When and how assessment matters: An update on the Treatment Utility of Clinical Assessment (TUCA)

Kamphuis, J.H.; Noordhof, A.; Hopwood, C.J.

DOI
10.1037/pas0000966

Publication date
2021

Document Version
Final published version

Published in
Psychological Assessment

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Assessment

ISSN: 1040-3590

2021, Vol. 33, No. 2, 122–132
http://dx.doi.org/10.1037/pas0000966

When and How Assessment Matters: An Update on the Treatment Utility of Clinical Assessment (TUCA)

Jan H. Kamphuis and Arjen Noordhof
University of Amsterdam

Christopher J. Hopwood
University of California, Davis

The Treatment Utility of Clinical Assessment (TUCA) has long been a controversial topic, with arguably more (strong) opinions than relevant, well-designed empirical research. We argue that this question has been tackled too broadly and that a more contextualized approach would likely be more informative. Instead of asking "what is the treatment utility of assessment," we suggest specifying and examining more closely the conditions by which assessment can—or cannot—contribute to treatment process and ultimately patient benefit. To this end, we present a heuristic model for conceptualizing the conditions under which clinical assessment may have treatment utility and illustrate its use by distinguishing four specific classes of assessment-driven interventions. We distinguish direct benefits from assessment from indirect TUCA as two principal pathways, emphasize the importance of having some a priori theory regarding working mechanisms, and stress the requirements of ensuring adequate variability of the presumed mediating variables. These considerations in turn argue for a broader view of pertinent outcome measures, the use of more powerful designs in TUCA research, and the implementation of some form of stepped assessment in clinical practice.

Public Significance Statement
The extent to which clinical assessment enhances subsequent treatment benefits for patients has long been controversial, and we suggest specifying and examining more closely the conditions by which assessment can—or cannot—contribute to treatment process and ultimately patient benefit. We argue that this can contribute to more informative research on this very important question and provide illustrative examples from clinical assessment practice.

Keywords: clinical assessment, treatment utility, mediation, research design

The treatment utility of clinical assessment (TUCA[1]) has long been a controversial topic, with arguably more (strong) opinions than relevant, well-designed empirical research (Hunsley & Mash, 2007; Meehl, 1959). A noteworthy recent addition to this body of evidence was provided by Olsson and Fridell (2018). These researchers reported on the 5-year cost and benefit of extended psychological and psychiatric assessment versus standard intake interview for women with comorbid substance use disorders treated in compulsory care in Sweden. In several respects this study is exemplary in its efforts to document the TUCA: It used a manipulated assessment design (albeit naturalistic quasi-experimental) and assessed long-term (i.e. five year) outcomes that have clear societal relevance in a large sample of patients (N = 227). The evidence did not support the clinical utility of assessment for improving treatment outcomes. Olsson and Fridell's (2018) negative findings are consistent with a number of previous studies that reported negative results regarding TUCA (e.g., Schulte, Kunzel, Pepping, & Schulte-Bahrenberg, 1992) but at odds with other studies that reported more favorable results (e.g., De Saeger et al., 2014; Lambert et al., 2003; Poston & Hanson, 2010). Such inconsistent findings suggest that some applications or forms of assessment do have demonstrable clinical utility, whereas others do not.

Importantly, the detectability of TUCA may also vary across different research designs. In a classic review Hayes, Nelson, and Jarrett (1987) offered a methodological typology for TUCA research. Most current research falls in the category of obtained differences: Patients differing in characteristics are post hoc compared with regard to treatment outcomes. A prospective design is the manipulated assessment approach in which patients are randomly assigned to alternative assessment procedures, like in the

This article was published Online First October 29, 2020.

Jan H. Kamphuis and Arjen Noordhof, Department of Psychology, University of Amsterdam; Christopher J. Hopwood, Department of Psychology, University of Californa, Davis.

Correspondence concerning this article should be addressed to Jan H. Kamphuis, Department of Psychology, University of Amsterdam, Postbus 15933, 1001 NK Amsterdam, the Netherlands. E-mail: j.h.kamphuis@uva.nl

[1] The current paper only discusses treatment utility, i.e., effects on outcomes. Assessment may have other important aims—e.g., administration, evaluation, or research—but these are beyond the scope of the current paper.

Olsson and Fridell (2018) study. An interesting alternative is the manipulated use of assessment, which allows researchers to specifically test the utility of obtained information (rather than general effects of the assessment process). In this design therapists get access to different information from an assessment procedure that is held equal between patients. TUCA research has further accumulated since the publication of the Hayes et al. review (1987), albeit mostly of the obtained differences type (see e.g., the review by Hunsley & Mash, 2007).

In this article, we provide a selective review of critical considerations in pertinent research, which collectively indicate that TUCA is a complex issue that is not well captured by an all-or-nothing approach. We argue that greater progress will be possible in evaluating questions related to TUCA if these complexities are better specified. We aim to move from the broad question—what, if any, is the pragmatic advantage of a personality assessment being known in advance by the therapist? (Meehl, 1959)—to a more specific one: What are the conditions under which and the mechanisms by which assessment does—or does not—contribute to treatment and beneficial outcomes? Our aims were (a) to highlight methodological issues that are critical but often overlooked in studies on TUCA and (b) to illustrate potential mechanisms and research designs for four assessment-driven-interventions (ADI) with significant promise for treatment utility. As our study presents extant research and does not involve new data collection, it was exempt from ethical review.

Five Methodological Issues in TUCA Research

The *when* and *how* of TUCA concern interrelated issues of moderation and mediation. Hypothesized processes or mechanisms by which an assessment is considered to yield positive outcomes (mediation) generally point toward the conditions and samples for which utility can be expected (moderation). This will become more evident in the specific research examples provided below. Conversely, considerations of key individual differences within certain samples or between certain clinical settings (moderation) can inspire development of assessment processes that will be able to capitalize on these differences to improve treatment. In the latter case, considerations of moderation yield hypotheses about possible mechanisms. Hence, in the following we will discuss issues of moderation and mediation intertwined, sometimes reasoning from mechanisms to conditions under which these mechanisms will likely work best and at other times reasoning from certain conditions toward the kind of processes that would produce positive results. With this as a preamble, we will now present five propositions that we believe will further subsequent TUCA related research.

Our first proposition is that it is useful to distinguish direct from indirect effects of clinical assessment. As shown in Figure 1, *direct effects* of clinical assessment (path a → d → e) include all mechanisms (e.g., remoralization, self-understanding) by which the assessment process itself has clinical benefits (e.g., Poston & Hanson, 2010). *Indirect effects* of clinical assessment (path a → b → c → e) occur along the pathways by which the assessment process and resulting data could be used to enhance subsequent effective clinical care. Such pathways might include the identification of key problem areas, establishing rapport, enhancing motivation, and treatment selection.

The distinction between direct and indirect effects, as depicted in Figure 1, has several implications for TUCA research. First, target outcomes and mediators should be carefully measured and distinguished and should also differ enough across cases to leave reliable variance to be explained. Second, the effects of the assessment on mediators should be examined separately from the effects of assessment on target outcomes. Third, to test for mediation, the effects of mediators on outcomes should be evaluated, net of the variance explained by assessment. Finally, studies should gather data longitudinally with sufficient appropriately timed assessment waves to allow for directional tests of mediation processes. More specifically, the timing of the assessments should be informed to the degree possible by a well-specified theory of the process being tested, and the data should be adequate for testing given (e.g., mediation) statistical models (see also Collins, 2006, or Luhmann, Orth, Specht, Kandler, & Lucas, 2014, for this complex issue). It is not possible to disentangle direct from indirect effects with a prepost study design. Later in this article, we will illustrate how these conditions are critical by reexamining some studies that have been presented as evidence for a generalized lack of TUCA.

