tried before," and "other." "Other" reasons included the desire for diagnostic clarification, for efficiency, to rule out malingering, obtain information, generate or confirm hypotheses, improve accuracy, charge more money, guide recommendations, meet ethical standards, and integrate diverse data.

[Insert Figure 2 about here]

As we expected, the most common reason that forensic clinicians did not use any structured tools in the cases in question was, "I trusted my clinical judgment in completing the evaluation." Other reasons included, "the time the tools(s) would have added to the evaluation was not justified," "there were no tool(s) available for the task I was asked to complete," "the tool(s) have a lot of limitations," "the cost the tool(s) would have added to the evaluation was not justified," "I am not familiar with the relevant tool(s)," and "other." "Other" included having enough data without using tools, no tools were validated for the language or culture of the evaluee, policy restrictions against using tools, testing was recently completed, evaluee was fatigued, too psychotic, or refused, cognitive limitations precluded use of tool, and tools were not flexible enough.

Discussion

This study adds new information to the knowledge base of forensic practice, including the relative frequency of various referrals, frequency and type of structured tools used in routine practice by forensic clinicians internationally, and practitioners' rationales for using or not using tools. Moreover, the sampling method that we used provides a different perspective than previous surveys that have asked clinicians to estimate the frequency of their use of tools for one or two specific types of forensic referral questions. This information is important for understanding big-picture trends in the diverse field of forensic mental health assessment.

As far as we know, this is the first study to document the relative frequency of various referral types completed by members of professional forensic organizations, as depicted in Figure 1. We were not surprised to see that CST evaluations were the most frequent referral, as this is consistent with what other sources have suggested (e.g., Golding, 1992; Zapf & Roesch, 2009). However, we were not sure what to expect in terms of the relative frequency of other kinds of referrals. The data show that risk assessments (both violence and sex offender risk) were the second and third most frequent type of referral, respectively – a finding we did not necessarily anticipate. We also found that several kinds of referrals about which there is substantial literature are actually infrequent (i.e., less than 1% of the sample), including Competence to Consent to Treatment, Juvenile Transfer to/from Adult Court, Fitness for Duty, and Capacity to Waive Miranda Rights evaluations.[8]

When interpreting the results, it is important to keep in mind that the data reflect our sample, which is comprised of members of forensic organizations. Many clinicians who do "forensic" work may not identify as "forensic clinicians" or may not be members of forensic associations. And those

[8] The data in Figure 1 might not fully represent the population of all forensic referrals. For example, given our sampling of forensic organization members, it may be expected that Competence to Consent to Treatment evaluations would be uncommon. These kinds of referrals may typically be completed by clinical psychologists or psychiatrists in hospitals who do not do other types of forensic evaluations and would not belong to any forensic organizations. Thus, these evaluations may be more common than it appears based on Figure 1.

who are members of these organizations may be more likely to follow professional forensic standards of practice. Therefore, these results should not be interpreted to represent general forensic practice, but rather practice by forensic specialists who identify with organizational standards.

The issue of representativeness within our sample merits attention as well. We do not know the rate of responding within each of the groups we sampled or how representative respondents were of their respective groups. This is a common limitation of online surveys. Nevertheless, our data are among the first to provide information about this topic and therefore contribute meaningfully to the literature in spite of the limitations of our online survey method. Studies examining this topic in the future should rely on other methods to balance out the limitations of online studies.

One more limitation deserves discussion. The size of some of our cells or groupings is relatively small for the questions addressed (e.g., under 50 and as low as 25). Group sizes in this range can produce misleading information about such matters as relative frequency of test use, especially given the fact that we cannot be sure about how representative the small samples are of their larger populations. These limitations must be kept in mind when interpreting our results and discussion.

Use of Tools

The fact that about three-fourths of the evaluations sampled employed tools to aid professional judgment is encouraging, given that structured tools improve clinical decision making (i.e., reduce bias, increase inter-rater reliability, and ostensibly increase validity). Moreover, most of the "top-ten" tools in each type of evaluation have a sufficient body of research on their validity to support their use. Other surveys have asked clinicians what tools are typically used in various kinds of forensic evaluations (e.g., Archer et al., 2006; Boccaccini & Brodsky, 1999; Keilin & Bloom, 1986; Lees-Haley et al., 1996; McLaughlin & Kan, 2014). The present results show that when we sample actual cases, we find a different picture of the use of tools in forensic work than what has been reported in these

previous studies. Rather than simply data about favorite tools, we find the involvement of literally hundreds of different tools in forensic evaluations.

