Running Head: Clinicians and Clients Disagree

Clinicians and Clients Disagree: Five Implications for Clinical Science

Douglas B. Samuel

Takakuni Suzuki

&

Sarah A. Griffin

In press; Journal of Abnormal Psychology

Authors' Notes:

Douglas B. Samuel, Takakuni Suzuki, and Sarah A. Griffin, Department of Psychological Sciences, Purdue University.

Correspondence concerning this article should be addressed to Douglas B. Samuel, Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907. E-mail: dbsamuel@purdue.edu

Abstract

Recent findings highlight the limited agreement between diagnostic ratings provided by practicing clinicians and the self-report and interview methods typically employed in research settings. Such discrepancies between the diagnoses assigned in research and applied settings greatly complicate the translation of empirical findings into practice. This review highlights these disagreements, offers explanations for these observed differences, and provides five implications for research. Specifically, we provide evidence that, despite criticisms, self-reported psychopathology may be at least as valid as clinicians' unstructured diagnoses. Further, we highlight the need for research that provides clinicians with the most valid tools, including those that focus on dimensional constructs, rather than diagnostic categories. In addition, we recommend that adult psychopathology research incorporate methodologies from general personality for unraveling informant discrepancies. We highlight recent work that has provided valuable tools for incorporating meta-perception—the extent to which one is aware of how they are perceived by others—for contextualizing these differences. We also underscore the utility of emerging technologies that provide rich data, such as ambulatory assessment, for overcoming the criterion problem. Finally, we recommend advances in combining data from multiple sources within the childhood psychopathology literature, such as examining the extent to which discrepancies themselves might aid in diagnosis, be incorporated into adult psychopathology research. In sum, we hope that these implications inspire research that improves the science of diagnostic assessment in a way that might ultimately improve practice.

General Scientific Summary: This paper reviews studies reporting limited agreement between diagnoses assigned in research and clinical settings. We highlight the implications of these

findings and offer recommendations for improving research on personality disorders and mental disorders, more broadly.

Keywords: Diagnosis, Multi-method, agreement, dimensional, personality disorder

Clinicians and Clients Disagree: Five Implications for Clinical Science

Research has long-demonstrated a schism between the methods used to diagnose mental disorders within clinical practice and those employed in research settings. The modal therapist relies on holistic, gestalt impressions formed on the basis of unstructured interviews and clinical interactions to assign diagnoses and rely much less heavily on explicit information about diagnostic criteria conveyed by clients. In contrast, diagnoses in research settings are almost exclusively based on either semi-structured interviews administered by research personnel or self-report questionnaires that systematically evoke information from the client to determine the presence of each diagnostic criterion. Thus there are crucial differences in terms of the source of the information, the method of gathering data, and the level of abstraction. It is perhaps not surprising then that a recent meta-analysis revealed broad differences in the diagnoses from each method (Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009).

Such inconsistencies have major implications for the translation of empirical findings from clinical science into evidence-based practice. Results from clinical trials of pharmaceutical or talk-based therapeutic interventions form the basis for practice guidelines (American Psychiatric Association, 2016) and specify the gold-standard of clinical care. Similarly, the American Psychological Association has formed an Advisory Steering Committee (chaired by Steven D. Hollon) that is charged with developing a set of Clinical Practice Guidelines that will guide treatment selections for various disorders. Ultimately, these formal practice guidelines, based on comprehensive reviews of the literature, provide recommendations about the interventions that are most likely to be effective and increasingly determine reimbursement from third-party payers.

Nonetheless, given the unique methods employed in clinical and research settings as well as the limited agreement across them, it stands to reason that groups of patients diagnosed by each may differ substantially. If the individuals diagnosed with a given disorder within clinical trials differ from those diagnosed in practice settings, this greatly complicates the degree to which empirical research findings will be applicable in real-world settings. For example, if a novel pharmaceutical or psychotherapy shows effectiveness in reducing symptoms among a group of individuals diagnosed with Major Depressive Disorder (MDD) within a clinical trial, this should inform clinical care. However, if those diagnosed with MDD within the clinical trial differ substantially from those who are being diagnosed in practice settings, then it reduces the confidence that the same intervention will work with a given patient.

As noted above, this discrepancy between research and clinical methods appears to be the rule, rather than the exception. Rettew and colleagues (2009) synthesized 38 studies that reported the agreement between structured diagnostic interviews and diagnoses made by standard (i.e., unstructured) clinical evaluations. They reported some discrepancy across classes of disorders with the highest agreement for eating disorders and psychotic disorders (K = .70 and .67, respectively), but the overall agreement across the 38 studies was quite meager (K = .27). This overall discordance suggests a rather limited capacity for the translation of research findings into practice. Unfortunately, given the routine divergence of these methods, even fewer studies have compared the relative validity for predicting outcomes.

In this manuscript, we further articulate the potential causes and consequences of this limited agreement and detail five key implications for clinical science. It is also important to note that we do not focus on the implications for clinical practice, nor do we make explicit recommendations for improving practical diagnosis as we feel these questions are beyond the

scope of this manuscript and have been covered elsewhere (e. g., Widiger & Samuel, 2005b). We do explore possible causes for the limited agreement from both sides, including those factors that might deter rigorous diagnoses in clinical settings. We hope that these observations will form hypotheses worthy of testing in future research to explicate the nature of disagreement across these methods.

In the majority of this paper, we utilize the personality disorder diagnoses as an exemplar for further probing the question of clinician-client agreement. A specific focus on personality disorder (PD) is relevant in this regard primarily as recent studies have further articulated the agreement between clinical and research methods, as well as their relative validity (Samuel, 2015; Samuel et al., 2013). Personality is also generally relevant as some have argued that the dimensions underlying PD have substantial links with broad spectra that cut across psychopathology more generally (Krueger & Markon, 2014). Thus, personality pathology makes a particularly compelling exemplar by which to probe the broad agreement between methods of diagnosis employed in clinical and research settings. In this way, the following material focuses on PD as a specific case, but with implications that apply across the diagnostic manual.

Samuel (2015) recently provided an in-depth examination of the agreement between clinicians' naturalistic PD diagnoses and those assigned by research methods. This extended the work of Rettew and colleagues by including self-report questionnaires as well as semi-structured interviews and by disaggregating agreement based on the type of instrument used to collect the clinicians' descriptions. The last point was necessary as clinicians often provide ratings using very brief rating forms with limited validity evidence. Focusing specifically on the few studies that employed longer and more robust clinician rating tools allowed for a method-matched comparison with self-report instruments that isolated source differences.

Aggregated across 27 studies, this review revealed weak correspondence across clinical and research methods (Samuel, 2015). The median kappa agreement for categorical diagnoses was .26, while the median correlation for dimensional ratings was .23. These overall values appeared to be moderated slightly by the method used to collect descriptions from the client as well as the therapist. Specifically, agreement increased slightly when clinicians' ratings were correlated with semi-structured interviews versus a self-report questionnaire. In addition, when clinicians' PD ratings were aggregated using a systematic measure, such as the Shedler-Westen Assessment Procedure (Shedler, 2015), the agreement increased slightly. In sum, it appeared that although using more systematic and structured methods improved the agreement between clinicians and clients, the overall magnitude was fairly modest. Indeed, Samuel (2015) noted that the agreement between clients and therapists was lower than the agreement between self- and peer-reported PD pathology (Klonsky, Oltmanns, & Turkheimer, 2002). This indicates, perhaps ironically, that clients' descriptions of their PD pathology overlap less with descriptions from their therapists than with those from other lay informants.