Our second proposition is that research would profit from theoretical consideration of the potential substantive mechanisms (paths b, c and d) by which the supposed direct or indirect effects come about. In our view this will result in research designs with a higher a priori chance of finding TUCA than blind empiricism. For example, if an assessment procedure is hypothesized to have a direct effect (path d) by improving insight into complex person-

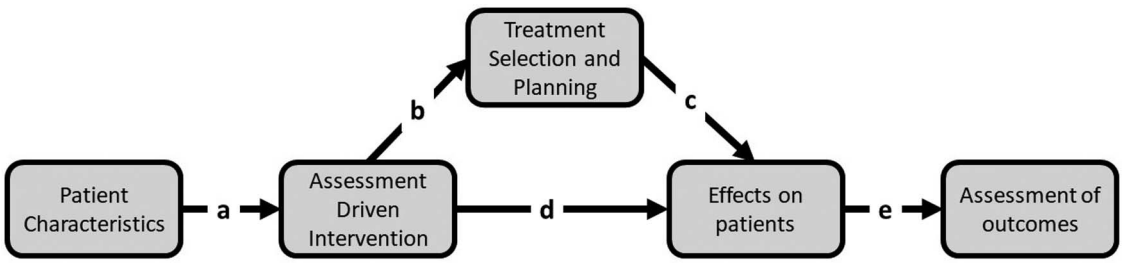

Figure 1. Schematic depiction of direct and indirect Treatment Utility of Clinical Assessment (TUCA). Each path indicates an important consideration for research designs: (a) Does assessment capture treatment relevant individual variance, (b) does assessment affect treatment and (c) thereby indirectly affect patients, or (d) does assessment directly affect patients, and (e) which outcomes are most relevant to assess?

ality dynamics, a potent research design might include patients who have rather severe personality problems, who are seen by experienced personality assessors, and in which the emergence of new insights was measured directly. To be clear, we are not suggesting that each TUCA study should stringently test by which mechanisms the outcomes are influenced by the assessment. We do however believe that it is very useful for improving research designs to think through the possible mechanisms by which an assessment is thought to improve outcomes.

Our third proposition is that there must be sufficient variability within samples and across treatments, outcomes, and potential mediators or moderators to adequately examine TUCA. Participating patients should have sufficient treatment-relevant heterogeneity (path a) to allow for predictions of target outcomes and potential mediators. Study designs in which patients are selected on narrow inclusion and exclusion criteria may be underpowered for TUCA research, due to the reduction of clinically informative variability in homogeneous samples. The classic MATCH study (Project MATCH Research Group, 1997) may be a case in point in this regard.[2] Mediators, such as interventions that are selected based on assessment data, also need to vary meaningfully (path b and c). One general finding from intervention research is that many treatments have relatively similar effects and may even work by the same mechanisms, even if they are based on different theories of change (e.g., A Tjak, Morina, Topper, & Emmelkamp, 2020). Assessment data cannot reasonably be expected to help clinicians effectively choose between two treatments that work about the same, and for the same reasons. Again, arguably, this was the case in the MATCH study. Moreover, treatment outcomes need to vary across clients to have a chance of being predicted by assessment or other techniques. Finally, and related to variability in each of the paths in Figure 1, there must be a plausible opportunity for assessment results to influence treatment (selection). For example, an assessment procedure might generally be rather adequate, but if the clinical setting is such that the influence of assessment on treatment is minimal, for example because only very few distinctive treatments are available, then one cannot expect to enhance outcomes based on assessment.

Our fourth proposition is that it is crucial to carefully consider which outcomes can be expected to be obtained at which moment in the therapeutic process (path e). In most empirical studies, TUCA has been indexed by (short-term) symptomatic improvement (e.g., Lima et al., 2005; Schulte et al., 1992), but many other outcomes may be considered such as relational and/or societal functioning (Olsson & Fridell, 2018), more distal outcomes like economic cost (e.g., health care consumption, relative productivity loss; Olsson & Fridell, 2018), more proximal/intermediate outcomes like various indicators of treatment readiness (e.g., De Saeger et al., 2014), or more personal outcomes like acceptance, insight, and meaning (De Saeger, Bartak, Eder, & Kamphuis, 2016; van Os, Guloksuz, Vijn, Hafkenscheid, & Delespaul, 2019). Another crucial outcome that we believe needs to be taken into consideration is drop-out. We therefore advocate for a more inclusive view of treatment utility (as did several prominent psychotherapy and assessment researchers, e.g., Bram, 2013; McWilliams, 2005).

Of note, several psychotherapy models describe necessary intermediate steps before ultimate outcomes in terms of symptom improvement and patient functioning can be achieved. One influential model is the Stages of Change model (Norcross, Krebs, & Prochaska, 2011). The central hypothesis advanced by the Stages of Change model is that the effectiveness of an intervention depends on the stage of change a particular client is currently in. Five subsequent stages are distinguished: precontemplation, contemplation, preparation, action, and maintenance. Interventions should be matched to stages; for example, in the contemplation stage raising awareness of a certain problem may be a useful intervention, while the same intervention may not be particularly effective in moving toward action. Conversely, a focus on actions to be taken does not match with a patient who is still in a precontemplation stage. The relevance of this for TUCA is that the transition of one stage to the next can be assessed using specific instruments (see Norcross, Krebs, et al., 2011), and hence it can be tested whether specific assessment interventions contribute to improving the probability of moving from one stage to the next. Down the line, such intermediate progress may well result in the prevention of drop-out or continuation of ineffective treatment.

The fifth proposition is that the ultimate aim of replicability ought to be at the heart of designing TUCA-studies. Treatment planning and selection decisions are potentially of great consequence to patients, and clinical implementation of research findings should therefore be based on replicated findings. Careful consideration of Type I and II errors is at the heart of this pursuit. The probability of Type I errors (i.e. false positives) can be minimized by adopting sufficiently stringent significance testing and proper correction for multiple testing. Chances for type II errors (i.e. false negatives) can be reduced by securing sufficient power.[3] Obvious ways to improve power are to increase the sample size and/or to increase number of repeated measures. Less obvious perhaps, power can also be improved by maximizing the probable effect-size of an intervention by careful consideration of context aspects. Below, we provide several specific examples of how this may be attempted (e.g., by using proximal outcomes, specific design choices in terms of setting and sampling). Furthermore, power can be improved by reducing the complexity of hypotheses to be tested. Complex interactions of multiple moder-

[2] Project MATCH (Matching Alcohol Treatment to Client Heterogeneity) is the largest effort to date to match patients to treatments based on pretreatment patient characteristics. The objective was to identify subgroups of substance use disorder patients who would respond differentially to three protocol-driven individual treatments of equal duration; i.e. 12-step facilitation therapy, cognitive behavioral coping skills therapy, and motivational enhancement therapy. Matching variables were theoretically plausible factors such as the severity of addiction or presence or absence of antisocial features. Minimal support was found for the favorable effects of matching clients to treatments: No supportive evidence was found for 15 out of 16 predicted matching associations, and each of the treatments performed about equally well. Accordingly, it was concluded that—with the possible exception of psychiatric severity—client characteristics need not be taken into account when assigning clients to each of these three treatment protocols.