Specifically, few referrals – even of the same type – relied on the same set of structured tools or information sources. Given the high number of evaluations using tools, and the fact that most used several tools, this means that dozens or even hundreds of different tools are used for the same referral question.[9] For example, 110 different structured tools were used in the 101 violence risk assessment evaluations in this sample.

Nevertheless, there is some agreement about tools that are appropriate for given referral questions based on the frequency with which particular tools were used for particular referrals (see Table 3). These data indicate there may be some consensus about the use of certain tools for certain referrals. For instance, more than one-third of all violence risk assessments relied on the HCR-20 (Webster & Eaves, 1995) and the PCL-R (Hare, 2003). Four other tools were used in more than 10% of all of the violence risk assessments: the VRAG (Quinsey et al., 1998), MMPI (Butcher et al., 2003), PAI (Morey & Boggs, 1991), and LS/CMI (Andrews, Bonta, & Wormith, 2004).[10]

Heilbrun, Rogers, & Otto's (2002) three-category typology is useful for sorting out the tool-use picture created by the data in this report. The first class is Forensic Assessment Instruments (FAIs), a classification offered by Grisso (2003) to identify tools specifically designed to assess more or less directly the abilities or propensities associated with the legal question, such as instruments designed to measure abilities associated with competence to stand trial. A second class of tool, which Heilbrun et al.

[9] We do not mean to imply that each test is intended to answer the referral question itself. A clinician can use the Beck Depression Inventory in a CST case, for example, not to answer the question of competence, but as a piece of data to test a hypothesis about the presence of depression as a possible reason for apparent functional competency deficits.

[10] We also do not mean to imply that tools are necessarily interchangeable. For instance, among the CST tools, the ECST-R and the MacCAT-CA are dramatically different in their content and styles, and clinicians may select to use one or the other (or a different tool or no tool) given the circumstances of a given case. Likewise, although the PCL-R and the LS/CMI are both commonly-used tools in violence risk assessments, they are entirely different in their concepts and purposes.

call Forensically Relevant Instruments (FRIs), measures clinical constructs that are sometimes pertinent to psycholegal concepts. For example, measures of psychopathy (e.g., the PCL-R; Hare, 2003) and malingering (e.g., the TOMM; Tombaugh, 1996) often map onto clinical questions of direct relevance to the court. The third class of tool identified by Heilbrun et al. is the Clinical Assessment Instrument (CAI), referring to standard psychological tests developed for use in diagnosis, symptom description, and intervention planning with clinical populations. While they are an inferential step further than FRIs (and two steps further than FAIs) from the forensic issue, they can be quite valuable in explaining clinical conditions underlying forensic cases. Examples of CAIs include the Wechsler Adult Intelligence Scale (WAIS-IV; Wechsler, 2008) and personality tests such as the MMPI (Butcher et al., 2003) and PAI (Morey & Boggs, 1991).

Many forensic referrals require information from various categories of tools. For instance, in a CST evaluation a person may do poorly on a FAI that assesses competency abilities (e.g., the MacCAT-CA), yet more information is needed to complete the evaluation. The evaluator will need to explain those deficits. Are they due to limited cognitive capacity? To thought disturbance associated with psychosis? Or possibly malingering? Many types of structured assessment tools other than FAIs are likely to be used to test various competing hypotheses. Similarly, an FAI risk tool may do a fine job of identifying risk, but the court may want to know more about how to manage the risk or what treatment the evaluee might require – questions that might require CAI or FRI supplements as part of the assessment process.

Is Our Diversity of Tools Beneficial or Problematic?

Our finding regarding the extraordinarily great number of tools that are used in this sample raises questions about the implications of this diversity. Is there a danger in too much diversity? And if so, does it threaten the value of it? Diversity has its advantages. If, on the one hand, examiners are

being selective in their use of those tools, choosing those that best test their hypotheses in a specific case, this could explain the wide range of tests. This could be seen as consistent with the training forensic clinicians typically receive with regard to relying on clinical judgment to select tools that best fit the given case rather than using a standard battery or tool in a rote fashion.

On the other hand, this subjective decision making process about how to approach each and every case, which allows for extreme flexibility, could be construed as a liability within the forensic mental health field. The lack of any standardized approach to various referral questions might contribute to lower inter-rater rates of agreement in forensic opinions as major variation in methods may substantially influence the outcome of evaluations. Moreover, attorneys and courts who use our information are required to become familiar with a bewilderingly wide range of tools, rather than becoming consumer-wise regarding a more limited number. The question these observations raise is whether the "typical training" of forensic evaluators to use a flexible approach to each case is the best approach to training.