Beyond the potential to complicate the translation of research on PD treatment into practice, disagreement across methods is not itself inherently problematic. Indeed, modest agreement across informants is found rather routinely for a variety of psychopathology constructs in adults (Achenbach, Krukowski, Dumenci, & Ivanova, 2005) and children (De Los Reyes & Kazdin, 2005). Moreover, a great deal of research in the child and adolescent literature has focused on the integration of information across sources and explicitly evaluated the utility of such discrepancies for understanding psychopathology (e.g., De Los Reyes, Salas, Menzer, & Daruwala, 2013). In contrast, there is very little known about integrating information across methods within the adult literature – although some initial results have also suggested unique

strengths (e.g., Hopwood et al., 2008). Nonetheless, a fundamental assumption underlying this integration is that information from various sources and methods have reciprocal validity for predicting outcomes of clinical interest. For personality pathology this has proved to be the case for self- and peer-informant reports, such that each source contributes meaningful variance beyond the other (Fiedler, Oltmanns, & Turkheimer, 2004; Klein, 2003; J. D. Miller, Pilkonis, & Clifton, 2005; South, Oltmanns, Johnson, & Turkheimer, 2011). Similarly, research from industrial-organizational psychology has routinely demonstrated incremental validity offered by multiple raters (Connelly & Ones, 2010). Interestingly, though, there has been very little research examining the reciprocal validity of PD ratings from clients and their treating clinicians.

A study by Samuel et al. (2013) is the only one which has explored validity of clinicians' PD diagnoses relative to those provided by self-report questionnaires and semi-structured interviews. In that study, 320 participants were rated for PDs by their treating clinician using a prototype matching approach that approximated the model initially proposed for DSM-5 (Shedler et al., 2011). The clients also completed a semi-structured diagnostic interview and a self-report questionnaire. These initial ratings were compared in terms of their ability to longitudinally predict two assessments of psychosocial functioning five years later using a series of regressions. Provocatively, Samuel and colleagues (2013) found that the semi-structured interview and self-report questionnaire routinely predicted incremental variance in psychosocial functioning beyond the clinician ratings, but the clinician ratings rarely predicted variance beyond those methods. These findings suggested that the research assessment methods might be more valuable than the clinician-assigned diagnoses, which form the basis of practice and have been considered a gold standard (Westen, DeFife, Bradley, & Hilsenroth, 2010). Clearly, the findings from Samuel and colleagues (2013) require replication using other samples,

instruments, and outcome criteria; however, the findings do indicate that disagreement between clinicians' ratings and other methods are, at the very least, not simply an indication of problems with self-report.

Taken together, we argue that these recent findings regarding the agreement and reciprocal validity between the methods of diagnosis typically used in clinical practice and those favored in research settings have important implications for clinical science research. Building on these findings, we outline five implications for improving assessment and diagnosis of psychopathology in research, using personality pathology as a specific exemplar.

Five Implications for Clinical Science Research

1. The Possible Limitations of Self-Report have been Overstated

Much has been written about the potential complications inherent to relying on a client's self-report to diagnose psychopathology, especially personality pathology. Clinical scholars have argued that a client's self-presentation may be colored by a number of biases, given the ego-syntonic nature of PD. Huprich and Bornstein (2007) have emphasized that self-reports may be limited by a lack of insight—that is to say that individuals may not accurately perceive their own difficulties or their stimulus value to others. Further, self-reports might be biased by deliberate attempts to portray oneself positively (e.g., in employment settings) or negatively (e.g., in forensic settings). Beyond issues of purposeful or inadvertent distortions, Westen (1997) and others have argued that clients may simply lack the ability to provide a valid description of their own personality pathology. Westen speculated that this might be due to skewed reference points for comparing oneself to others or the lack of sophistication on the part of a layperson to understand and describe these constructs. Similarly, McAdams and Pals (2006) have noted that humans are unique in their ability to craft self-narratives. It is possible that these narratives, in

turn, skew an individual's portrayal by prompting them to emphasize aspects that are most consistent with the story. Finally, it has been noted that a self-report, in a given instance, may be susceptible to fluctuations based on current mood states or the product of implicit processes, such as priming (Huprich, Bornstein, & Schmitt, 2011). In sum, potentially compelling reasons have been suggested for presuming that factors conspire to limit the validity of a self-report and, as a result, some have argued against a reliance on their use for PD assessment in research and clinical settings. Given this context, critics would likely suggest that limited correlation between clients and clinicians present further evidence for the inadequacy or inaccuracy of self-report methods. We argue that such a conclusion would be erroneous and that the presumed limitations of self-reports have been overstated.

As Meehl (1945) pointed out, accurate insight is a not a necessary ingredient for a valid self-report -- particularly when questionnaires rely on non-obvious items. As long as individuals answer questions in predictable ways and those responses are consistently linked with an outcome of interest, then the self-report responses can have predictive validity even without insight. In the case of empirically-keyed instruments, it need not be that a client (or the test developer) understands the link between the response and the outcome of interest for the test to be valid. Nonetheless, most modern instruments do not employ an empirical criterion-keying approach without any theoretical connection between item content and the construct of interest (Simms, 2008). Even still, items on modern self-report questionnaires can be written so as to suggest a milder variant of the characteristic that carries less negative connotation or requires less insight into the severity of the problem. For example, an individual with excessive antagonism might be more willingly endorse a mild item like "I guess you could say I look out for my own interests" as opposed to "I manipulate people to get what I want from them." In this

way, it is possible that well-constructed self-report questionnaires might validly assess constructs without adequate insight on the part of the respondent.

Withstanding the above caveat, there is also compelling data that diminish historical concerns about self-reported psychopathology. Notably, Carlson, Vazire, and Oltmanns (2013) reported that individuals are able to accurately report how others perceive them on narcissism, undermining a central concern about lack of insight in self-report. In addition, a meta-analysis found that self-report psychopathy scales were not strongly related to indicators of positive response bias (Ray et al., 2013). Similarly, Watts et al. (2016) performed a wide variety of analyses in a large offender sample and found very limited evidence that response distortion moderated or suppressed correlations between self-reported psychopathy and relevant external criteria. In short, these converging findings counter apprehensions about the validity of self-report questionnaires (Huprich et al., 2011) for personality pathology and refute the view that self-report is inherently limited by response style or inaccurate self-perception.

This should not be taken to indicate that self-report is infallible or without biases. We emphasize that no single method will provide a comprehensive picture and argue that ideal measurement should always employ multiple methods. However, recent research suggests that self-reports have been unfairly singled out as problematic. In fact, it might be the case that the biases present in typical ratings by clinicians outstrip whatever biases exist within self-reports. Although rarely discussed, there are a number of structural reasons why typical diagnoses and ratings by clinicians might be limited. First, despite expertise in the diagnostic constructs, most clinicians do not systematically apply the diagnostic criteria (Westen, 1997), instead relying on global gestalts and other heuristics (Morey & Benson, in press; Morey & Ochoa, 1989). This approach has been shown to result in a series of biases. For example, clinicians over-weight the

presence of certain salient symptoms (e.g., self-harm for borderline PD) when assigning diagnoses (Kim & Ahn, 2002). In addition, clinicians appear to differentially diagnose cases on the basis of demographic features, such as gender (e.g., Samuel & Widiger, 2009) and ethnicity (Mikton & Grounds, 2007). These findings illustrate systematic propensities in clinical diagnostic practices that might limit the validity of the resulting diagnoses.