[3] Alternatively, type I and type II errors can be properly weighted a priori as well as a posteriori by adopting a Bayesian approach that indicates the amount of evidence for as well as against TUCA (Wagenmakers et al., 2012). In this approach, small samples will simply appear to provide insufficient evidence and hence be interpreted as inconclusive rather than as negative evidence or non-replications. There are important advantages of using a Bayesian approach, especially in the area of application of research to clinical practice, but a more thorough discussion of this topic is beyond the scope of this article.

ators predicting different mechanisms, trajectories, and outcomes are often readily conceivable in theory but unlikely to be testable with statistical rigor. Considerations for minimally adequate sample sizes for multivariate moderation effects have been examined by several research groups (see e.g., Cohen & DeRubeis, 2018; Luedtke, Sadikova, & Kessler, 2019 for simulation analyses), and the overall conclusion is that robust effects require sample sizes well beyond those typically employed in TUCA research or RCTs in general (greater equal than 100 patients per study arm for large effects, or well beyond that for the detection of smaller interaction effects). Power is particularly complicated for tests of moderation, but informative issues and guidelines have been presented elsewhere (see e.g., Memon et al., 2019; Preacher & Sterba, 2019; Zedeck, 1971).

A creative study on TUCA conducted by Schulte et al. (1992) might serve as a brief illustration of the importance of several of the methodological propositions just described. In an experimental study, patients ($N = 90$) who met diagnostic criteria for specific phobia but no other (comorbid) diagnoses were randomly allocated to (a) standardized Evidence-Based Treatment (EBT) consisting of exposure techniques along with self-verbalizations, (b) tailored treatments, based on the patient's individualized behavioral analysis, or (c) individualized treatment according to the behavioral analysis of another patient. The best results were observed for patients who received the standard EBT, and less favorable, equivalent results were seen for patients who received treatment according to individually tailored behavioral analyses, be it based on their own history of presenting complaints or that of another patient. These results have sometimes been presented as general evidence against TUCA. However, this extended assessment scenario was a priori extremely unlikely to produce extra client benefit. Given that clients were selected with a single psychiatric disorder that is relatively well understood in terms of the underlying pathological mechanisms, and for which a highly effective EBT is available, assessment-driven tailoring of treatment was most likely to deviate from the EBT format and thus accrue suboptimal results.[4]

In the remainder of this paper we will present in more detail how TUCA-related research efforts have been conducted in four prevailing approaches to assessment: (a) actuarial assessment-driven interventions, (b) diagnostic assessment, (c) clinical personality assessment, and (d) collaborative and therapeutic approaches. Collectively, these classes cover most TUCA pertinent research. We will also suggest an agenda for each of these approaches and discuss an illustrative study that might have profited from (some of) the methodological considerations summarized in our Figure 1. In so doing, we hope to show the heuristic value of our model for enhancing research designs by carefully considering the conditions under which and the mechanisms by which treatment utility can be obtained.

Actuarial Assessment

Clinical assessment research can serve TUCA by identifying personality and clinical characteristics that predict differential treatment outcome across patients. When such a variable predicts treatment outcome irrespective of treatment condition (say in the context of a randomized clinical trial, or in a single treatment design), the variable is called a *prognostic factor*; when it predicts differential treatment outcome contingent upon type of treatment, is called a *prescriptive factor*. For example, presence of a comorbid cluster C personality disorder was shown to be a (modest) negative *prognostic* predictor for cognitive behavior therapy (CBT) outcomes in patients with panic disorder (Telch, Kamphuis, & Schmidt, 2011); as an example of a prescriptive factor, consider Blatt's classic work on subtypes of depression, in which he hypothesized that anaclitic patients profit more from psychotherapy, whereas introjective patients profit more from psychoanalysis (Blatt, 1992). Statistically, such moderating variables present as interaction effects in prediction studies; that is, outcome for a certain treatment is shown to significantly depend on the level of some other (premorbid) variable in the patient. As such, clinical application of prescriptive factors is more straightforward than for prognostic factors, as presumably the former can be of use to select optimal treatment conditions. Prognostic factors are often more difficult to link to specific ways of intervening.

A pitfall for research targeting obtained differences is that these effects often do not replicate well, probably due to several (cumulative) design limitations including sampling characteristics, Type I error, and multicollinearity (Bohart & Wade, 2013; Zilcha-Mano, 2019). Moreover, both univariate and multivariate designs are fraught with inherent limitations and challenges. With respect to univariate analyses, it is unlikely that one single moderator can by itself adequately answer Gordon Paul's (1967) "what works for whom" question (considering all other things being equal; *ceteris paribus*). However, when multiple prescriptive variables are identified, guidance on how to combine and integrate such information into choice of preferred treatment is often arbitrary, in absence of a formal multivariate model that can dictate the treatment selection.

Several lines of research follow an actuarial approach rationale. First, clinical screening is often a standard component of clinical practice. Provided sufficient sensitivity, such questionnaires (e.g., PHQ9 for depression; Kroenke & Spitzer, 2002) can be useful in efficiently selecting the right patients for (scarce) treatment resources. TUCA, perhaps especially at the cost-effectiveness level, may well exist relative to formal comprehensive assessments across all patients; studies adhering to a manipulated assessment design (Hayes et al., 1987) could answer this question. Furthermore, brief scales measuring broadband domains of psychopathology (e.g., internalizing, externalizing, psychotic problems and subdomains) might be valuable in cost-efficiently signaling co-occurring areas of symptomatology or personality problems that may remain undetected in informal intakes. Thus, an important line of future research might involve the incremental treatment utility of relatively short and cost-effective scales in settings with low assessment resources (e.g., counseling). One specifically interesting topic would be to what extent patients for whom more extensive assessment would be useful can be distinguished from those who for whom this would be unnecessary (i.e. stepped assessment). In similar vein, in forensic assessment, use of standardized instruments such as the PCL-R to predict recidivism is common practice, although clinical judgment is also still widely used to make appraisals, either exclusively or to modify actuarial

[4] In fairness, Schulte (1996) later published a thoughtful reflection on the complementary tasks of planning standardized and tailor-made behavior therapy.

inputs (Smid, Kamphuis, de Wever, & Van Beek, 2013, 2014; Hanson & Morton-Bourgon, 2005). The public health and safety impact of TUCA would be demonstrated if actuarial prediction yields lower recidivism rates than clinical judgment (only) assessment.

A more complex but highly successful and influential assessment-driven intervention paradigm relies on actuarial tracking of longitudinal outcome data, rather than on premorbid patient characteristics. In patient focused research (Lambert, Hansen, & Finch, 2001), therapists compare the weekly progress of the individual patient on a standardized outcome measure to normative progress/recovery curves of the patient population they are a member of. When a particular patient lags in progress relative to the successful completers, the therapist (and in some applications the patient as well) receives a warning that treatment planning needs modification. Several candidate mechanisms have been proposed to target profitable treatment modification (path b). Specifically, clinical support tools (Whipple et al., 2003) have been developed to guide the therapist in determining critical domains for inspection, such as the quality of the treatment relationship, agreement on goals, and the nature of the case-conceptualization.

Finally, some recently developed statistical models have shown great promise for TUCA in their potential to predict treatment response and in turn offer individualized treatment recommendations. For example, DeRubeis and colleagues (Cohen & DeRubeis, 2018; DeRubeis et al., 2014) developed the personalized advantage index as an instrument geared for personalized medicine. Applying this technology to the treatment selection of patients with major depressive disorder, they used data from a randomized controlled trial comparing the efficacy of cognitive behavior therapy versus antidepressant medication. Sampling from a large pool of potential prospective (which patients do better/worse, irrespective of treatment condition) and predictive variables (which patients do better in a specific treatment arm), prediction formulas of response were statistically determined for each treatment. Next, each patient received a score that reflected their outcome for each treatment (i.e. derived from the respective formula). These scores were combined into a personalized advantage index, the value of which points to the "optimal" treatment. Post hoc simulations (DeRubeis et al., 2014) suggested that overall treatment outcome may be enhanced by a moderate effect-size when individuals are allocated according to this index score (relative to treatment selection as usual). A manipulated assessment design (Hayes et al., 1987) may prospectively test the TUCA of this method, for example by comparing treatment allocation according to patient preference versus treatment allocation according to personalized advantage index informed shared decision making.