Perhaps there is a compromise between these two positions. The compromise begins by noting that many of the tools examiners reported using across various referral questions have serious psychometric limitations. It is outside the scope of this paper to classify these 286 tools into those that have good empirical foundation and those that do not (see e.g., Grisso, 2003 and Heilbrun et al., 2002 for more information about sound psychometric tools within the forensic field). But when tools with good psychometric properties and sound empirical foundations are available for a given referral question (or for sub-questions relevant to the overall referral), these sound tools should be used rather than tools with questionable validity and reliability. The compromise, then, is for our training (a) to continue to emphasize the importance of selecting tools that are appropriate for the specific case, but (b) to move examiners to favor those tools that have the best psychometric properties.

Heilbrun et al. (2002) developed a useful "checklist" to help examiners decide whether a given tools would be appropriate for use in a forensic evaluation. Their recommendations included that the tool (a) be commercially published and distributed, (b) have an available test manual, (c) have demonstrated and adequate levels of reliability and validity for the purpose for which it will be used, (d) have undergone successful peer review, and (e) have known decision making formulas. We recommend forensic evaluators make use of this checklist in deciding what tools to use.

How Much Data is Too Much?

Consistent with best practice guidelines (e.g., Packer & Grisso, 2011), our data show that forensic mental health evaluators are using a variety of sources of information to answer the courts' referral questions. Almost every evaluation relied on an interview with the examinee and most relied on mental health or medical records as well. Use of other sources of information varied by referral question, and the pattern of variation makes sense. For instance, whereas justice system records are commonly relied on for cases in the criminal justice system, civil cases are less likely to use justice system records. Employment records were much more common for Disability evaluations than any of the other kinds of forensic referral questions. And our data indicate that, at least in this sample of forensic specialists, most cases involve the use of quite a number of assessment tools—on average, 3 to 5 and, at the first standard deviation, 5 to 7 per case. But is all this testing and data collection necessary or wise, based on what we know about optimal decision-making?

This approach to broad information gathering is advantageous in some ways. Forensic examiners are guided to obtain data from multiple sources (e.g., tests and collateral sources) and to be able to cross-check psychometric results across multiple tools. This is believed to reduce method-based error. But one can argue that sometimes this may lead to excessive data-gathering. Decision science suggests that identifying and then relying on only about four to six variables essential to the issue at

question is the optimal approach to valid and reliable decision making (e.g., Kahneman, 2011). These four to six data points should overlap as little as possible so as to provide unique data, and they should be highly reliable and valid indicators (e.g., Faust & Ahern, 2012; Gawande, 2009). Regardless of how much information might be potentially relevant to a particular decision task, human brains can only analyze the patterns and interrelationships among approximately five variables when it comes time to integrate all those data into a decision, conclusion, or forensic opinion (Faust & Ahern, 2012; Simon, 1956).

Various tools and pieces of information can be integrated to form one of the overarching four-to-six variables important to the referral. An intelligence test, achievement test, adaptive behavior test, and school records, for example, might be combined to represent "cognitive capacity." Cognitive capacity might then be one of the four-to-six pieces of information to consider when forming one's conclusion or opinion. Thus, we are not arguing that forensic evaluators only *gather* four to six facts. Rather, we suggest that forensic evaluators should consider what information is most essential to the referral question and then use due diligence to gather and integrate that relevant information into their conclusions and opinions.

Considering what the most essential four-to-six variables might be, and how they might be best assessed for any given referral question, is a topic for future discussion. In the meantime, the lack of guidance for forensic evaluators leads to subjective decisions in every case about what these variables should be and how they should be indexed. This situation likely contributes to lower rates of inter-rater reliability in forensic mental health assessments.

Conclusions

With regard to both of the issues we raise (diversity and quantity), our results suggest that the field of forensic mental health assessment has evolved to a level that raises different questions than

those that faced the field several decades ago. Our agenda in past decades was to develop methods to move us forward from an approach almost entirely dependent on clinical judgment unaided by structured tools. Now we have many such tools as well as standards that have made diverse sources of data commonplace in our forensic evaluations. We conclude that the field has evolved sufficiently to begin focusing on guidance for forensic clinicians in using our methods *with optimal efficiency*. This approach would lead us to seek the most effective use of our tools and methods. This would include favoring those that are most psychometrically sound, reducing the amount of testing when it is not essential to ruling out case-relevant hypotheses, choosing only those that add incremental validity to answer the question at hand, and developing structured decision methods that may achieve best results while reducing the actual amount of data required to obtain them.

References

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., ... & Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341-382. doi: 10.1177/0011000005285875

Andrews, D. A., Bonta, J., & Wormith, S. J. (2004). *The Level of Service/ Case Management Inventory (LS/CMI).* Toronto, Ontario, Canada: Multi-Health Systems

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of personality assessment, 87*(1), 84-94.