Beyond these diagnostic biases, there are a number of factors inherent to the typical clinical practice that might limit the validity of clinician diagnoses in extant studies (and in routine practice). For instance, the current reimbursement structure of third party-payers provides little incentive for therapists to spend appreciable time on thorough assessment (Butcher, 2006). Without financial incentive, it stands to reason that busy clinicians would not prioritize this complex enterprise unless it was central to the therapeutic intervention (C. V. Wright et al., in press). Compensation for therapeutic services also may differ by the client's reported presenting problem, such that therapists may be more likely to receive third-party compensation for treating salient behavioral problems or symptoms as opposed to treating the underlying personality pathology, possibly further disincentivizing PD assessment.

Beyond these systemic factors, there may also be practical limitations to the information available to clinicians. Mental health practitioners almost exclusively interact with clients in a single, highly-controlled setting that features well-defined roles and a clear power differential. Such a context assuredly restricts the range of behaviors exhibited by the client and thus limits the information available to the therapist (Carlson, 2015). In sum, it might be the case that clinicians generally have restricted access to valuable information that could inform diagnosis. Thus, even when asked to provide their best diagnoses of clients within research studies, clinicians may simply lack the relevant information to provide maximally valid ratings. Our

point in highlighting these potential limitations is neither to denigrate clinicians nor to suggest they are shirking an important duty. We are merely drawing attention to aspects of clinical diagnosis that deserve consideration and further empirical attention. Regardless of the reasons, current evidence suggests that self-reports appear to have advantages over clinician ratings.

One significant advantage of self-reported personality pathology is the availability of a number of comprehensive measures that have undergone considerable validity examination within normative samples. In contrast, clinicians' typical method of aggregating information and arriving at PD diagnoses is to simply record categorical diagnoses within a chart. Westen and Weinberger (2004) had argued for the necessity of providing clinicians with systematic methods for aggregating their clinical insights. However, the SWAP (Shedler, 2015) is the only measure that has been routinely used to collect reports from clinicians that is comparably systematic. The SWAP consists of 200 statements relevant to PD that a clinician sorts into one of eight successively smaller piles, indicative of how descriptive it is of the target. For example, half of the items must be marked as "0" or inapplicable to a given client, while eight items are sorted into the highest rating. The SWAP has been criticized on the basis of that fixed distribution (Block, 2008; Wood, Garb, Nezworski, & Koren, 2007), but has shown the ability to modestly improve the convergence of therapist ratings with other methods (Samuel, 2015). Ultimately, to determine the true validity of therapist ratings relative to self-report methods, they must be collected via equivalent scales to remove measurement confounds and isolate source effects.

We wholeheartedly agree with calls for diagnostic practices that emphasize multimethod assessment (Ganellen, 2007; Lenzenweger, Loranger, Korfine, & Neff, 1997; Widiger & Samuel, 2005b). Indeed, there have been repeated demonstrations of the value of information from multiple sources and methods for overcoming limitations of any single assessment (Carlson

et al., 2013; Klein, 2003; McCrae, 2013; J. D. Miller et al., 2005; Samuel & Widiger, 2010; Tromp & Koot, 2010). In this respect, we do not suggest that self-reports, clinician ratings, or any other method should necessarily be preferred over any other. Rather, we encourage a multimethod battery in order to maximize the depth and breadth of information available. Nonetheless, the findings we highlight here counter long-standing concerns about the validity of self-reports and suggest several reasons to believe that for PD—and likely for much of psychopathology—self-reports should remain a highly valued component of any diagnostic battery.

2. Research Must Focus on Homogeneous Dimensions Rather than Heterogeneous Categories

It is worth noting that clinicians' naturalistic diagnoses focus on the categorical mental disorders that appear in DSM-5, but research has repeatedly demonstrated the limitations of these categories (Clark, 2007; Skodol, 2014; Trull & Durrett, 2005). Amid widely acknowledged complications such as heterogeneity within the categories, diagnostic instability, as well as slow progress in identifying etiologies, biomarkers, or specific treatments for the DSM-defined syndromes, researchers have pushed toward dimensional alternatives that cut across traditional diagnoses (Caspi et al., 2014; Cuthbert & Kozak, 2013). For example, a broad literature has demonstrated that the symptoms of a wide-variety of common mental disorders can be sorted into the higher order internalizing, externalizing, and psychotic spectra (Krueger, 1999; A. G. C. Wright et al., 2013). Additionally, seeking to better integrate findings from neurobiology, the National Institute of Mental Health has proposed a novel set of Research Domain Criteria (RDoC) that seeks to more cleanly parse the universe of psychopathology (Sanislow et al., 2010). Research has repeatedly indicated that continuous measures of symptoms have more

robust psychometric properties and offer promising potential for diagnosing mental disorders (Markon, Chmielewski, & Miller, 2011). These alternative structural models will likely become increasingly integrated into future diagnostic nomenclatures (Casey et al., 2013; Widiger & Samuel, 2005a).

The diagnosis of PD is no exception and has long been considered the portion of the diagnostic manual that is most in need of a revision (Bernstein, Iscan, & Maser, 2007). It also has been suggested as an ideal proving ground for incorporating dimensions into the diagnostic process (Rounsaville et al., 2002). The traditional PD categories include an abundance of clinical information, but have a myriad of well-known flaws as currently organized (Skodol, 2014). These have been sufficiently elaborated elsewhere (Clark, 2007; Krueger & Markon, 2014; Samuel & Griffin, 2015; Skodol, Morey, Bender, & Oldham, 2013) and include issues such as inadequate coverage of the universe of PD pathology, notorious heterogeneity within categories, excessive overlap across categories, and dissatisfaction by the clinicians who use it. It should go without saying that it will be impossible to accurately detect agreement across parties when the constructs being rated have problematic reliability or validity. Thus, a crucial implication is that the literature reviewed by Samuel (2015) might have underestimated the true agreement between clinical and research methods. This also leads to the recommendation that future research incorporate dimensional models that more clearly demarcate homogeneous units of PD pathology and thus more precisely detect the agreement (and disagreement) across sources.

The last twenty years have seen a great deal of research on the specifics of an alternative dimensional model of PD (Widiger & Simonsen, 2005). Although the specific lower-order traits (i.e., facets) differ across models, this research has largely coalesced around five broad higher-order domains that are common to a variety of dimensional models (e .g., Harkness, Finn,

McNulty, & Shields, 2012; Livesley & Jackson, 2009; Widiger & Mullins-Sweatt, 2009). These five domains reasonably capture the variance in the existing PD constructs (Widiger & Trull, 2007) and display sizeable relations with psychopathology more broadly (Kotov, Gamez, Schmidt, & Watson, 2010). The DSM-5 Section III alternative PD model reflects these domains, which are highly similar to the PSY-5 (Anderson et al., 2013) and have been show to reflect maladaptive, extreme variants of the Five Factor Model (FFM) that describe general personality functioning (Krueger & Markon, 2014; Suzuki, Griffin, & Samuel, in press).