Actuarial Assessment Research Example: Prescriptive Psychotherapy

Over the past decades, an extensive literature has emerged concerning replicable interactions between patient characteristics and treatment outcomes, which is reflected in efforts toward a prescriptive psychotherapy (e.g., Beutler & Clarkin, 1990; Beutler & Harwood, 2000) in which patient characteristics are used to direct treatment considerations for better therapist-client-therapy fit. A particularly good example of this approach is a recent meta-analysis by Beutler, Kimpara, Edwards, and Miller (2018) showing a medium effect size ($d = .60$) for interactions of treatment (insight-focus vs. symptom-focus) by patient characteristics (externalizing vs. internalizing coping styles). Such a large effect size is impressive when compared to the magnitude of (often nonsignificant) effect sizes of comparisons between different treatment protocols. The 18 included studies demonstrated good replicability of results. Such robustness of findings has been obtained for a number of variables in obtained difference studies (see a review by Norcross & Wampold, 2011a). Furthermore, the association between patient characteristics, treatment types, and outcomes points to a pathway for using assessment results to improve treatment (path a $\rightarrow$ b $\rightarrow$ c $\rightarrow$ e). That noted, the observed associations were not tested a priori, so there is an a-b-c-e connection, but not a clear a $\rightarrow$ b $\rightarrow$ c $\rightarrow$ e causal pathway. To establish a causal pathway, a manipulated assessment design as described by Hayes et al. (1987) is required. Given that the associations (a-b-c-e) have already been well-researched and are grounded in some theoretical considerations, these findings present themselves as likely candidates for TUCA studies.

What considerations might be useful in designing such a prospective TUCA study? First, sufficient heterogeneity in internalizing and externalizing coping styles is essential (path a), as well as the availability of an accurate assessment to distinguish between the two. Second, one needs to ensure that clinicians are capable and willing to use the advice in order to adapt their therapy (path b) in the experimental condition, but not in the control condition (absence of path a $\rightarrow$ b). To set specific hypotheses, it should be noted that it is a priori unlikely that the effect sizes found in obtained difference studies (mean $d = .60$) generalize to a prospective design. For the assessment and prescriptive advice to have incremental utility, the clinicians need to make sufficient errors of judgment in comparison to actuarial assignment (i.e. *mismatch* on arrow a). Although it is well known that in general actuarial approaches tend to outperform clinical judgment in terms of reliable predictions (Meehl, 1954), it is not known whether the actual difference in accuracy (in this case) is sufficiently large and consequential that it would also result in substantial improvement in clinical outcomes.

Diagnostic Assessment

Probably the most common assessment goal is to derive categorical diagnoses, typically in the rubric of the diagnostic and statistical manual of mental disorders (*DSM–5*; American Psychiatric Association, 2013) or international classification of diseases (ICD 11; World Health Organization, 2018). Aside from extratherapeutic aims (e.g., reimbursement or administrative purposes), the presumed utility is indirect in that proper diagnosis is hypothesized to result in proper treatment-selection (i.e. path b), ideally rooted in evidence from randomized controlled intervention trials observing those same DSM-based inclusion and exclusion criteria. Consequently, the utility of DSM assessment may appear so obvious that it hardly needs empirical demonstration. Indeed, the vast majority of research on psychiatric diagnosis is focused on testing and improving classificatory accuracy rather than evaluating and improving clinical utility. Given the far-reaching consequences of DSM diagnoses—not only in terms of treatment, but also regarding social status, identity, stigma, education, insurance, legal issues, and so forth—improving diagnostic accuracy is valuable in

and of itself, but establishing TUCA would be based on the impact diagnoses have on treatment course down the line (i.e. paths b and c). Empirical research testing the extent to which accurately diagnosing DSM syndromes contributes to treatment outcomes is surprisingly scant, and virtually no research addresses how such effects are obtained. For example, to our knowledge no published research can document whether introducing structured clinical interviews in clinical practice improves treatment outcome in comparison with unstructured interviews.[5]

In fact, it is not known and has often been questioned (van Os et al., 2019; Hopwood et al., in press) whether DSM-syndromes provide optimal criteria for treatment planning. It is not so easy to rigorously test whether DSM-distinctions do or do not contribute to treatment outcome. Paradoxically, the very popularity of the approach, as reflected in the dominance of the DSM-paradigm within treatment research, is an obstacle against testing whether it enhances clinical practice. One promising approach to testing the TUCA of DSM syndromes cuts against the grain of its hegemony by comparing it to alternative methods for treatment selection such as the *DSM–5* alternative model for personality disorders (AMPD; American Psychiatric Association, 2013; see also Weekers, Hutsebaut, Bach, & Kamphuis, 2020, for a multimethod individual assessment protocol) and the hierarchical taxonomy of psychopathology (HiToP; Hopwood, 2018; Hopwood et al., in press; Mullins-Sweatt et al., 2020).

Diagnostic Assessment Research Example: Levels of Information

We suggest it may be time to change the focus of research from broad questions about whether or not to diagnose to more specific questions about (a) how specific diagnoses should be and (b) which diagnostic distinctions are most valuable for treatment. Both the AMPD and HiToP models use broad, "higher-order," domains to account for the high levels of overlap found between DSM disorder categories and employ comprehensive dimensional models of individual differences to help explain the heterogeneity commonly found within diagnostic categories.[7] The use of such alternative models may increase power to identify TUCA by allowing for finer grained and theoretically embedded tests of incremental clinical utility. First, using a comprehensive dimensional system can help circumvent problems associated with narrow diagnostic inclusion and exclusion criteria. Second, the availability of different dimensions at different levels of abstraction ensures potentially informative sources of patient variability in the assessment data. Third, evidence-based dimensions may provide researchers with greater power and sensitivity relative to psychometrically problematic diagnostic categories.

Most pertinent to the topic of this paper, hierarchical dimensional structures stimulate different, more specific TUCA-related research questions. Indeed, there are a wide range of clinically promising but largely untested hypotheses about how evidence-based dimensions of psychopathology could inform and enhance interventions and clinical outcomes. For example, which dimensions (e.g., those related to negative affective experiences vs. those related to social engagement) are most useful to consider given the clinical issue at hand? Hopwood et al. (in press) recently delineated a set of hypotheses about how different dimensions of psychopathology might be responsive to different assessment and intervention approaches.[6] Second, what level of abstraction (i.e. broad traits or narrow symptom domains) is most useful for a given clinical question? One might use a manipulated use of assessment design (Hayes et al., 1987) to examine which level of information is most clinically beneficial for a specific population by randomly assigning assessors to reports exclusively based on higher order information, versus those who get the more specific level information. A promising and readily testable hypothesis is that it will be more effective to select treatments based on relatively broad dimensions rather than on fine-grained diagnostic distinctions (e.g., Barlow, Allen, & Choate, 2016; van Os et al., 2019), but that variability within those broad dimensions might be informative about clinically relevant patient heterogeneity. For instance, it may be the case that internalizing problems, in general, tend to respond to clinical strategies based on exposure and response prevention principles but that people with certain strengths related to responsibility and impulse control would be most likely to comply with homework and other aspects of treatment.

