COMMENT

Unresolved Questions Concerning the Effectiveness of Psychological Assessment as a Therapeutic Intervention: Comment on Poston and Hanson (2010)

Scott O. Lilienfeld Emory University Howard N. Garb Lackland Air Force Base, San Antonio, Texas

James M. Wood University of Texas at El Paso

In a recent article in this journal, Poston and Hanson (2010) reported a meta-analysis of 17 studies on the use of psychological assessment as a therapeutic intervention (PATI) and concluded that "psychological assessment procedures—when combined with personalized, collaborative, and highly involving test feedback—have positive, clinically meaningful effects on treatment" (Poston & Hanson, 2010, p. 203). Although extant data suggest that PATI can sometimes exert positive effects, Poston and Hanson's (2010) meta-analysis may overstate the magnitude of these effects because the authors (a) included several studies that combined assessment with treatment components that are irrelevant to PATI, sometimes rendering it impossible to attribute any observed effects to PATI per se and (b) excluded numerous nonsignificant results. Moreover, the studies Poston and Hanson (2010) reviewed neglected to rule out Barnum effects as alternative explanations for client improvement, raising the possibility that PATI works for reasons other than those proposed by its advocates. We conclude that Poston and Hanson's (2010) review leaves a number of lingering questions concerning the treatment utility of PATI unanswered.

Keywords: assessment, intervention, treatment utility, Barnum effect

Does the practice of psychological assessment yield enhanced treatment outcomes? The answer to this question bears significant implications for clinical psychology because if administering and interpreting psychological measures, such as the Minnesota Multiphasic Personality Inventory—2 (MMPI-2), the Beck Depression Inventory, or the Rorschach Inkblot Test, do not contribute to improved treatment outcomes, it would raise serious questions concerning the routine use of clinical assessment in psychotherapy (Garb, Lilienfeld, Nezworski, Wood, & O'Donohue, 2009). In contrast, if the administration and interpretation of such measures—including sharing these interpretations with clients—is associated with better treatment outcomes, it would go a substantial way toward affirming the value of psychological assessment in clinical settings.

Most authors who have reviewed the literature on the treatment utility (Hayes, Nelson, & Jarrett, 1987) of assessment have concluded that the ability of psychological measures to enhance treat-

ment outcomes has barely been examined, let alone established, for most widely used psychological tests (Garb et al., 2009; Hayes et al., 1987; Hunsley, 2003; Nelson-Gray, 2003; Wood, Garb, Lilienfeld, & Nezworski, 2002; cf. Butcher & Rouse, 1996, p. 101). In a recent review, Garb et al. (2009) concluded that treatment utility has been demonstrated for brief measures of interim treatment progress in psychotherapy that enhance treatment outcomes when fed back regularly to therapists (Lambert, Whipple, Hawkins, Vermeesch, Nielson, & Smart, 2003). Nevertheless, they also argued that treatment utility has not been convincingly demonstrated for other uses of psychological assessment, including the routine practice of administering measures of personality or psychopathology in the early phases of treatment. As recently as 1997, Finn and Tonsager (1997), who pioneered much of the research on the treatment utility of psychological assessment, conceded that "the empirical evidence for the treatment utility of assessment is weaker than many of us might want" (p. 385). Indeed, it is remarkable that more than 90 years since the introduction of the first major psychological test, the Woodworth Personal Data Sheet (Woodworth, 1918), during World War I, the treatment utility of psychological assessment remains in dispute.

In a recent article in this journal, Poston and Hanson (2010)—who did not cite a number of negative reviews (e.g., Hunsley, 2003; Wood et al., 2002) on treatment utility—arrived at a sharply different conclusion. They conducted a meta-analysis of 17 pub-

Scott O. Lilienfeld, Department of Psychology, Emory University; Howard N. Garb, Lackland Air Force Base, San Antonio, Texas; James M. Wood, Department of Psychology, University of Texas at El Paso.