Boccaccini, M. T., & Brodsky, S. L. (1999). Diagnostic test usage by forensic psychologists in emotional injury cases. *Professional Psychology: Research and Practice, 30*, 253-259.

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk-20. Professional guidelines for assessing risk of sexual violence.* Vancouver, British Columbia: Institute Against Family Violence.

Briere, J. (1996). *Trauma symptom inventory.* Odessa, FL: Psychological Assessment Resources.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., & Dahlstrom, W. G. (2003). *MMPI-2: Minnesota Multiphasic Personality Inventory-2*. University of Minnesota Press.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668-1674.

Dolan, M., & Doyle, M. (2000). Violence risk prediction Clinical and actuarial measures and the role of the Psychopathy Checklist. *The British Journal of Psychiatry, 177*, 303-311. doi:10.1192/bjp.177.4.303

Douglas, K.S., Ogloff, R.P., Nicholls, T.L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening Version. *Journal of Consulting and Clinical Psychology, 67*, 917-930.

Faust, D. & Ahern, D.C. (2012). Clinical judgment and prediction. In D. Faust, *Coping with Psychiatric and Psychological Testimony, Sixth Edition,* 147-208. New York: Oxford University Press.

Faust, D., & Ziskin, J. (1988). The expert witness in psychology and psychiatry. *Science, 241*, 31-35.

Gawande, A. (2009). *The checklist manifesto: How to get things right.* New York: Picador.

Golding, S.L. (1992). Studies of incompetent defendants: Research and social policy implications. *Forensic Reports, 5*, 77-83.

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

Grisso, T. (2005). *Evaluating Juveniles' Adjudicative Competence: A Guide for Clinical Practice.* Sarasota, FL: Professional Resource Press.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: a meta-analysis. *Psychological assessment, 12*, 19. doi: 10.1037/1040-3590.12.1.19

Guy, L.S. (2008). *Performance indicators of the structured professional judgment approach for assessing risk for violence to others: A meta-analytic survey.* Unpublished Doctoral Dissertation, Simon Frasier University, Vancouver, CAN.

Guy, L.S., Packer, I.K., & Warnken, W. (2012): Assessing Risk of Violence Using Structured Professional Judgment Guidelines, *Journal of Forensic Psychology Practice, 12*:3, 270-283

Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders* (User Report No.1999–02). Ottawa: Department of the Solicitor General of Canada.

Hare, R. D. (2003). *The Hare Psychopathy Checklist-Revised technical manual (2nd ed.).* Toronto: Multi-Health Systems.

Harris, A. J., Scott, T., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project (Vol. 5).* Ottawa: Public Safety Canada.

Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J.G. (2009). Research electronic data capture (REDCap): A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform, 42*, 377-81.

Hart, S. D., Kropp, P. R., Laws, D. R., Klaver, J., Logan, C., & Watt, K. A. (2003). *The Risk for Sexual Violence Protocol (RSVP).* Mental Health, Law and Policy Institute of Simon Fraser University, Pacific Psychological Assessment Corporation & British Columbia Institute Against Family Violence.

Haynes, A. B., Weiser, T. G., Berry, W. R., Lipsitz, S. R., Breizat, A. H. S., Dellinger, E. P., ... & Gawande, A. A. (2009). A surgical safety checklist to reduce morbidity and mortality in a global population. New England Journal of Medicine, 360, 491-499. doi: 10.1056/NEJMsa0810119

Heilbrun, K., Rogers, R., & Otto, R. K. (2002). Forensic assessment. In *Taking psychology and law into the twenty-first century* (pp. 119-146). New York, NY: Springer.

Hoge, S.K., Bonnie, R.J., Poythress, N., & Monahan, J. (1999). *The MacArthur Competence Assessment Tool-Criminal Adjudication (MacCAT-CA).* Odessa, FL: Psychological Assessment Resources.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Keilin, W. G., & Bloom, L. J. (1986). Child custody evaluation practices: A survey of experienced professionals. *Professional Psychology: Research and Practice, 17*, 338–346.

Lees-Haley, P.R., Smith, H.H., Williams, C.W., & Dunn, J.T. (1996). Forensic neuropsychological test usage: An empirical survey. *Archives of Clinical Neuropsychology, 11,* 45-51.

Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law, 7*, 409. doi:10.1037/1076-8971.7.2.409

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

Montgomery, K. (2005). *How doctors think: Clinical judgment and the practice of medicine*. New York, NY: Oxford University Press.