Clinical Personality Assessment

In many clinical settings, intake procedures involve extensive testing beyond an intake interview and/or DSM evaluation, possibly involving multimethod assessment, which aims at a more individualized case conceptualization than a DSM classification affords. Hence, the underlying idea is that some patients have particularly complex personality dynamics that are crucial to understand before making treatment plans, or that these dynamics require further scrutiny when treatment does not adequately progress, in order to tailor treatment to an individual patient. For example, a therapist about to treat a patient presenting with panic disorder (perhaps in combination with secondary depressive symptomatology) may not derive great benefit from prior knowledge of extensive clinical personality assessment, whereas a therapist about to treat a patient with complex personality issues around self-esteem in combination with poorly understood interpersonal hypervigilance and acting out behavior may fare very well. There are several standardized frameworks for reliably articulating person-specific models of personality dynamics (Eells et al., 1995; Luborsky & Diguer, 1998; Schacht & Henry, 1994). However, it remains untested whether their application leads to hypothesized indirect effect on outcomes (path b and c). Furthermore, it is assumed that these psychological structures and dynamics can be captured by the integration of information from a number of assessment instruments into a case-formulation. Relatedly, several authors have explicated and advocated the incremental validity and

[5] The presumed mechanism would be that unstructured diagnostic clinical interviewing is prone to biases, for example by overly focusing on one primary diagnosis to the possible exclusion of co-occurring syndromes (e.g., Zimmerman & Mattia, 2005). Structured interviews presumably reduce such biases and increase interrater reliability, thereby preventing both type I and type II errors in classification.

[6] A striking parallel is evident between the development ambitions of the HiTOP model and the MMPI-2-RF. Like HiTOP, the MMPI-2-RF stemmed from an effort to go beyond binary, a-theoretical convenience code-typing, and to derive a systematic hierarchical portrait of key psychopathology dimensions of patients instead. Indeed, the MMPI-2-RF bears strong structural similarity to HiTOP, although a recent article specifically stipulates that currently "no single measure [..] fully captures the HiTOP model" (p. 1078; Ruggero et al., 2019).

utility of multi method assessment (Finn, 1996; Hopwood & Bornstein, 2014; Meyer et al., 2001). Numerous case studies document how integration of self-report and performance-based instruments were crucial in answering poorly understood individualized assessment questions. Again, to our knowledge, there are no published studies that test the incremental value of adding performance-based instruments to self-report tests (or vice versa) in terms of TUCA.

A second precondition for TUCA is that meaningful and straightforward connections between assessment findings and treatment adaptation can be forged. Bram (2013) posited that in general, clinical personality assessment can (and should) do better in terms of making treatment relevant recommendations. He champions a treatment-centered approach to clinical personality assessment that draws on the seminal work of Norcross and Wampold (2011a, 2011b) on the so-called common factors of psychotherapy. Using test data and observations from patient-therapist interaction, clinical personality assessment can provide the therapist with information on those factors shown to contribute to better psychotherapy outcomes. Broadly, these factors can be divided into aspects pertaining to the therapeutic relationship versus those that describe ways in which the treatment may need adaptation to the specific client. Aspects pertaining to the therapeutic relationship include: alliance, empathy, collecting client feedback (deemed demonstrably effective), goal consensus, collaboration, positive regard (deemed probably effective), congruence/genuineness, repairing alliance ruptures, and managing countertransference (promising). Methods of adapting treatment to particular clients include reactance/resistance level, preferences, culture, religion and spirituality (demonstrably effective), stages of change, coping style (probably effective), expectations, and attachment style (promising). These factors may serve as a roadmap or heuristic tool for assessors seeking to make treatment-relevant recommendations that presumably foster TUCA.

Finally, clinical personality assessment may be beneficial in terms of its ability to enhance client preparation and motivation for the tasks of subsequent psychotherapy and to provide focus and goals for that therapy. Eventually—if these are indeed crucial moderators and mediators of therapeutic effectiveness—one would expect measurable effects on symptoms and daily functioning as well. However, given the complexity of the chain of effects between assessment and long-term outcomes, we would encourage researchers to focus on proximal outcomes first, and test distal outcomes later.

Clinical Personality Assessment Research Example: Incremental Utility of the Minnesota Multiphasic Personality Inventory (MMPI)-2

Lima et al. (2005) studied the incremental validity that MMPI-2 information might have for therapists planning and selecting treatments for patients in a university clinic. This study followed a manipulated assessment design (Hayes et al., 1987) in which therapists did or did not have access to MMPI-2 information to develop a case-conceptualization and derive treatment planning and selection. As reflected in the title—"The Incremental Validity of the MMPI-2: When Does Therapist Access Not Enhance Treatment Outcome?"—the authors found no beneficial effect for inclusion of the MMPI-2 for any of the included outcome measures (illness severity, improvement ratings, number of sessions attended, premature termination). Several design strengths are notable. First, the study took place in a community clinic, which suggests adequate variability in presenting problems, as was confirmed by examining the sample composition in terms of diagnoses (sufficient heterogeneity on path a). Another strength was the multioutcome assessment that included both therapist-ratings of symptoms (clinical global impression) and dysfunction (global assessment of functioning), as well as number of sessions attended and premature termination (path e).

When viewed from our conceptual framework as depicted in Figure 1, however, some limitations of the study become evident. First, MMPI-2 information might only serve to alter treatment allocation from one evidence based treatment (EBT) to another. Most research would suggest that the group effects of EBT are rather similar, and very few studies provide guidance on which patient characteristics are predictive of differential outcome across EBTs (path b → c → e). In other words, little variance is expected in terms of outcome, and no theory or empirical guidance is available for the therapist to base modification of treatment selection on. Moreover, it is unclear (except for reported clinical impressions) whether the clinicians indeed used the MMPI-2 info to modify treatment planning or selection (path b). Another limitation was that the MMPI-2 data were added to extensive record review and other routine assessment. Hence findings only pertain to incremental validity. From the study, it remained unclear whether the older record might contain MMPI-2 information, but regardless, TUCA can only be expected for cases in which the MMPI-2 info pointed in a different direction than the remainder of the available pretreatment information. Moreover, outcome (path e) was measured at the end of treatment, that is, quite some time post-assessment and was rather distal in nature (i.e. in terms of post treatment symptoms and dysfunction). Other salient outcomes, depending on one's theory, might include client satisfaction, quality of life, or other intermediate targets. Collectively, these limitations may well have hampered the power to yield or detect TUCA. Of note, most of these points were explicitly recognized by the authors, who in fact also argued for a more conditional appraisal of TUCA (Lima et al., 2005).

Remedying the noted limitations of studies like this will improve the validity of TUCA testing. This might involve (a) checking how the clinicians used the MMPI-2 information (i.e. was there indeed modification of the treatment planning and/or selection, corresponding to path b), (b) providing more diverse and explicit options for how MMPI-2 information might impact salient treatment characteristics (also path b), or developing and adhering to a theory of how the specific MMPI-2 indices might be relevant to treatment selection, for example operationalized by means of decision rules for one or more specific diagnostic groupings. Finally, one may enlist MMPI-2 experts who put their reputation on the line rather than graduate students.

Assessment as Treatment

The three classes of assessment driven interventions discussed above are similar in that assessment is explicitly aimed at improving outcomes by improving treatment (path a → b → c → e). In other words, direct effects of assessment may be present but are typically neither sought nor measured. The fourth and final class of

clinical assessment is distinctive in that these approaches explicitly aim for direct therapeutic benefits (path a → d → e). Evidence has steadily accrued that assessment can have beneficial clinical outcomes, as was documented in a meta-analytic study by Poston and Hanson (2010). Pooling 17 published studies from a diversity of samples, they compared the benefits of collaborative forms of psychological assessment to various control conditions and found strong effects were found for therapy process variables ($d = 1.11$) and smaller effects for therapy outcomes ($d = .37$). The authors concluded that "psychological assessment procedures—when combined with personalized, collaborative, and highly involving test feedback—have positive, clinically meaningful effects on treatment, especially regarding treatment processes" (Poston & Hanson, 2010, p203).