The few studies that have examined agreement via the trait dimensions of the FFM that appear in DSM-5 Section III have indicated slightly improved agreement across clinical and self-reported PD ratings (e.g., Few et al., 2013). In addition, Samuel and Widiger (2010) compared the agreement across a variety of sources, including self- and clinician-report, and found that the FFM traits achieved greater convergent and discriminant validity than the DSM PD categories. However, the modest magnitude of this lift clearly indicates that the dimensional trait model will not be a panacea and that genuine and meaningful differences do exist across these sources. Nonetheless, the dimensional traits appear to be a promising upgrade over the categories and it is only through studies that carefully control extraneous factors, such as the internal validity of the measures, that the field will be able to truly discern the overall nature of the agreement as well as the specific factors that may influence it.

One such method of maximizing internal validity in that way would be to collect descriptions from therapists using a well-validated measure of the dimensional traits. Within our lab, we have begun to do this by providing clinicians with an informant version of the Personality Inventory for DSM-5 (PID-5-IRF; Markon, Quilty, Bagby, & Krueger, 2013) that is largely equivalent to the self-report operationalization of the DSM-5 Section III traits. This has

the advantage of providing a commensurate instrument for both parties to isolate source effects for the domains and facets of the dimensional trait model. Although results are still being evaluated, it appears that the use of a well-validated trait instrument does further enhance the convergence and sharpen distinctions across sources.

In sum, a key implication of past research is that the use of heterogeneous categories undermines the agreement across sources as it combines variance across distinguishable constructs (Smith, McCarthy, & Zapolski, 2009). Instead, research must focus on the homogenous dimensions that will allow greater clarity regarding the true agreement and fully articulate the relative value of different sources for assessing those constructs. In order to accurately assess the agreement between clinician and client ratings, we must assess that agreement using constructs that are clearly articulated and equitably assessed by both clinician- and client-rated measures. The research to date suggests that the best way to achieve this is to increase utilization of dimensional models of psychopathology in both clinician and client assessments.

3. Embrace Meta-Perception to Articulate Limitations of Self-Knowledge

As noted previously, a chief concern about the potential limitations of self-report for individuals with PD is that they lack self-knowledge and have limited awareness of their own stimulus value. Within a rater-agreement context, this can reasonably be summarized as whether a person can adequately understand how others perceive them. When self and informant ratings do not agree, it may be the case that the target cannot appreciate that they perceive their own personality in a way that is discrepant from how others perceive them. In other words, that their self-report and meta-perception (i.e., how they believe they are perceived by others) are not identical. Alternatively, it is possible that the target is well-aware of how they are described by

an informant, yet simply does not agree with that description. This distinction has important consequences for understanding the value of self-report. If the former situation is true and targets cannot accurately appreciate how they are perceived by others, it suggests that self-reports may indeed be flawed in important ways. If the latter is true, it suggests that self-informant discrepancies are based on genuinely different views, rather than on faulty self-perceptions.

Recent advances from general personality have provided a novel ability to arbitrate these two possibilities. Vazire's (2010) work on what she terms Self-Other Knowledge Asymmetries (SOKA) has provided data regarding informant discrepancies for general personality traits. Vazire hypothesized that targets and informants have unique areas of strength in their perception of the target and that these strengths are largely understandable based on the inherent differences in the type of information available to each source. Specifically, individuals have greater knowledge about personality traits that are unobservable and low in evaluativeness –favoring self-report for their assessment—while observers are better positioned to provide accurate ratings of highly observable and/or evaluative traits. Initial tests have supported this view and found that self-report was most accurate for more internal traits, such as neuroticism. Likewise, informant reports were more accurate for highly evaluative traits like creativity. Self- and informant-reports appear to be relatively equally valid for highly observable and unevaluative traits like extraversion (Vazire, 2010).

Given the known links between general personality traits of the Five Factor Model and pathological personality (e.g., Gore & Widiger, 2013; Samuel & Widiger, 2008; Suzuki, Samuel, Pahlen, & Krueger, 2015; A. G. C. Wright & Simms, 2014) it stands to reason that similar effects might apply to PD. Then again, it is possible that biases in self-report might become more apparent among individuals with pathological levels of traits that are likely more evaluative (e.g.,

narcissistic grandiosity). Recent research has examined differences in self and other perception with PD constructs (e. g., Carlson et al., 2013) and presented similar results, reinforcing the notion that self-reports tend to be most accurate for PD pathology focused on internalizing symptoms of PD whereas informant reports have more value for externalizing pathology related to agreeableness and conscientiousness (Carlson et al., 2013).

In addition to differential validity for specific traits across sources, one particularly interesting methodology for arbitrating the value of self-report is asking individuals not only to describe themselves, but also how they believe they are perceived by others. This meta-perception (Kenny & Depaulo, 1993) can be particularly valuable for integrating information across targets and informants. A common way this is implemented is to have the target make self-report ratings of their personality (or personality pathology) as well as ratings from the perspective of an informant (e.g., a spouse or peer). Secondly, that informant also provides ratings of the target's personality. These perspectives index the degree to which the two raters genuinely agree (i.e., self-report with informant report) as well as how well the target can appreciate how the informant sees him/her (i.e., the relation between the meta-perception and the informant report).

This type of method has great power to characterize the level of insights among informants and at times have revealed results that counter the prevailing wisdom. Carlson, Vazire, and Oltmanns (2011) collected these ratings for narcissism and found that individuals possessed the ability to perceive their stimulus value accurately (i.e., reporting that others perceived them less positively than they regarded themselves) and were aware of how others perceived their levels of narcissism. Nonetheless, it does appear to be the case that meta-perception is decreased at elevated levels of personality pathology (Carlson & Oltmanns, 2015),

supporting the contention that individuals with PD pathology are less aware of how others perceive them. Interestingly, though, the direction of the effect indicated that self-reports tended to be *more* maladaptive than the observers' reports. This means that, contrary to popular concerns about underreporting, individuals with elevated PD pathology tended to show a *negativity* bias and over-reported pathology relative to informants. This stands in stark contrast to the typical view that individuals with PDs are unaware of how their behaviors negatively impact others (i.e., the self-absorbed narcissist). Rather, it appears that they *over*estimate the degree to which others see them negatively. Miller and colleagues (2011) reported an analogous finding in that levels of self-reported psychopathy – a disorder thought to be characterized by lack of insight – were not significantly different than reports from informants, providing further refutation of the view that individuals see themselves less pathologically than do informants.

There are two specific ways in which these advances might be profitably employed to improve clinical science research. First, as noted by Carlson (2015), clinicians might be able to tap into the valid meta-perceptions of their clients by asking them to describe themselves as perceived by others. Meta-perception assessment has promise to provide useful information that is beyond self-report, in that it is possible that such an avenue might allow a person to overcome self-presentation biases. Second, and most important for understanding the relative validity of each source, it would be of great interest to collect meta-perception data from the client and therapist dyads. This would entail asking a client to provide his own self-description and ratings of how he is perceived by his therapist. Similarly, the therapist would provide ratings of her perception of the client's personality as well as her rating of how the client perceives himself. Such a method would provide rich data that could answer the question about who knows what about whom within the therapeutic relationship. Not only would this reveal the extent of the

client's meta-perception (i.e., the degree to which they know how they are seen by the therapist), it would also provide a potential index for the precision of therapist ratings.