Correspondence concerning this article should be addressed to Scott O. Lilienfeld, Room 473, Department of Psychology, Emory University, 36 Eagle Row, Atlanta, GA 30322. E-mail: slilien@emory.edu

lished studies, incorporating 52 effect sizes, on the use of psychological assessment as a therapeutic intervention (PATI). For the purposes of their meta-analysis, they operationalized PATI as "the process of completing any formal psychological test/measure and receiving feedback on the results" (p. 205). Poston and Hanson focused largely on a collaborative feedback model called therapeutic assessment (Finn & Tonsager, 1992) and concluded that "psychological assessment procedures—when combined with personalized, collaborative, and highly involving test feedback—have positive, clinically meaningful effects on treatment" (p. 203). Nevertheless, they also included studies that did not examine therapeutic assessment per se but that nonetheless examined the use of feedback on PATI. Hence, in the remainder of this article, we use the acronym PATI when referring more broadly to the use of assessment feedback as a therapeutic intervention and reserve the term therapeutic assessment for the specific form of PATI developed by Finn and Tonsager (1992).

Across the studies they analyzed, Poston and Hanson (2010) reported an overall effect size (Cohen's d) of .423, which was statistically significant and close to medium in magnitude according to the criteria outlined by Cohen (1969). Nevertheless, as Cohen (1988) himself noted, his widely used effect size descriptors (small, medium, and large) are only rough guidelines: "These qualitative adjectives. . . may not be reasonably descriptive in any specific area. Thus, what a sociologist may consider a small effect may be appraised as medium by a clinical psychologist" (p. 277). Indeed, in the domain of psychotherapy, in which one deals frequently with difficult to treat or largely intractable conditions that can produce marked distress, impairment, or both, even effects of d = 0.1 or 0.2 may be far from trivial in magnitude (see also Valentine & Cooper, 2003). Poston and Hanson's conclusions, if correct, would be a major step toward resolving one of the most longstanding controversies in psychological assessment (Hayes et al., 1987) and buttress the claim that psychological assessment, including assessment feedback to clients, should play a key role in treatment planning and execution (Harkness & Lilienfeld, 1997).

As Poston and Hanson (2010) observed, the predominant model of PATI stems from an important article on therapeutic assessment by Finn and Tonsager (1992), who found that sharing and discussing the results of MMPI-2 protocols with clients was associated with decreases in symptomatic distress and increases in selfesteem and hopefulness, both immediately following the session and at a 2-week follow-up (see also Newman & Greenway, 1997). Although other authors have developed slightly different models of PATI, all share a focus on "(a) developing and maintaining empathic connections with clients, (b) working collaboratively with clients to define individualized assessment goals, and (c) sharing and exploring assessment results with clients" (Finn & Tonsager, 1997, p. 378; see also Finn, 1996, 2007). According to Poston and Hanson, their meta-analytic results demonstrated that the adoption of this model of assessment "significantly enhances the treatment process" (p. 211).

We agree with Poston and Hanson (2010) that PATI comprises a promising class of techniques that merit additional investigation and that several well-conducted studies (e.g., Finn & Tonsager, 1992) suggest that PATI may exert positive therapeutic effects in some cases. Nevertheless, close inspection of Poston and Hanson's analyses raises concerns that the authors have overstated the magnitude of the effectiveness of PATI. In particular, Poston and

Hanson (a) included several studies that combined assessment with treatment components that are irrelevant to PATI, sometimes rendering it impossible to attribute any observed effects to PATI per se and (b) excluded numerous nonsignificant results for reasons that were not explained. Moreover, the studies Poston and Hanson reviewed neglected to rule out Barnum effects as alternative explanations for client improvement, raising the possibility that PATI may sometimes work for reasons other than those argued by its proponents.