A subcategory of the studies included in Poston and Hanson's review concerns therapeutic assessment (TA; Finn, 2007). TA is a semistructured clinical procedure that seeks to generate direct therapeutic effects (i.e. direct TUCA) by itself. To date, investigations specific to TA have exclusively focused on these direct effects, that is, immediate post TA. Such effects are to be attributed to administrating TA as such, and the specific value of diagnostic information obtained can therefore not be disentangled from general benefits of assessment as an intervention (see Lilienfeld, Garb, & Wood, 2011). Symptom relief was observed in some patient samples (e.g., Finn & Tonsager, 1992; Newman & Greenway, 1997; numerous case studies), but not in others (e.g., De Saeger et al., 2014, in patients with DSM personality disorder, or Peters, 2001, in female eating disorder patients). In the latter cases, patient benefit did manifest in treatment readiness, alliance (De Saeger et al., 2014; Hilsenroth, Peters, & Ackerman, 2004), and satisfaction (De Saeger et al., 2014), or progression in terms of stages of change (i.e. moving from precontemplation to contemplation; Peters, 2001). As argued by Kamphuis and Finn (2019) TA may also be (directly) effective in terms of another outcome variable: that is, restoring epistemic trust or reducing epistemic hypervigilance (Fonagy, Luyten, & Allison, 2015). In sum, consistent with our model, the nature and extent of TA outcome benefit appears to be contingent upon the specific population, and may well be discordant across outcome domains (see Kamphuis & Finn, 2019 or De Saeger et al., 2014, for a more detailed review of the pertinent evidence). For future research, examination of how TA exerts its benefits, that is, its mechanisms of action, and its long-term impact should be priorities (Lilienfeld et al., 2011).

Assessment as Treatment Research Example: TA in Patients With Personality Disorders

De Saeger et al. (2014) reported on a randomized controlled clinical trial ($N = 74$) allocating patients awaiting treatment in a specialized clinic for personality disorders to either four sessions of Therapeutic Assessment (TA) or four sessions of a structured goal-focused pretreatment intervention. By pitching TA against a credible, protocol-driven comparison condition of equal dosage in a randomized controlled trial, it followed the manipulated assessment design advocated by Hayes et al. (1987). The first author of the current paper was involved in this study, and some of the ideas presented below are derived from reflection on this research experience.

The study has several positive features and satisfies a number of the conditions proposed in our model. First, the study targeted a specific population for whom aspects of TA may credibly have incremental utility. TA's emphasis on sustained empathy, emotional attunement and agency may be particularly salient to patients with personality disorders. Moreover, the study included a diversity of patients with severe personality problems (path a) and employed clinicians with adequate experience with providing the respective treatments. To track salient changes, the study conducted multidimensional assessments that also included intermediate outcomes (measured in this study as alliance ratings, satisfaction, focus, and expectation for subsequent treatment), consistent with the notion that change in personality problems may require sustained efforts (well beyond the 4 sessions offered in each arm of the RCT), which may also qualitatively differ from those picked up by standard, more narrowly defined symptom-list assessments (path e).

The principal shortcoming of the study was that only direct effects were assessed (path a → *d* → e), and indirect TUCA through treatment planning and selection was impossible in this patient group already awaiting their respective assigned treatments. In addition, no specific theory was a priori stipulated or tested as to how TA might outperform the control condition. Key improvements flow from these considerations. First and foremost, in an ideal world, both interventions would have informed subsequent treatments, in terms of treatment planning (treatment implications, such as key individual problem areas, central dilemmas, pitfalls), as well as in terms of treatment selection (path a → b → c → e). Moreover, a hypothesis regarding the mechanism by which TA would exert its positive impact would strengthen its theoretical implications, and that mechanism should be operationalized and longitudinally assessed. Specifically, if the theory would have it that TA works to increase epistemic trust (Fonagy et al., 2015) above and beyond the motivating effects of a credible motivational package, indices of this process should be periodically collected (which sets an agenda for research, as to our knowledge epistemic trust still awaits psychometrically adequate measurement).

Conclusion

We argue that discussing the TUCA makes little sense, just like discussing the reliability of a test (or testing) makes little sense: Such properties are always contextualized and contingent upon the nature and composition of the sample, setting, treatment options, outcomes sought, and so forth. To return to the Olsson and Fridell (2018) study we previously discussed, their findings may only be generalized to the population (women with comorbid substance use disorder), setting (mandatory care programs), and the type of incremental assessment conducted (mandatory information gathering assessment, with specified instruments). Furthermore, in studying TUCA, it makes sense to distinguish direct from indirect effects. Empirical support for direct TUCA (i.e. path a → d → e) has now accrued into a fairly convincing body of evidence, summarized in part in the meta-analysis of Poston and Hanson (2010). We encourage researchers to carefully match the nature of the targeted outcomes to the nature of the study sample and to choose a model of assessment that has a good record of yielding direct TUCA. Indirect TUCA has certain prerequisites, which were graphically illustrated in Figure 1. Indirect TUCA is all but im-

possible when there is minimal outcome-relevant individual variation in patients, treatments, or outcomes. That is to say, TUCA will be less likely when treatment generally fails or generally succeeds, when assessment cannot effectively tailor treatment planning or treatment selection, or when no salient individual differences are evident within the sample. We believe that a research agenda on the utility of extensive assessment (i.e. beyond DSM-diagnosis, intake interview and routine outcome monitoring) should therefore neither in research nor clinical practice be directed to broadly applying assessment-procedures to large clinical samples. Rather, we think that study designs should evaluate the preconditions before embarking on (symptom-based) evaluations of TUCA and furthermore have some theory of how assessment would contribute to (some form of) TUCA.

Furthermore, we believe that future research should pay more attention to the nature of outcomes sought. TUCA tends to be defined and tested in terms of short-term symptom improvement. We advocate inclusion of intermediate outcomes, and point to such variables in the context of the stages of change framework (Norcross, Krebs, et al., 2011), as well research on the potentiating effects on treatment alliance (Hilsenroth et al., 2004). In similar vein, we urge mindfulness about how assessment and treatment are expected to work to achieve their effects. For example, in patients with longstanding personality pathology, improvements in social learning (Fonagy et al., 2015) may lead to better relationship choices, which may be necessary before adjustment improves. A longer-term focus is then required to detect TUCA down the line, and of course such research is currently exceedingly scarce.

It needs to be acknowledged that conducting TUCA research is complex and presents with numerous pitfalls for questionable research practices. Several facets of the open science movement can be marshaled toward this end. As discussed, preregistering (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012) can help researchers and consumers clearly distinguish between confirmatory and exploratory effects; sharing materials and data openly can be useful for confirming effects and follow-up explorations of the data, and adversarial collaborations can be used to generate rigorous and compelling research designs. Especially when multiple outcome measures are considered for the evaluation of TUCA, explicitly preregistering primary and secondary outcome measures is crucial, along with the specific hypotheses to be tested. Doing so certainly does not preclude exploratory additional analyses, but these need to be identified as such to avoid risk of cherry-picking post hoc findings that are unlikely to replicate. In sum, replication of preregistered outcomes should precede clinical implementation, and exploratory approaches—essential for theory development and innovation—should be acknowledged accordingly.