Finally, it would be ideal to incorporate this meta-perception methodology along with independent, method-neutral criteria, such as treatment outcome. This would help to determine which ratings, from which source, are most valid. For example, it is possible that the most valid rating for predicting an outcome might be the client's rating of how they are viewed by the therapist.

4. Transcend the Criterion Problem to Arbitrate the Validity of Sources

One major issue with arbitrating the validity of different sources is that the criteria themselves often are confounded with source-specific variance. For example, it would hardly be surprising for a self-report rating to have greater predictive validity than an informant rating, if the outcome of interest is itself self-reported. Although mechanisms to get around this potential method overlap exist, PD pathology research needs more objective, or method-neutral, criteria. The PD literature all too often employs hetero-method criteria that are, in many cases, simply alternative measures of similar constructs. Establishing agreement across instruments and sources is surely necessary for a field that is reaching consensus on the constructs that should be identified (Krueger & Markon, 2014), but ultimately the field will benefit most from identifying external criteria that transcend method.

There are a number of examples of how this has been accomplished within personality and personality pathology research. Chief among these are the naturalistic outcomes that occur within certain systems. A wonderful example is the study from Fiedler et al. (2004), which utilized military service records to demonstrate that informant ratings of PD pathology incremented self-reports for predicting major outcome such as formal discipline and discharge.

Similar efforts for PDs in non-military samples might include functional indicators in employment or scholastic settings that are objectively available (e.g., days of missed work, GPA, salary, etc.). Other indicators relevant to PD pathology might include interactions with social agencies, such as arrest records, hospitalizations, traffic citations, or health care utilization. Although each of these – and nearly all external indicators – are multiply determined such that any specific indicator will never account for a large portion of variance, they have value as objective manifestations of personality pathology when aggregated across time and individuals.

Ideally, the criteria might themselves be either a target or outcome of treatment. There are a number of treatment outcome variables that are logical criteria of interest within outpatient treatment settings. For example, it would be informative to determine how client or therapist ratings of personality relate to the number of sessions attended (versus those missed), symptom reduction during treatment, homework completion rate, or successful/mutual termination. Not only would these outcomes provide powerful indicators of the relative validity of diagnostic ratings from each source, but it would do so using criteria that are pragmatically relevant to therapeutic encounters. One potential example of this approach would be to integrate self-report and therapist-reported personality ratings into ongoing treatment outcome studies. Such an arrangement would represent an ideal entry-point for research on the relative value of informants, as psychotherapy studies feature well-characterized outcome tracking that would serve as objective criteria. Further, such studies already have a captive audience of therapists and clients so inserting diagnostic ratings from clients and therapists would require minimal additional expense or burden to make significant contributions to the literature.

Behavioral observations have also long been considered potential validators of personality assessment and modern technology has made this even more accessible as methods

for ambulatory assessment have advanced rapidly (Trull & Ebner-Priemer, 2013). In addition to well-tested methods like ecological momentary assessment that provide information about the dynamic interplay of symptoms and behaviors across time, applications within participants' existing smartphones also allow the collection of rich data regarding activity level and movements across space using GPS (G. Miller, 2012). In addition to cell phones, wearable technologies increasingly provide unique information about real-world behavior. For example, the Electronic Activated Recorder (EAR) is a recorder that can be worn by targets that records sounds at random intervals throughout the day (Mehl, Pennebaker, Crow, Dabbs, & Price, 2001). These ambient recordings are coded to produce summary indicators of an individual's behavior and have been used to determine the relative validity of alternative sources (Vazire & Mehl, 2008). The incorporation of video recordings as technology advances will allow even greater ability to encode feelings and behaviors as they occur in naturalistic settings.

Finally, a wide variety of psychobiological methods might provide additional sources of validation for PD information across various diagnostic methods. Some examples include links between traits and MRI scans (DeYoung et al., 2010), psychophysiological scalp EEG recordings (Hill, Samuel, & Foti, in press), behavioral tasks (Hunt, Hopko, Bare, Lejuez, & Robinson, 2005), and heart-rate variability (Cukic & Bates, 2015). These markers fit within the units of analysis of the RDoC approach and connote variables that should be considered largely outside of conscious control and so represent more method-neutral alternatives for arbitrating across reporting sources (Patrick et al., 2013).

5. Integrate Successful Strategies from Childhood Psychopathology Literature

Research from developmental psychopathology has a long history of informing the understanding of adult psychopathology. One well-known example is the externalizing and

internalizing spectra, which were established first in child psychiatric patients and have greatly enriched the understanding of the dimensions that cut across adult psychopathology (Achenbach, 1966; Krueger, 1999). Integrating data across multiple informants is another area where the developmental psychopathology literature has made significant strides that can be applied to the assessment of adult psychopathology. Specifically, there is a rich literature on integrating data across multiple informants that has great potential to enlighten the understanding of the client and therapist discrepancies. In contrast to the assessment of adults, which typically focuses on interviews and self-report questionnaires, the assessment of children and adolescents has long relied on the perspectives of knowledgeable informants. For example, it is routine to have adolescent diagnoses informed by ratings from a parent and/or teachers (Hunsley & Mash, 2007). A child psychologist whose diagnostic assessment included only a clinical interview of the child and his parents would be considered inadequate. Instead, clinicians request ratings from teachers or even conduct behavioral observations in the scholastic setting. This additional data is seen as fundamental for fully understanding the case. Such a robust strategy is unfortunately all too uncommon in the assessment of adult psychopathology (Klonsky et al., 2002).

Given this long-standing practice of collecting information from multiple sources, scholars in the developmental psychopathology field have taken an active interest in understanding the unique strengths of individual informants, with studies describing how their perspectives overlap and differ (De Los Reyes & Kazdin, 2005), as well as how discrepancies can themselves inform diagnosis (Laird & De Los Reyes, 2013). Whereas research on adults has largely ignored these sources of unique variance, the childhood psychopathology literature has effectively dealt with these challenges for quite some time, providing a valuable resource for determining the ideal methods of integrating information across sources.

For example, research on discrepancies across sources from the childhood literature has hypothesized several types of metrics to best probe these disagreements. For instance, Laird and De Los Reyes (2013) suggested that the interaction terms that combine child and parent data are associated with increased adolescent psychopathology. Still others have demonstrated that the standardized difference scores across sources appear to work quite well for predicting outcomes, such as youth externalizing behavior (Tackett, Herzhoff, Reardon, Smack, & Kushner, 2013). Clearly more research is needed, but the existing findings suggest that multiple indicators of these discrepancies might be a fruitful avenue to pursue for understanding differences across sources of adult psychopathology. Future research in the adult field that employs similar metrics for contextualizing the non-shared variance between the impressions of clinicians and their clients has the potential to inform a greater understanding of how to partition the non-shared variance to enhance clinical prediction.

Conclusions

Findings across a number of studies indicate that diagnostic ratings assigned by clinicians and clients show broad disagreement. These studies suggest that although agreement is improved when clinicians complete systematic instruments that assess dimensional constructs, the two sources still share only a modest amount of variance. The studies reviewed herein primarily concern personality pathology, yet the same issues are likely relevant across psychopathology. We offered a number of potential explanations for the observed diagnostic discrepancy as well as five implications for clinical science. First, we suggested that the concerns about limitations of self-report have been overstated and that clinicians' reports may incorporate similar levels of bias. Second, we highlighted the importance of utilizing homogenous dimensions, rather than heterogeneous diagnostic categories. We also called for greater integration of methodological

advances from other fields to help better understand the relative value of different sources in adult psychopathology. By utilizing these techniques, clinical science can determine how to most fruitfully integrate these perspectives and ultimately refine applied diagnosis.