Confounding of PATI With Extraneous Treatment Components

When reviewing research on the treatment utility of psychological assessment, Poston and Hanson (2010) included at least three studies that combined assessment and non-PATI-related treatment components, rendering it impossible to draw clear-cut conclusions regarding the therapeutic effects of PATI per se. Therapeutic assessment as outlined by Finn and Tonsager (1992) often includes explicit intervention components, such as role-playing, support, explanation of client problems, and a discussion of how clients' problems revealed by the assessment play out in their everyday lives. Nevertheless, it does not include miscellaneous treatment components derived from formal schools of therapy, such as cognitive-behavioral or psychodynamic interventions. Nor does it include treatment components that are not tied to assessment feedback. If one includes such components in a PATI package, it becomes difficult or impossible to evaluate the treatment utility of PATI per se because such components are irrelevant to PATI.

In one study (Jobes, Wong, Conrad, Drozd, & Neal-Walden, 2005) included in Poston and Hanson's (2010) meta-analysis, the authors retrospectively reviewed medical records for clients who had received "treatment as usual" (TAU) and for clients who had received treatment from therapists trained in a novel clinical approach to identifying, assessing, and managing suicidal outpatients, called the Collaborative Assessment and Management of Suicidality (CAMS; Jobes & Drozd, 2004). As described by the authors, CAMS contains substantial therapeutic components that go considerably beyond assessment feedback. For example, the authors noted that CAMS "integrates psychodynamic, cognitive, behavioral, humanistic, existential, and interpersonal theory into a structured clinical format emphasizing the importance of the therapist and patient working together to elucidate and understand the 'functional' role of suicidality in the patient's phenomenological world" (Jobes et al., 2005, p. 484). In addition, CAMS incorporates a "treatment planning process where [sic] the patient and clinician coauthor an outpatient treatment plan" (p. 484).

Jobes et al.'s (2005) results, which yielded significantly greater reductions in suicidality for clients who received CAMS, compared with clients who received TAU, seemingly provided support for the efficacy of CAMS. Nevertheless, their findings should not be interpreted as supporting the assertion that "psychological assessment procedures have positive clinically meaningful effects on treatment" (Poston & Hanson, 2010, p. 203), as Jobes et al. confounded the effects of assessment and treatment by administering an intervention that incorporates components that go well beyond the assessment administered to clients.

Hilsenroth, Peters, and Ackerman (2004) compared the effects of PATI in a group of outpatients at a university-based community

clinic with a matched group of outpatients who received a standard information gathering (IG) model of assessment. Their results revealed significantly higher levels of patient-rated therapeutic alliance for clients who received therapeutic assessment versus those who received the IG assessment. As acknowledged by Poston and Hanson (2010), Hilsenroth et al. did not compare the groups in symptomatic improvement (also see Poston & Hanson, 2010, Table 1, p. 206), so their study does not demonstrate a positive effect of PATI on client outcomes (as opposed to the therapeutic alliance, which is traditionally regarded as a mediator of therapy outcomes; Kazdin & Nock, 2003) is questionable. More to the point, as part of their PATI procedure, the therapists in Hilsenroth et al.'s study engaged clients in an exploration of a core conflictual relationship theme (CCRT; Luborsky & Crits-Cristoph, 1997). As described in an earlier publication coauthored by two members of the same research team (Ackerman, Hilsenroth, Baity, & Blagys, 2000), the "exploration of the CCRT helped the clinician focus on collaboration, alliance building [emphasis added], examination of factors contributing to the maintenance of life problems (often relational) and potential solutions" (p. 94). Clearly, the CCRT contains substantial therapeutic elements that could have accounted for the significant between-group differences in therapeutic alliance, especially given that CCRT is designed explicitly to strengthen the client-therapist alliance. Rather than concluding that Hilsenroth et al.'s findings demonstrated the therapeutic efficacy of PATI, one could offer the more parsimonious interpretation that a procedure designed in part to enhance the therapeutic alliance in fact succeeds in doing so.