Extensive assessment is probably most profitably reserved for cases in which regular treatment is not working, or did not work in the past. Such patients possibly need more individualized, multimethod case-conceptualizations. Stepped assessment (Kamphuis, 2010)—that is, tailoring the dose of CA to the surmised complexity of the psychopathology of the patient—may thus contribute to the efficient allocation of scarce resources. A final implication of this line of reasoning is that it may well be the case that a substantial amount of assessment-as-usual in clinical practice does not actually contribute much to improved treatment outcomes (but may serve other purposes like normative description, tracking change, etc.; see Meyer et al., 2001). Of course, this raises the question: Without assessment of patient characteristics, how do we know for whom the assessment will be of most value?

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). Washington, DC: Author.

Barlow, D. H., Allen, J. D., & Choate, M. L. (2016). Toward a unified treatment for emotional disorders. *Behavior Therapy, 47,* 838–853. http://dx.doi.org/10.1016/j.beth.2016.11.005

Beutler, L. E., & Clarkin, J. F. (1990). *Systematic treatment selection: Toward targeted therapeutic interventions*. New York, NY: Routledge.

Beutler, L. E., & Harwood, T. M. (2000). *Prescriptive Psychotherapy: A practical guide to systematic treatment selection*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/med:psych/9780195136692.001.0001

Beutler, L. E., Kimpara, S., Edwards, C. J., & Miller, K. D. (2018). Fitting psychotherapy to patient coping style: A meta- analysis. *Journal of Clinical Psychology, 74,* 1980–1995. http://dx.doi.org/10.1002/jclp.22684

Blatt, S. (1992). The differential effect of psychotherapy and psychoanalysis with anaclitic and introjective patients: The Menninger psychotherapy research project revisited. *Journal of the American Psychoanalytic Association, 40,* 691–724. http://dx.doi.org/10.1177/000306519204000303

Bohart, A. C., & Wade, A. G. (2013). The client in psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 219–257). Hoboken, NJ: Wiley.

Bram, A. D. (2013). Psychological testing and treatment implications: We can say more. *Journal of Personality Assessment, 95,* 319–331. http://dx.doi.org/10.1080/00223891.2012.736907

Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment selection in depression. *Annual Review of Clinical Psychology, 14,* 209–236. http://dx.doi.org/10.1146/annurev-clinpsy-050817-084746

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology, 57,* 505–528. http://dx.doi.org/10.1146/annurev.psych.57.102904.190146

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, R. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The personalized advantage index (PAI): translating research on prediction into individualized treatment recommendations: A demonstration. *PLoS ONE, 9,* e83875. http://dx.doi.org/10.1371/journal.pone.0083875

De Saeger, H., Bartak, A., Eder, E. E., & Kamphuis, J. H. (2016). Memorable experiences in therapeutic assessment: Inviting the patient's perspective following a pretreatment randomized controlled trial. *Journal of Personality Assessment, 98,* 472–479. http://dx.doi.org/10.1080/00223891.2015.1136314

De Saeger, H., Kamphuis, J. H., Finn, S. E., Smith, J. D., Verheul, R., van Busschbach, J. J. V., . . . Horn, E. (2014). Therapeutic assessment promotes treatment readiness but does not affect symptom change in patients with personality disorders: Findings from a randomized clinical trial. *Psychological Assessment, 26,* 474–483. http://dx.doi.org/10.1037/a0035667

Eells, T., Horowitz, M., Singer, J., Salovey, P., Daigle, D., & Turvey, C. (1995). The role- relationship models method: A comparison of independently derived case formulations. *Psychotherapy Research, 5,* 154–168. http://dx.doi.org/10.1080/10503309512331331276

Finn, S. E. (1996). Assessment feedback integrating MMPI-2 and Rorschach findings. *Journal of Personality Assessment, 67,* 543–557. http://dx.doi.org/10.1207/s15327752jpa6703_10

Finn, S. E. (2007). *In our clients' shoes: Theory and techniques of Therapeutic Assessment*. New York, NY: Erlbaum.

Finn, S. E., & Tonsager, M. E. (1992). Therapeutic effects of providing MMPI-2 feedback to college students awaiting therapy. *Psychological Assessment, 4,* 278–287. http://dx.doi.org/10.1037/1040-3590.4.3.278

Fonagy, P., Luyten, P., & Allison, E. (2015). Epistemic petrification and the restoration of epistemic trust: A new conceptualization of borderline personality disorder and its psychosocial treatment. *Journal of Personality Disorders, 29,* 575–609. http://dx.doi.org/10.1521/pedi.2015.29.5.575

Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73,* 1154–1163. http://dx.doi.org/10.1037/0022-006X.73.6.1154

Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist, 42,* 963–974. http://dx.doi.org/10.1037/0003-066X.42.11.963

Hilsenroth, M. J., Peters, E. J., & Ackerman, S. J. (2004). The development of therapeutic alliance during psychology assessment: Patient and therapist perspectives across treatment. *Journal of Personality Assessment, 83,* 332–344. http://dx.doi.org/10.1207/s15327752jpa8303_14

Hopwood, C. J. (2018). A framework for treating *DSM–5* alternative model of personality disorder features. *Personality and Mental Health, 12,* 107–125. http://dx.doi.org/10.1002/pmh.1414

Hopwood, C., & Bornstein, R. (2014). *Multimethod clinical assessment*. New York, NY: Guilford Press.

Hopwood, C. J., Bagby, R. M., Gralnick, T., Ro, E., Ruggero, C., Mullins-Sweatt, S., . . . Zimmermann, J. (in press). Using the hierarchical taxonomy of psychopathology to inform psychotherapy. *Journal of Psychotherapy Integration*.

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

Kamphuis, J. H. (2010). *Therapeutische diagnostiek (therapeutic assessment). Public acceptance lecture of chair in Clinical Psychology*. Amsterdam, the Netherlands: University of Amsterdam.

Kamphuis, J. H., & Finn, S. E. (2019). Therapeutic assessment in personality disorders: Toward the restoration of epistemic trust. *Journal of Personality Assessment, 101,* 662–674. http://dx.doi.org/10.1080/00223891.2018.1476360

Kroenke, K., & Spitzer, R. L. (2002). The PHQ9: A new depression diagnostic and severity measure. *Psychiatric Annals, 32,* 509–515. http://dx.doi.org/10.3928/0048-5713-20020901-06

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Clinical and Consulting Psychology, 69,* 159–172. http://dx.doi.org/10.1037/0022-006X.69.2.159

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science and Practice, 10,* 288–301. http://dx.doi.org/10.1093/clipsy.bpg025

Lilienfeld, S. O., Garb, H. N., & Wood, J. M. (2011). Unresolved questions concerning the effectiveness of psychological assessment as a therapeutic intervention: Comment on Poston & Hanson (2010). *Psychological Assessment, 23,* 1047–1055. http://dx.doi.org/10.1037/a0025177

Lima, E. N., Stanley, S., Kaboski, B., Reitzel, L. R., Richey, A., Castro, Y., . . . Joiner, T. E., Jr. (2005). The incremental validity of the MMPI-2: When does therapist access not enhance treatment outcome? *Psychological Assessment, 17,* 462–468. http://dx.doi.org/10.1037/1040-3590.17.4.462

Luborsky, L., & Diguer, L. (1998). The reliability of the CCRT measure: Results from eight samples. In L. Luborsky & P. Crits-Christoph (Eds.), *Understanding transference: The Core Conflictual Relationship Theme method* (pp. 97–107). Washington, DC: American Psychological Association.