References

Achenbach, T. M. (1966). Classification of Childrens Psychiatric Symptoms - a Factor-Analytic Study. *Psychological Monographs, 80*(7), 1-&.

Achenbach, T. M., Krukowski, R. A., Dumenci, L., & Ivanova, M. Y. (2005). Assessment of Adult Psychopathology: Meta-Analyses and Implications of Cross-Informant Correlations. *Psychological Bulletin, 131*(3), 361-382. doi:10.1037/0033-2909.131.3.361

American Psychiatric Association. (2016). *Practice Guidelines for the Psychiatric Evaluation of Adults* (Third Edition ed.): American Psychiatric Association.

Anderson, J. L., Sellbom, M. R., Bagby, R. M., Quilty, L. C., Veltri, C. O. C., Markon, K. E., & Krueger, R. F. (2013). On the convergence between PSY-5 domains and PID-5 domains and facets: Implications for assessment of DSM-5 personality traits. *Assessment, 20*(3), 286-294.

Bernstein, D. P., Iscan, C., & Maser, J. (2007). Opinions of personality disorder experts regarding the DSM-IV personality disorders classification system. *Journal of Personality Disorders, 21*(5), 536-551. doi:10.1521/pedi.2007.21.5.536

Block, J. (2008). *The Q-sort in character appraisal : encoding subjective impressions of persons quantitatively* (1st ed.). Washington, DC: American Psychological Association.

Butcher, J. N. (2006). Assessment in clinical psychology: A perspective on the past, present challenges, and future prospects. *Clinical Psychology-Science and Practice, 13*(3), 205-209. doi:DOI 10.1111/j.1468-2850.2006.00025.x

Carlson, E. N. (2015). Are Clinicians Asking the Right Questions? The Role of Metaperceptions as an Assessment Tool. *Clinical Psychology-Science and Practice, 22*(1), 25-28. doi:10.1111/cpsp.12091

Carlson, E. N., & Oltmanns, T. F. (2015). The Role of Metaperception in Personality Disorders: Do People with Personality Problems Know How Others Experience Their Personality? *Journal of Personality Disorders, 29*(4), 449-467.

Carlson, E. N., Vazire, S., & Oltmanns, T. F. (2011). You Probably Think This Paper's About You: Narcissists' Perceptions of Their Personality and Reputation. *Journal of Personality and Social Psychology, 101*(1), 185-201. doi:Doi 10.1037/A0023781

Carlson, E. N., Vazire, S., & Oltmanns, T. F. (2013). Self-Other Knowledge Asymmetries in Personality Pathology. *Journal of Personality, 81*(2), 155-170. doi:10.1111/j.1467-6494.2012.00794.x

Casey, B. J., Craddock, N., Cuthbert, B. N., Hyman, S. E., Lee, F. S., & Ressler, K. J. (2013). DSM-5 and RDoC: progress in psychiatry research? *Nature Reviews Neuroscience, 14*(11), 810-814. doi:10.1038/nrn3621

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., . . . Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders. *Clinical Psychological Science, 2*(2), 119-137.

Clark, L. A. (2007). Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual Review of Psychology, 58,* 227-257. doi:DOI 10.1146/annurev.psych.57.102904.190200

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-Analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*(6), 1092-1122. doi:10.1037/A0021212

Cukic, I., & Bates, T. C. (2015). The Association between Neuroticism and Heart Rate Variability Is Not Fully Explained by Cardiovascular Disease and Depression. *Plos One, 10*(5). doi:UNSP e0125882

10.1371/journal.pone.0125882

Cuthbert, B. N., & Kozak, M. J. (2013). Constructing Constructs for Psychopathology: The NIMH Research Domain Criteria. *Journal of Abnormal Psychology, 122*(3), 928-937. doi:Doi 10.1037/A0034028

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*(4), 483-509. doi:10.1037/0033-2909.131.4.483

De Los Reyes, A., Salas, S., Menzer, M. M., & Daruwala, S. E. (2013). Criterion Validity of Interpreting Scores From Multi-Informant Statistical Interactions as Measures of Informant Discrepancies in Psychological Assessments of Children and Adolescents. *Psychological Assessment, 25*(2), 509-519. doi:10.1037/a0032081

DeYoung, C. G., Hirsh, J. B., Shane, M. S., Papademetris, X., Rajeevan, N., & Gray, J. R. (2010). Testing Predictions From Personality Neuroscience: Brain Structure and the Big Five. *Psychological Science, 21*(6), 820-828. doi:Doi 10.1177/0956797610370159

Few, L. R., Miller, J. D., Rothbaum, A. O., Meller, S., Maples, J., Terry, D. P., . . . MacKillop, J. (2013). Examination of the Section III DSM-5 Diagnostic System for Personality

Disorders in an Outpatient Clinical Sample. *Journal of Abnormal Psychology, 122*(4), 1057-1069. doi:10.1037/A0034878

Fiedler, E. R., Oltmanns, T. F., & Turkheimer, E. (2004). Traits associated with personality disorders and adjustment to military life: Predictive validity of self and peer reports. *Military Medicine, 169*(3), 207-211.

Ganellen, R. J. (2007). Assessing normal and abnormality personality functioning: Strengths and weaknesses of self-report, observer, and performance-based methods. *Journal of Personality Assessment, 89*(1), 30-40.

Gore, W. L., & Widiger, T. A. (2013). The DSM-5 dimensional trait model and five-factor models of general personality. *Journal of Abnormal Psychology, 122*(3), 816-821.

Harkness, A. R., Finn, J. A., McNulty, J. L., & Shields, S. M. (2012). The Personality Psychopathology-Five (PSY-5): recent constructive replication and assessment literature review. *Psychological Assessment, 24*(2), 432-443. doi:10.1037/a0025830

Hill, K. E., Samuel, D. B., & Foti, D. (in press). Contextualizing individual differences in error monitoring: Links with impulsivity, negative affect, and conscientiousness. . *Psychophysiology*.

Hopwood, C. J., Morey, L. C., Edelen, M. O., Shea, M. T., Grilo, C. M., Sanislow, C. A., . . . Skodol, A. E. (2008). A comparison of interview and self-report methods for the assessment of borderline personality disorder criteria. *Psychological Assessment, 20*(1), 81-85. doi:10.1037/1040-3590.20.1.81

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology, 3*, 29-51. doi:10.1146/annurev.clinpsy.3.022806.091419

Hunt, M. K., Hopko, D. R., Bare, R., Lejuez, C. W., & Robinson, E. V. (2005). Construct validity of the Balloon Analog Risk Task (BART) - Associations with psychopathy and impulsivity. *Assessment, 12*(4), 416-428. doi:10.1177/1073191105278740

Huprich, S. K., & Bornstein, R. F. (2007). An overview of issues related to categorical and dimensional models of personality disorders assessment. *Journal of Personality Assessment, 89*(1), 3-15.