Another study included in Poston and Hanson's (2010) metaanalysis is a classic investigation by Miller, Benefield, and Tonigan (1993), who compared three groups: (a) immediate feedback with a directive-confrontational style, (b) immediate feedback with a client-centered style, and (c) delayed feedback (waiting-list control) with later assignment to the directive-confrontational or client-centered style. Participants in all three groups first received a 2 hr evaluation that included a breath test to ensure sobriety at the time of testing, measures sensitive to early alcohol-related risk and impairment, including a neuropsychological test battery, and the drawing of a serum sample to be assayed for biological indexes of alcohol impairment. One week later, participants in the two immediate feedback groups were afforded the opportunity to discuss their test results. In the directive-feedback condition, clinicians confronted clients by emphasizing the evidence of their alcohol problems. In the client-centered feedback condition, clinicians responded to clients in an empathic manner. Participants in both immediate feedback conditions displayed a significant reduction in drinking relative to controls. After 1 year, the directiveconfrontational style yielded worse alcohol drinking outcomes than did the client-centered style, although both groups displayed improvement relative to their baseline rate of alcohol consumption.

Nevertheless, it is doubtful whether Miller et al.'s (1993) findings bear on the efficacy of PATI, as Poston and Hanson (2010) claim. Indeed, as Miller et al. themselves noted, the client-centered feedback condition "adhered closely to what has been described elsewhere as motivational interviewing" (p. 456), an intervention widely regarded as a form of psychotherapy derived from Rogerian person-centered principles (Arkowitz & Miller, 2008). The direct-confrontational feedback condition also comprised substantial therapeutic components that are irrelevant to assessment feedback; as described by Miller et al., "interviewers were instructed to confront client resistance by emphasizing the evidence of alcohol problems, giving direct advice, and disagreeing with client minimization of problems" (p. 456). Again, the inclusion of these therapeutic elements in Miller et al.'s study renders it difficult or impossible to draw meaningful conclusions regarding the unique value of PATI per se.

It is worth noting that the effect sizes (ds) of two of these studies (Miller et al., 1993; Hilsenroth et al., 2004) were medium to large

Table 1
Nonsignificant Findings and Effect Sizes Omitted From Poston and Hanson's (2010)
Meta-analysis

Study	N	Dependent measure	d
Allen, Montgomery, Tubman, Frazier,			
& Escovar (2003)	83	State self-esteem: Performance	0.37
		State self-esteem: Social	0.40
		State self-esteem: Appearance	0.40
Folds & Gazda (1966)	44	Self-concept	unknowna
Hanson, Claiborn, & Kerr (1997)	26	Thought listing elaboration	0.3
		Thought listing favorability	0.80
		Session smoothness	0.09
		Session arousal	0.46
		Session positivity	0.49
Jobes et al. (2005)	37	Outcome questionnaire	0.1
		Suicidality	-0.06
		Global assessment of functioning	-0.12
		Total sessions	0.26
		Revisiting Poston and Hanson (2010)	28
		Treatment days	0.44
		Number of cancellations	0.31
		Direct costs	0.045
Wild, Cunningham, & Roberts (2006)	306	Drinking quantity among non-problem drinkers	unknowna

^a Effect size cannot be computed from information provided in the original article.

in magnitude under Cohen's (1969) guidelines (0.543 and 1.022, respectively). Defenders of PATI might contend that the operationalization of this procedure should be sufficiently broad to encompass such procedures as treatment planning for suicidality, generation of potential solutions to life problems, empathic reflection of the client's feelings, and confronting clients about their problematic behaviors. Nevertheless, extending the operationalization of the PATI construct to include therapeutic techniques that go well beyond assessment feedback per se raises troubling questions regarding the boundaries of PATI (see also Smith, Handler, & Nash, 2010, for a study that incorporated a family intervention session, which itself often includes psychodrama and family board games, as part of a PATI intervention). Doing so also would appear to conflict with Poston and Hanson's (2010) explicit operationalization of PATI as "the process of completing any formal psychological test/measure and receiving feedback on the results" (p. 205). Indeed, we might well ask, "What therapeutic procedures would not fall under the umbrella of PATI?" Expanding the boundaries of the already broad PATI rubric to include a variety of well-established therapy techniques would render claims concerning the efficacy of PATI difficult or impossible to falsify because doing so permits investigators to "count" findings derived from assessment procedures that incorporate therapy techniques as supportive of PATI's efficacy.