Morey, L. C., & Boggs, C. (1991). *Personality assessment inventory (PAI).* PAR. Lutz, FL: Psychological Assessment Resources

Packer, I.K. & Grisso, T.G. (2011). *Specialty Competencies in Forensic Psychology.* New York, NY: Oxford University Press.

Pinals, D.A., Tillbrook, C.E., & Mumley, D.L. (2006). Practical application of the MacArthur Competence Assessment Tool-Criminal Adjudication (MacCAT-CA) in a public sector forensic setting. *Journal of the American Academy of Psychiatry and the Law, 34*, 179-188.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk.* American Psychological Association.

Rogers, R., Tillbrook, C. E., & Sewell, K. W. (2004). *Evaluation of Competency to Stand Trial-Revised (ECST-R) and professional manual.* Lutz, FL: Psychological Assessment Resources.

Simon, H.A. (1956). Rational choice and the structure of environments. *Psychological Review, 63*, 129-138.

Skeem, J., Schubert, C., Stowman, S., Beeson, S., Mulvey, E., Gardner, W., & Lidz, C. (2005). Gender and risk assessment accuracy: underestimating women's violence potential. *Law and Human Behavior, 29,* 173. doi: 10.1007/s10979-005-3401-z

Tombaugh, T. N. (1996). *Test of memory malingering: TOMM.* North Tonawanda, NY: Multi-Health Systems.

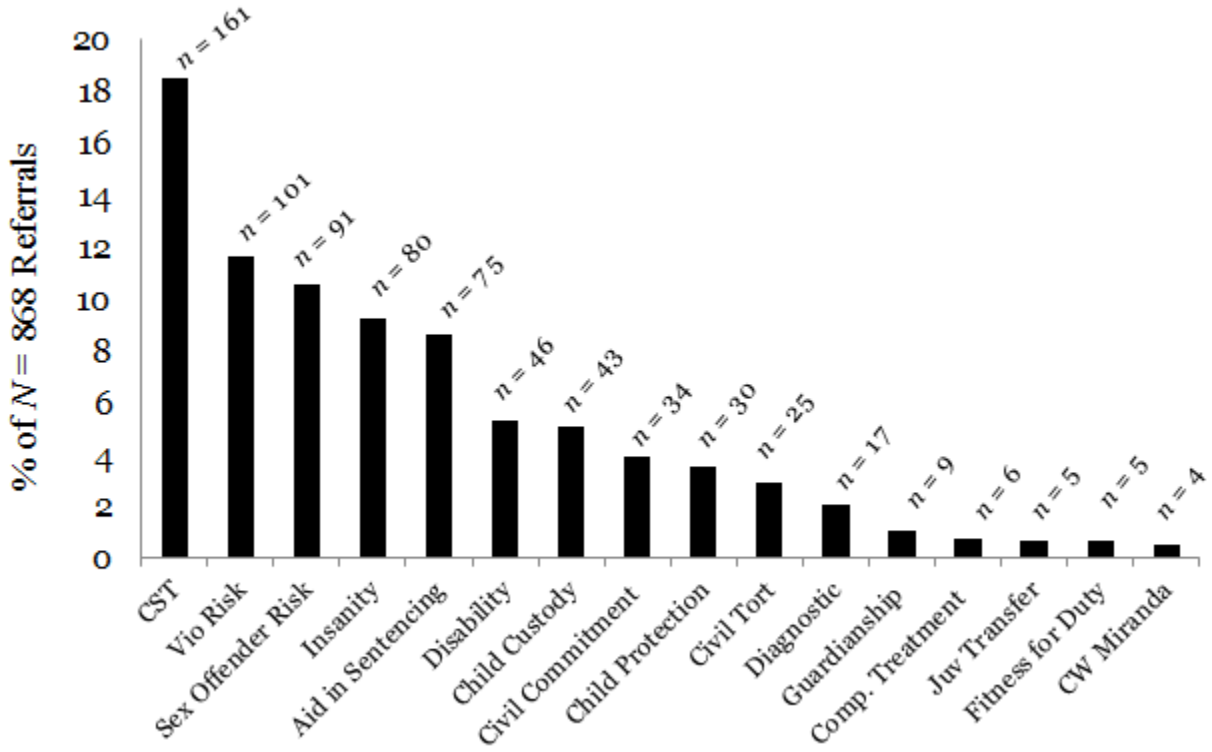
Weathers, F., Huska, J., & Keane, T. (1991). *The PTSD checklist military version (PCL-M).* Boston, MA: National Center for PTSD.

Webster, C. D., & Eaves, D. (1995). *The HCR-20 scheme: The assessment of dangerousness and risk.* Mental Health, Law and Policy Institute, Department of Psychology, Simon Fraser University and Forensic Psychiatric Services Commission of British Columbia.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition.* San Antonio, TX: Pearson.