Huprich, S. K., Bornstein, R. F., & Schmitt, T. A. (2011). Self-Report methodology is insufficient for improving the assessment and classification of axis II personality disorders. *Journal of Personality Disorders, 25*(5), 557-570. doi:10.1521/pedi.2011.25.5.557

Kenny, D. A., & Depaulo, B. M. (1993). Do People Know-How Others View Them - an Empirical and Theoretical Account. *Psychological Bulletin, 114*(1), 145-161. doi:Doi 10.1037//0033-2909.114.1.145

Kim, N. S., & Ahn, W. K. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology-General, 131*(4), 451-476. doi:10.1037//0096-3445.131.4.451

Klein, D. N. (2003). Patients' versus informants' reports of personality disorders in predicting 7 1/2-year outcome in outpatients with depressive disorders. *Psychological Assessment, 15*(2), 216-222. doi:10.1037/1040-3590.15.2.216

Klonsky, E. D., Oltmanns, T. F., & Turkheimer, E. (2002). Informant-reports of personality disorder: Relation to self-reports and future research directions. *Clinical Psychology-Science and Practice, 9*(3), 300-311.

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin, 136*(5), 768-821. doi:Doi 10.1037/A0020327

Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry, 56*(10), 921-926. doi:DOI 10.1001/archpsyc.56.10.921

Krueger, R. F., & Markon, K. E. (2014). The Role of the DSM-5 Personality Trait Model in Moving Toward a Quantitative and Empirically Based Approach to Classifying Personality and Psychopathology. *Annu Rev Clin Psychol, 10*, 477-501. doi:10.1146/annurev-clinpsy-032813-153732

Laird, R. D., & De Los Reyes, A. (2013). Testing Informant Discrepancies as Predictors of Early Adolescent Psychopathology: Why Difference Scores Cannot Tell You What You Want to Know and How Polynomial Regression May. *Journal of Abnormal Child Psychology, 41*(1), 1-14. doi:10.1007/s10802-012-9659-y

Lenzenweger, M. F., Loranger, A. W., Korfine, L., & Neff, C. (1997). Detecting personality disorders in a nonclinical population: Application of a 2-stage for case identification. *Archives of General Psychiatry, 54*(4), 345-351.

Livesley, W. J., & Jackson, D. N. (2009). *Dimensional Assessment of Personality Pathology - Basic Questionnaire*. Port Huron, MI: Sigma Assessment Systems.

Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin, 137*(6), 1093-1093. doi:Doi 10.1037/A0025727

Markon, K. E., Quilty, L. C., Bagby, R. M., & Krueger, R. F. (2013). The development and psychometric properties of an informant-report form of the Personality Inventory for DSM-5 (PID-5). *Assessment, 20*(3), 370-383.

McAdams, D. P., & Pals, J. L. (2006). A new big five - Fundamental principles for an integrative science of personality. *American Psychologist, 61*(3), 204-217. doi:10.1037/0003-066x.61.3.204

McCrae, R. R. (2013). Exploring Trait Assessment of Samples, Persons, and Cultures. *Journal of Personality Assessment, 95*(6), 556-570. doi:10.1080/00223891.2013.821075

Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods Instruments & Computers, 33*(4), 517-523. doi:Doi 10.3758/Bf03195410

Mikton, C., & Grounds, A. (2007). Cross-cultural clinical judgment bias in personality disorder diagnosis by forensic psychiatrists in the UK: A case-vignette study. *Journal of Personality Disorders, 21*(4), 400-417. doi:DOI 10.1521/pedi.2007.21.4.400

Miller, G. (2012). The Smartphone Psychology Manifesto. *Perspectives on Psychological Science, 7*(3), 221-237. doi:10.1177/1745691612441215

Miller, J. D., Pilkonis, P. A., & Clifton, A. (2005). Self- and other-reports of traits from the five-factor model: Relations to personality disorder. *Journal of Personality Disorders, 19*(4), 400-419. doi:10.1521/pedi.2005.19.4.400

Morey, L. C., & Benson, K. T. (in press). An investigation of adherence to diagnostic criteria, revisited: Clinical diagnosis of the DSM-IV/DSM-5 section II personality disorders. . *Journal of Personality Disorders*.

Morey, L. C., & Ochoa, E. S. (1989). An investigation of adherence to diagnostic criteria: Clinical diagnosis of the DSM-III personality disorders. *Journal of Personality Disorders, 3*(3), 180-192. doi:10.1521/pedi.1989.3.3.180

Patrick, C. J., Venables, N. C., Yancey, J. R., Hicks, B. M., Nelson, L. D., & Kramer, M. D. (2013). A construct-network approach to bridging diagnostic and physiological domains: application to assessment of externalizing psychopathology. *Journal of Abnormal Psychology, 122*(3), 902-916. doi:10.1037/a0032807

Ray, J. V., Hall, J., Rivera-Hudson, N., Poythress, N. G., Lilienfeld, S. O., & Morano, M. (2013). The Relation Between Self-Reported Psychopathic Traits and Distorted Response Styles: A Meta-Analytic Review. *Personality Disorders-Theory Research and Treatment, 4*(1), 1-14. doi:Doi 10.1037/A0026482

Rettew, D. C., Lynch, A. D., Achenbach, T. M., Dumenci, L., & Ivanova, M. Y. (2009). Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research, 18*(3), 169-184. doi:10.1002/mpr.289

Rounsaville, B. J., Alarcon, R. D., Andrews, G., Jackson, J. S., Kendell, R. E., & Kendler, K. (2002). Basic nomenclature issues for DSM-V. In D. J. Kupfer, M. B. First, & D. A. Regier (Eds.), *A Research Agenda for DSM-V.* (pp. 1-29). Washington, DC: American Psychiatric Association.

Samuel, D. B. (2015). A Review of the Agreement Between Clinicians' Personality Disorder Diagnoses and Those From Other Methods and Sources. *Clinical Psychology-Science and Practice, 22*(1), 1-19. doi:Doi 10.1111/Cpsp.12088

Samuel, D. B., & Griffin, S. A. (2015). A Critical Evaluation of Retaining Personality Categories and Types. In S. K. Huprich (Ed.), *Personalitiy Disorders: Toward Theoretical and Empirical Integration of Diagnosis and Assessment*. Washington, DC: American Psychological Association.

Samuel, D. B., Sanislow, C. A., Hopwood, C. J., Shea, M. T., Skodol, A. E., Morey, L. C., . . . Grilo, C. M. (2013). Convergent and Incremental Predictive Validity of Clinician, Self-Report, and Structured Interview Diagnoses for Personality Disorders Over 5 Years. *Journal of Consulting and Clinical Psychology, 81*(4), 650-659. doi:Doi 10.1037/A0032813

Samuel, D. B., & Widiger, T. A. (2008). A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: A facet level analysis. *Clinical Psychology Review, 28*(8), 1326-1342. doi:10.1016/j.cpr.2008.07.002

Samuel, D. B., & Widiger, T. A. (2009). Comparative gender biases in models of personality disorder. *Personality and Mental Health, 3*(1), 12-25. doi:10.1002/pmh.61

Samuel, D. B., & Widiger, T. A. (2010). Comparing personality disorder models: Cross-method assessment of the FFM and DSM-IV-TR. *Journal of Personality Disorders, 24*(6), 721-745. doi:10.1521/pedi.2010.24.6.721

Sanislow, C. A., Pine, D. S., Quinn, K. J., Kozak, M. J., Garvey, M. A., Heinssen, R. K., . . . Cuthbert, B. N. (2010). Developing constructs for psychopathology research: Research domain criteria. *Journal of Abnormal Psychology, 119*(4), 631-639. doi:Doi 10.1037/A0020909

Shedler, J. (2015). Integrating clinical and empirical perspectives on personality: The Shedler-Westen Assessment Procedure (SWAP). In S. K. Huprich (Ed.), *Personality Disorders:*

Toward Theoretical and Empirical Integration in Diagnosis and Assessment (pp. 225-252). Washington, DC: American Psychologial Association.