Exclusion of Nonsignificant Findings

Further complicating the interpretation of Poston and Hanson's (2010) results is the fact that the authors, for unclear reasons, excluded numerous—by our count, 17—nonsignificant results (from five of their 17 articles) from their analysis. Table 1 lists these omitted nonsignificant findings and their accompanying effect sizes in the articles; readers may wish to compare this table with Table 1 in Poston and Hanson (p. 206). Because Poston and Hanson provide no rationale for the omission of such findings—indeed, they do not even mention these findings in their article—the results and conclusions of their meta-analysis are difficult to interpret with confidence. As can be seen in the table, the omitted effect sizes range from -0.12 to 0.80; in two of the five studies, effect sizes could not be calculated from the nonsignificant findings provided.

In one of the studies in Poston and Hanson's (2010) meta-analysis, Allen, Montgomery, Tubman, Frazier, and Escovar (2003) reported 12 results for examiner, rapport-related, and self-related measures of treatment outcome. Yet Poston and Hanson included only the 8 results that were statistically significant and excluded the 3 nonsignificant results for state self-esteem (see Allen et al., 2003, Table 2, p. 174). Yet Allen et al. had explicitly predicted changes in state self-esteem in their hypotheses (see Allen et al., 2003, p. 169), so it is unclear why Poston and Hanson excluded these negative results from their analysis, especially because they did include the lone significant finding for trait self-esteem.

In the investigation by Folds and Gazda (1966); Poston and Hanson (2010) reported two positive findings (for self-understanding and goals) but neglected to note that the groups did not differ significantly on the predicted outcome variable of concept of self (see Folds & Gazda, 1996, p. 322). The effect size for this variable cannot be calculated from the information provided by Folds and Gazda, but it was not mentioned by Poston and

Hanson, even though changes in this variable were one of the key hypotheses of the study.

In another study (Hanson, Claiborn, & Kerr, 1997), the authors reported results for seven variables on the effects of career counseling, but Poston and Hanson (2010) included only the two statistically significant results (see Hanson et al., 1997, Table 1, p. 403). Strikingly, although Hanson et al. reported data from four ratings of reactions to counseling sessions, namely, depth, smoothness, arousal, and positivity, Poston and Hanson reported only the effect size for depth (d=1.57), which was by far the largest of the four and the only one of the four that was statistically significant (see Table 1 for the other three effect sizes).

In a third study, Jobes et al. (2005) reported statistically significant and positive results for two of four dependent measures of suicidality, using the CAMS model of suicide assessment and feedback discussed earlier. Yet Poston and Hanson (2010) did not include the data from a number of other relevant psychological variables, including a table (see Jobes et al., 2005, p. 490) that reported seven of eight nonsignificant findings, such as measures of client symptoms, client progress in psychotherapy, global psychological functioning, and health care costs (in the other direction, Poston and Hanson also omitted a table consisting of four of four significant findings for medical utilization, as well as a positive finding for medical appointments per year; see Jobes et al., 2005, pp. 490 & 492). Perhaps most important, Poston and Hanson omitted the crucial effect size for in-session suicidality, which was actually slightly negative (d = -0.06). Yet significant differences on these very outcome variables had been hypothesized by the authors: "we hypothesized that CAMS patients would have better categorical treatment outcomes, less [sic] psychiatric symptoms, meet criteria for clinical recovery more quickly, and have lower overall mental health care costs than TAU [treatment as usual] patients" (pp. 484-485).

In a fifth study, Wild, Cunningham, and Roberts (2006) reported significant effects of PATI on binge drinking in problem drinkers but nonsignificant effects in nonproblem drinkers. Yet Poston and Hanson (2010) reported only the results for the former group, even though the latter group consisted of drinkers in the general population who were interested in obtaining self-help materials on alcohol. In fact, the findings for the latter group ran in the opposite direction, with higher adjusted means for drinking in both men and women exposed to the assessment-feedback intervention (see Wild et al., 2006, Table 2, p. 247). Although the original authors had anticipated larger effects for PATI on problem drinkers (p. 242),