Zapf, P.A. & Roesch, R. (2009). *Evaluation of Competence to Stand Trial.* New York, NY: Oxford University Press.

Figure 1. Relative Percentages of Various Referrals



Note: CST = Competence (Fitness) to Stand Trial. Vio Risk = Violence/Recidivism Risk. Insanity = Criminal Responsibility/Mental State at Time of Offense. Child Protection (e.g., child abuse, termination of parental rights); Civil Torts (i.e., psychiatric, psychological, or emotional disability evaluations in civil suits). Comp. Treatment = Competence to Consent to Treatment. Juv Transfer = Transfer (waiver) of a Juvenile to/from Adult Court. CW Miranda = Capacity to Waive Miranda Rights.

Table 1. Structured Tools Used, Time, and Report Length for Various Referrals.

	% using any structured tool	Average # tools used (if any used) *M* (*SD*) [Range]	# of different tools used	How long evaluation took (days) *M* (*SD*)	Report length (pages) *M* (*SD*)
Competence to Stand Trial	58.4	3.21 (2.68) [1-15]	65	24.98 (22.05)	13.02 (9.39)
Violence Risk	89.0	3.98 (2.95) [1-15]	110	35.91 (28.76)	19.22 (16.27)
Sex Offender Risk	96.9	4.17 (2.69) [1-15]	78	35.63 (25.68)	19.45 (15.56)
Insanity	71.8	4.24 (3.51) [1-16]	69	34.02 (27.75)	21.02 (18.04)
Aid in Sentencing	82.1	4.29 (3.04) [1-15]	84	35.45 (27.57)	15.53 (13.42)
Disability	65.5	4.42 (2.46) [1-9]	41	17.88 (21.57)	15.81 (15.76)
Child Custody	79.1	3.77 (2.38) [1-9]	35	44.13 (32.48)	32.19 (24.42)
Civil Commitment	83.9	3.62 (2.23) [1-8]	38	36.00 (27.28)	20.29 (13.16)
Child Protection	92.6	4.65 (2.41) [1-9]	38	34.83 (31.19)	23.87 (22.16)
Civil Tort	66.7	4.60 (5.07) [1-18]	42	29.44 (29.07)	17.37 (14.95)
Average Across All Referrals	*74.2*	*4.0 (2.95) [1-18]*	*286**	*30.76 (27.28)*	*16.96 (15.48)*

Note: We included more detailed information here only for referrals with $n \geq 25$ reports. Structured tools were described as "Structured Assessment Tools (e.g., tests, instruments, checklists, rating systems)." *This represents the total number of unique tools among all tools mentioned across the 10 areas.

Table 2. Sources of Information used for Various Referral Questions (Percent of evaluations using each source)

	CST	Violence Risk	Sex Offender Risk	Insanity	Aid in Sentence	Disability	Child Custody	Civil Commit	Child Protect	Civil Tort	*Average Across Referrals*
Examinee Interview	99.4%	99.0%	100%	100%	100%	100%	100%	100%	100%	100%	*99.2%*
Mental Health/Medical Records	90.1	90.9	79.5	97.4	89.6	100	72.1	100	81.5	83.3	*88.7*
Justice System Records	91.3	95.1	97.6	96.2	97.0	17.2	48.8	93.5	55.6	33.3	*77.1*
Non-Professional Collat. Interviews	41.6	27.0	35.2	55.1	40.3	10.3	86.1	41.9	59.3	41.7	*38.5*
Professional Collateral Interviews	43.5	54.5	25.0	37.9	31.3	6.9	72.1	54.8	44.4	16.7	*40.1*
Educational Records	26.7	26.0	30.7	26.9	40.3	3.4	39.5	19.4	25.9	20.8	*25.8*
Additional observation of examinee	28.0	25.7	5.7	25.6	17.9	31.0	46.5	3.2	66.7	25.0	*22.4*
Biological Tests	13.0	12.0	12.5	26.9	20.9	13.8	11.6	19.4	14.8	16.7	*14.3*
Visit to relevant environment	0.6	7.0	4.5	3.8	4.5	0	32.6	3.2	29.6	8.3	*5.8*
Employment Records	0.1	0	0	2.6	0	24.1	0	0	0	0	*3.4*
Other	0.1	0	0	2.6	0	0	11.6	0	0	0	*0.2*