Shedler, J., Beck, A. T., Fonagy, P., Gabbard, G. O., Kernberg, O., Michels, R., & Westen, D. (2011). Revision of the Personality Disorder Model for DSM-5 Response. *American Journal of Psychiatry, 168*(1), 97-98. doi:10.1176/appi.ajp.2010.10101466r

Simms, L. J. (2008). Classical and modern methods of psychological scale construction. *Social and Personality Psychology Compass, 2*(1), 414-433.

Skodol, A. E. (2014). Personality disorder classification: stuck in neutral, how to move forward? *Curr Psychiatry Rep, 16*(10), 480. doi:10.1007/s11920-014-0480-x

Skodol, A. E., Morey, L. C., Bender, D. S., & Oldham, J. M. (2013). The Ironic Fate of the Personality Disorders in DSM-5. *Personality Disorders-Theory Research and Treatment, 4*(4), 342-349. doi:Doi 10.1037/Per0000029

Smith, G. T., McCarthy, D. M., & Zapolski, T. C. B. (2009). On the value of homogeneous constructs for construct validation, theory testing, and the description of psychopathology. *Psychological Assessment, 21*(3), 272-284. doi:Doi 10.1037/A0016699

South, S. C., Oltmanns, T. F., Johnson, J., & Turkheimer, E. (2011). Level of Agreement Between Self and Spouse in the Assessment of Personality Pathology. *Assessment, 18*(2), 217-226. doi:Doi 10.1177/1073191110394772

Suzuki, T., Griffin, S. A., & Samuel, D. B. (in press). Capturing the DSM-5 alternative personality disorder model traits in the Five-Factor Model's nomogoical net. *Journal of Personality*.

Suzuki, T., Samuel, D. B., Pahlen, S., & Krueger, R. K. (2015). DSM-5 Alternative Personality Disorder Model Traits as Maladaptive Extreme Variants of the Five-Factor Model: An Item-Response Theory Analysis. *Journal of Abnormal Psychology, 124*(2), 343-354.

Tackett, J. L., Herzhoff, K., Reardon, K. W., Smack, A. J., & Kushner, S. C. (2013). The Relevance of Informant Discrepancies for the Assessment of Adolescent Personality Pathology. *Clinical Psychology-Science and Practice, 20*(4), 378-392. doi:10.1111/cpsp.12048

Tromp, N. B., & Koot, H. M. (2010). Self- and Parent Report of Adolescent Personality Pathology: Informant Agreement and Relations to Dysfunction. *Journal of Personality Disorders, 24*(2), 151-170.

Trull, T. J., & Durrett, C. A. (2005). Categorical and dimensional models of personality disorder. *Annual Review of Clinical Psychology, 1*, 355-380. doi:10.1146/annurev.clinpsy.1.102803.144009

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory Assessment. *Annual Review of Clinical Psychology, Vol 9, 9*, 151-176. doi:10.1146/annurev-clinpsy-050212-185510

Vazire, S. (2010). Who Knows What About a Person? The Self-Other Knowledge Asymmetry (SOKA) Model. *Journal of Personality and Social Psychology, 98*(2), 281-300. doi:Doi 10.1037/A0017908

Vazire, S., & Mehl, M. R. (2008). Knowing Me, Knowing You: The Accuracy and Unique Predictive Validity of Self-Ratings and Other-Ratings of Daily Behavior. *Journal of Personality and Social Psychology, 95*(5), 1202-1216. doi:10.1037/a0013314

Watts, A. L., Lilienfeld, S. O., Edens, J. F., Douglas, K. S., Skeem, J. L., Verschuere, B., & LoPilato, A. C. (2016). Does Response Distortion Statistically Affect the Relations

Between Self-Report Psychopathy Measures and External Criteria? *Psychological Assessment, 28*(3), 294-306. doi:10.1037/pas0000168

Westen, D. (1997). Divergences between clinical and research methods for assessing personality disorders: Implications for research and the evolution of axis II. *American Journal of Psychiatry, 154*(7), 895-903.

Westen, D., DeFife, J. A., Bradley, B., & Hilsenroth, M. J. (2010). Prototype personality diagnosis in clinical practice: A viable alternative for DSM–5 and ICD–11. *Professional Psychology: Research and Practice, 41*(6), 482-487. doi:10.1037/a0021555

Westen, D., & Weinberger, J. (2004). When clinical description becomes statistical prediction. *American Psychologist, 59*(7), 595-613. doi:10.1037/0003-066X.59.7.595

Widiger, T. A., & Mullins-Sweatt, S. N. (2009). Five-Factor Model of Personality Disorder: A Proposal for DSM-V. *Annual Review of Clinical Psychology, 5*, 197-220. doi:10.1146/annurev.clinpsy.032408.153542

Widiger, T. A., & Samuel, D. B. (2005a). Diagnostic categories or dimensions? A question for the diagnostic and statistical manual of mental disorders-fifth edition. *Journal of Abnormal Psychology, 114*(4), 494-504. doi:10.1037/0021-843X.114.4.494

Widiger, T. A., & Samuel, D. B. (2005b). Evidence-based assessment of personality disorders. *Psychological Assessment, 17*(3), 278-287. doi:10.1037/1040-3590.17.3.278

Widiger, T. A., & Simonsen, E. (2005). Alternative dimensional models of personality disorder: Finding a common ground. *Journal of Personality Disorders, 19*(2), 110-130.

Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder - Shifting to a dimensional model. *American Psychologist, 62*(2), 71-83.

Wood, J. M., Garb, H. N., Nezworski, M. T., & Koren, D. (2007). The Shedler-Westen Assessment Procedure-200 as a basis for modifying DSM personality disorder categories. *Journal of Abnormal Psychology, 116*(4), 823-836. doi:10.1037/0021-843x.116.4.823

Wright, A. G. C., Krueger, R. F., Hobbs, M. J., Markon, K. E., Eaton, N. R., & Slade, T. (2013). The Structure of Psychopathology: Toward an Expanded Quantitative Empirical Model. *Journal of Abnormal Psychology, 122*(1), 281-294. doi:Doi 10.1037/A0030133

Wright, A. G. C., & Simms, L. J. (2014). On the structure of personality disorder traits: Conjoint analyses of the CAT-PD, PID-5, and NEO-PI-3 trait models. *Personality Disorders: Theory, Research, and Treatment, 5*(1), 43-54. doi:10.1037/per0000037

Wright, C. V., Beattie, S. G., Galper, D. I., Church, A. S., Bufka, L. F., Brabender, V. M., & Smith, B. L. (in press). Assessment practices of professional psychologists: Results of a National Survey. *Professional Psychology: Research and Practice*.