Luedtke, A., Sadikova, E., & Kessler, R. C. (2019). Sample size requirements for multivariate models to predict between-patient differences in best treatment for major depressive disorder. *Clinical Psychological Science, 7,* 445–461. http://dx.doi.org/10.1177/2167702618815466

Luhmann, M., Orth, U., Specht, J., Kandler, C., & Lucas, R. E. (2014). Studying changes in life circumstances and personality: It's about time. *European Journal of Personality, 28,* 256–266. http://dx.doi.org/10.1002/per.1951

McWilliams, N. (2005). Preserving our humanity as therapists. *Psychotherapy, 42,* 139–151. http://dx.doi.org/10.1037/0033-3204.42.2.139

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press. http://dx.doi.org/10.1037/11281-000

Meehl, P. E. (1959). Some ruminations on the validation of clinical procedures. *Canadian Journal of Psychology/Revue Canadienne de psychologie, 13,* 102–128. http://dx.doi.org/10.1037/h0083769

Memon, M. A., Cheah, J. H., Ramayah, T., Ting, H., Chuah, F., & Cham, T. H. (2019). Moderation analysis: Issues and guidelines. *Journal of Applied Structural Equation Modeling, 3,* i–xi. http://dx.doi.org/10.47263/JASEM.3(1)01

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., . . . Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56,* 128–165. http://dx.doi.org/10.1037/0003-066X.56.2.128

Mullins-Sweatt, S. N., Hopwood, C. J., Chmielewski, M., Meyer, N. A., Min, J., Helle, A. C., & Walgren, M. D. (2020). Treatment of personality pathology through the lens of the hierarchical taxonomy of psychopathology: Developing a research agenda. *Personality and Mental Health, 14,* 123–141. http://dx.doi.org/10.1002/pmh.1464

Newman, M. L., & Greenway, P. (1997). Therapeutic effects of providing MMPI-2 feedback to clients at a university counseling service: A collaborative approach. *Psychological Assessment, 9,* 122–131. http://dx.doi.org/10.1037/1040-3590.9.2.122

Norcross, J., Krebs, P. M., & Prochaska, J. O. (2011). Stages of change. *Journal of Clinical Psychology, 67,* 143–154. http://dx.doi.org/10.1002/jclp.20758

Norcross, J. C., & Wampold, B. E. (2011a). Evidence-based therapy relationships: Research conclusions and clinical practices. *Psychotherapy, 48,* 98–102. http://dx.doi.org/10.1037/a0022161

Norcross, J. C., & Wampold, B. E. (2011b). What works for whom: Tailoring psychotherapy to the person. *Journal of Clinical Psychology, 67,* 127–132. http://dx.doi.org/10.1002/jclp.20764

Olsson, T. M., & Fridell, M. (2018). The five-year costs and benefits of extended psychological and psychiatric assessment versus standard intake interview for women with comorbid substance use disorders treated in compulsory care in Sweden. *BMC Health Services Research, 18,* 53. http://dx.doi.org/10.1186/s12913-018-2854-y

Paul, G. (1967). Strategy of outcome research in psychotherapy. *Journal of Counseling Psychology, 31,* 109–118. http://dx.doi.org/10.1037/h0024436

Peters, J. M. (2001). The effect of therapeutic assessment on women with eating disorders [Unpublished dissertation]. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 61,* 4422.

Poston, J. M., & Hanson, W. E. (2010). Meta-analysis of psychological assessment as a therapeutic intervention. *Psychological Assessment, 22,* 203–212. http://dx.doi.org/10.1037/a0018679

Preacher, K. J., & Sterba, S. K. (2019). Aptitude-by-treatment interactions in research on educational interventions. *Exceptional Children, 85,* 248–264. http://dx.doi.org/10.1177/0014402918802803

Project MATCH Research Group. (1997). Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol, 58,* 7–29. http://dx.doi.org/10.15288/jsa.1997.58.7

Ruggero, C. J., Kotov, R., Hopwood, C. J., First, M., Clark, L. A., Skodol, A. E., . . . Zimmermann, J. (2019). Integrating the hierarchical taxonomy of psychopathology (HiTOP) into clinical practice. *Journal of Consulting and Clinical Psychology, 87,* 1069–1084. http://dx.doi.org/10.1037/ccp0000452

Schacht, T., & Henry, W. (1994). Modeling recurrent patterns of interpersonal relationship with structural analysis of social behavior: The SASB-CMP. *Psychotherapy Research, 4,* 208–221. http://dx.doi.org/10.1080/10503309412331334042

Schulte, D. (1996). Tailor-made and standardized therapy: Complementary tasks in behavior therapy a contrarian view. *Journal of Behavior Therapy and Experimental Psychiatry, 27,* 119–126. http://dx.doi.org/10.1016/0005-7916(96)00015-8

Schulte, D., Kunzel, R., Pepping, G., & Schulte-Bahrenberg, T. (1992). Tailor-made versus standardized therapy for phobic patients. *Advances in Behaviour Research and Therapy, 14,* 67–92. http://dx.doi.org/10.1016/0146-6402(92)90001-5

Smid, W. J., Kamphuis, J. H., de Wever, E. C., & Van Beek, D. (2013). Treatment referral for sex offenders based on clinical judgment versus actuarial risks assessment: Match and analysis of mismatch. *Journal of Interpersonal Violence, 28,* 2273–2289. http://dx.doi.org/10.1177/0886260512475313

Smid, W. J., Kamphuis, J. H., Wever, E. C., & Van Beek, D. (2014). A comparison of the predictive properties of nine sex offender risk assessment instruments. *Psychological Assessment, 26,* 691–703. http://dx.doi.org/10.1037/a0036616

Telch, M. J., Kamphuis, J. H., & Schmidt, N. B. (2011). The effects of comorbid personality disorders on cognitive behavioral treatment for panic disorder. *Journal of Psychiatric Research, 45,* 469–474. http://dx.doi.org/10.1016/j.jpsychires.2010.08.008

Tjak, A., Morina, J., Topper, N., & Emmelkamp, M. (2020). *One year follow-up and mediation in cognitive behavioral therapy and acceptance and commitment therapy for adult depression*. Manuscript submitted for publication.

van Os, J., Guloksuz, S., Vijn, T. W., Hafkenscheid, A., & Delespaul, P. (2019). The evidence-based group-level symptom-reduction model as the organizing principle for mental health care: Time for change? *World Psychiatry, 18,* 88–96. http://dx.doi.org/10.1002/wps.20609

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7,* 632–638. http://dx.doi.org/10.1177/1745691612463078

Weekers, L., Hutsebaut, J., Bach, B., & Kamphuis, J. H. (2020). Scripting the *DSM–5* alternative model for personality disorders assessment procedure: A clinically feasible multi-informant multi-method approach. *Personality and Mental Health, 14,* 304–318. http://dx.doi.org/10.1002/pmh.1481

Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment and problem-solving strategies in routine practice. *Journal of Counseling Psychology, 50,* 59–68. http://dx.doi.org/10.1037/0022-0167.50.1.59

World Health Organization. (2018). *International classification of diseases for mortality and morbidity statistics* (11th Revision). Retrieved from https://icd.who.int/browse11/l-m/en

Zedeck, S. (1971). Problems with the use of "moderator" variables. *Psychological Bulletin, 76,* 295–310. http://dx.doi.org/10.1037/h0031543

Zilcha-Mano, S. (2019). Major developments in methods addressing for whom psychotherapy may work and why. *Psychotherapy Research, 29,* 693–708. http://dx.doi.org/10.1080/10503307.2018.1429691

Zimmerman, M., & Mattia, J. I. (2005). Psychiatric diagnosis in clinical practice: Is comorbidity being missed? *Comprehensive Psychiatry, 40,* 182–191. http://dx.doi.org/10.1016/S0010-440X(99)90001-9

Received January 14, 2020
Revision received September 1, 2020
Accepted September 4, 2020 ■