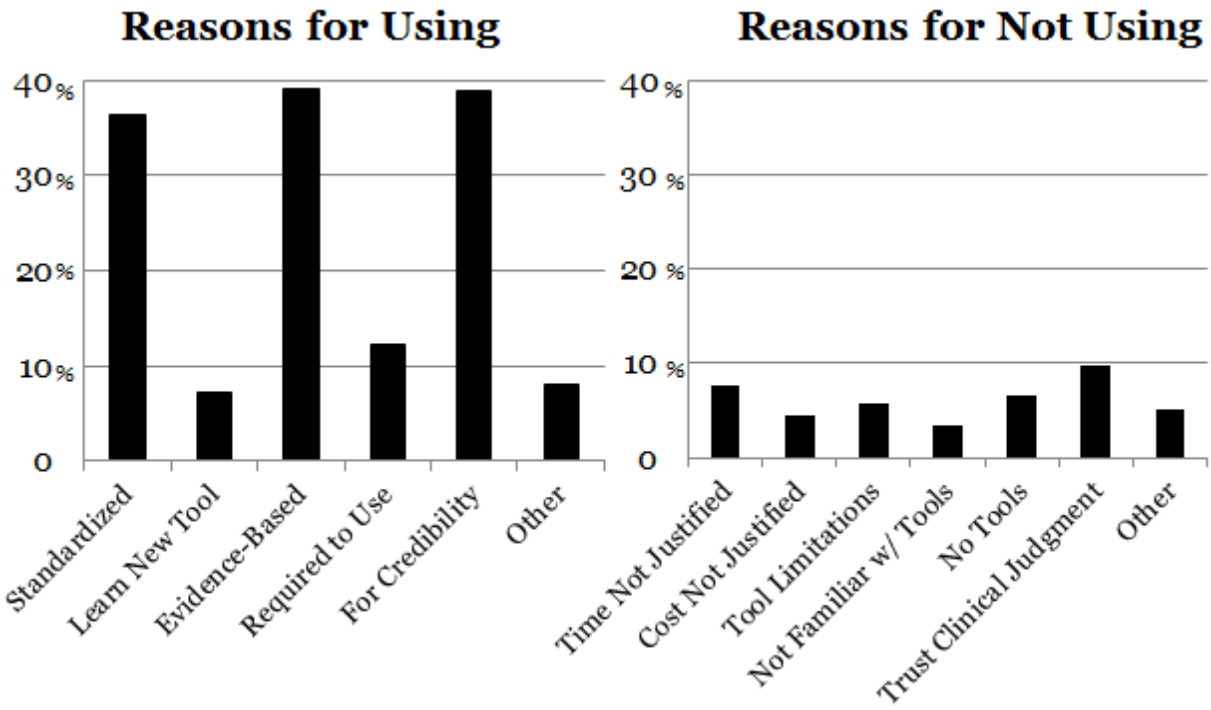
Note: We included more detailed information here only for referrals with $n \geq 25$ reports. Structured Assessment Tools are represented in Table 1.

Table 3. Ten Most Frequently Used Tools per Referral Question (Percent of evaluations using each tool)

	Average	CST	Violence Risk	Sex O. Risk	Insanity	Aid in Sentence	Disability	Child Custody	Civil Commit	Child Protect	Civil Tort
1	*MMPI - 15.2%*	WAIS - 11.8%	HCR-20 -35.6%	Static 65.9%	MMPI - 21.8%	MMPI - 35.8%	MMPI -51.7%	MMPI - 60.5%	Static - 64.5%	MMPI - 44.4%	Trauma -25.0%
2	*PAI - 9.6*	TOMM - 9.9	PCL-R - 35.6	PCL-R - 35.2	PAI - 17.9	MCMI - 19.4	Trauma - 27.6	MCMI - 27.9	PCL-R -24.3	MCMI - 40.7	TSI - 20.8
3	*Static - 7.4*	M-FAST - 9.3	VRAG - 17.8	MMPI - 27.3	TOMM - 16.7	WAIS - 19.4	BAI - 17.2	PAI - 25.6	SVR-20 - 13.5	PSI - 25.9	MMPI - 20.8
4	*PCL-R - 6.9*	MMPI - 8.1	MMPI - 13.9	PAI - 22.7	WAIS - 15.4	PCL-R - 16.4	BDI - 17.2	PSI - 25.6	MnSOST - 10.8	Rorschach - 22.2	PAI - 16.7
5	*MCMI - 6.7*	ECST-R - 6.8	PAI - 13.9	MCMI - 17.1	M-FAST - 12.8	PAI - 14.9	PAI - 13.8	Rorschach - 20.9	MMPI - 10.8	CAP - 18.5	BAI - 8.3
6	*WAIS - 6.7*	CBCL - 6.8	LS/CMI - 12.9	SVR-20 - 13.6	PCL-R - 12.8	SASSI - 14.9	SIMS - 13.8	CAP - 18.6	PAI - 10.8	PAI - 18.5	BDI - 8.3
7	*HCR-20 - 3.2*	PAI - 6.2	Static - 8.9	RSVP - 10.2	Rorschach - 9.0	LS/CMI - 10.5	TOMM - 13.8	CBCL - 11.6	RRASOR - 10.8	BASC-2 - 14.8	BNT - 8.3
8	*TOMM - 3.2*	WRAT - 6.2	LSI-R - 7.9	Stable - 9.1	SIMS - 7.7	WRAT - 9.0	PCL-M - 10.3	STAXI - 11.6	WAIS - 10.8	TSI - 14.8	COWAT - 8.3
9	*Rorschach -3.2*	MacCAT-CA - 5.6	SASSI - 6.9	SORAG - 7.9	SIRS - 7.7	CBCL - 7.5	Dep. - 10.3	BASC-2 - 9.3	MSI - 8.1	PDS - 11.1	Trails - 8.3
10	*M-FAST - 2.8*	JACI - 5.0	WASI - 6.9	VRAG - 7.9	VIP - 7.7	LSI-R - 7.5	Pain - 10.3	PDS - 9.3	WCST - 8.1	STAXI - 11.1	WAIS - 8.3

Note: We categorized tools together if they were different versions of the same tool (e.g., MMPI-2, MMPI-2-RF, MMPI-2 RC, and MMPI-A all reported as "MMPI.") In both the Disability and Civil Tort columns, "Trauma" refers to "other trauma measures," such as the Detailed Assessment of Posttraumatic Stress, the Clinician-Administered PTSD Scale, the PENN PTSD scale, PTSD Checklist, Davidson PTSD scale, Posttraumatic Distress Scale, Morel Emotional Numbing Test for PTSD, and the Mississippi Scale for Combat-Related PTSD. In the Disability column, "Dep." Refers to other depression inventories, including the Hamilton and Zung Depression Scales, and "Pain" refers to the Pain Patient Profile and the Pain Catastrophizing Scale. The full names of the tools in alphabetical order by acronym (*that are not cited in the results text*) are as follows: BAI = Beck Anxiety Inventory; BASC = Behavior Assessment System for Children, Second Edition; BDI = Beck Depression Inventory (includes both the first and second editions); BNT = Boston Naming Test; CAP = Child Abuse Potential Inventory; CBCL = Child Behavior Checklist; COWAT = Controlled Oral Word Association Test; LS/CMI = Level of Service/Case Management Inventory (also included the YLS/CMI (Youth LS/CMI) and the LS/RNR (Level of Service/Risk, Need, Responsivity); LSI-R = Level of Service Inventory – Revised; M-FAST = Miller Forensic Assessment of Symptoms Test; MCMI = Millon Clinical Multiaxial Inventory (also includes the MACI, Millon Adolescent Clinical Inventory); MMPI = MMPI-2, MMPI-A (Adolescent), MMPI-2-RF (Restructured Form), MMPI-2 RC (Restructured Clinical Scales); MnSOST (Minnesota Sex Offender Screening Tool); MSI (Multiphasic Sex Inventory, includes versions I and II); PAI (includes PAI-A, Adolescent version); PCL-R (includes the 1st and 2nd Editions and the PCL-YV, Youth Version); PDS = Paulhus Deception Scales; PSI = Parenting Stress Index (includes all versions); Rorschach = Rorschach Inkblot Test (includes the R-PAS, Rorschach Performance Assessment System); RRASOR = Rapid Risk Assessment for Sex Offense Recidivism; SASSI = Substance Abuse Subtle Screening Inventory (includes all versions); SIMS = Structured Inventory of Malingered Symptomatology; SIRS = Structured Interview of Reported Symptoms (includes both the first and second versions); Static = Static-99, 99-R, 2002R, 2007; STAXI = State-Trait Anger Expression Inventory (includes all versions); Trails = Trail Making Test; TSI (includes all versions of the TSI and the TSCC, Trauma Symptom Checklist for Children); VIP = Validity Indicator Profile; WCST = WCST-64, Wisconsin Card Sorting Task; WAIS = Wechsler Adult Intelligence Scale (also included WISC, or Wechsler Intelligence Scale for Children); WASI = Wechsler Abbreviated Scale of Intelligence; WRAT = Wide Range Achievement Test.

Figure 2. Reasons Forensic Clinicians Reported for Using and Not Using Structured Tools



Note: The values refer to the percentages of forensic clinicians who endorsed each reason for using/not using tools.