¹ In other cases, Poston and Hanson (2010) appear to have been inconsistent in their reporting of significant versus nonsignificant outcome and process variables. In the study by Newman and Greenway (1997), they reported the two significant findings for psychological symptoms and self-esteem but omitted the nonsignificant finding for "feelings toward examiner," which captured "the extent to which a client felt accepted, liked, and respected by the examiner" (p. 127). Although one might legitimately question whether clients' attitude toward the clinician is a relevant outcome or process variable for ascertaining the efficacy of PATI, Poston and Hanson counted the positive finding by Hilsenroth et al. (2004) on what is arguably a similar variable, namely, the strength of therapeutic alliance, as supportive of the efficacy of PATI. In any case, Poston and Hanson did not justify the inclusion in their meta-analysis of the latter positive finding but not the former negative finding.

they offered no explicit grounds for neglecting positive results for nonproblem drinkers, had they been found.

The practice of omitting nonsignificant findings runs counter to widely accepted recommendations for conducting meta-analyses and other literature reviews (Boutron, Dutton, Ravaud, & Altman, 2010; Hewitt, Mitchell, & Torgeson, 2008), which mandate reporting of all relevant outcomes, including those that are not statistically significant (Moher, Liberati, Tetzlaff, Altman, The PRISMA Group, 2009). Neglecting this recommendation boosts the risk of upwardly biasing the magnitude of the calculated effects. Holding sample size equal, nonsignificant findings will be smaller in magnitude than significant findings. Rather than exclude nonsignificant findings, one should ideally use the meta-analysis to clarify the boundary conditions associated with positive versus negative results. If authors exclude nonsignificant findings, it is incumbent on them to offer a clear a priori rationale for doing so. As a consequence of their omission of many negative findings, Poston and Hanson's (2010) conclusions cannot be interpreted with confidence and may overstate the magnitude of the effects of PATI.

The Barnum Effect

Meehl (1956), crediting his University of Minnesota colleague Donald G. Patterson, coined the term P.T. Barnum effect, after the circus entrepreneur who famously said that he liked to "give a little something to everybody" in his acts. Today, most psychologists define the P.T. Barnum effect as individuals' acceptance of highbase rate but nonobvious personality descriptors, such as "You have a need for other people to like and admire you" and "At times you have serious doubts as to whether you have made the right decision" (Forer, 1949). A substantial body of research demonstrates that most individuals find Barnum descriptions subjectively compelling (Dickson & Kelly, 1985; Furnham & Schofield, 1987; Snyder, 1974) and that Barnum effects may account largely for widely held beliefs in the familiar but poorly supported personality "profiles" of adult children of alcoholics (Logue, Sher, & Frensch, 1992) and adult victims of child sexual abuse (Emery & Lilienfeld, 2004). Barnum descriptions presumably derive much of their effectiveness from the fact that most people "read" meaning into them, finding personal significance in vague and ambiguous statements (Dutton, 1988).

Poston and Hanson (2010) did not mention the Barnum effect except in one case, namely, as an exclusion criterion in their meta-analysis. Specifically, they noted that to be included in their meta-analysis, studies needed to "utilize authentic test results/data (i.e., no Barnum-type results)" (Poston & Hanson, 2010, p. 205). But they do not address or even acknowledge the possibility that client improvement in the PATI studies they reviewed might have been obtained with generalized personality feedback that applies to most or virtually all individuals (see also Goodyear, 1990; Roback, 1972; Wood et al., 2002). If so, it would not call into question the effectiveness of PATI, but it would suggest that any positive effects of PATI could instead be obtained with any number of nonspecific interventions that give clients a sense of understanding of their problems—even if this understanding is erroneous. It would also suggest that Poston and Hanson's (2010) conclusion

that "psychological assessment procedures—when combined with personalized [emphasis added], collaborative, and highly involving test feedback—have positive, clinically meaningful effects on treatment" (p. 203) requires qualification and revision because Barnum feedback is generalized to all individuals, not personalized. Goodyear (1990) conjectured that "if the counselor's goal is to help the client make changes in his or her life, it may not be necessary for the feedback he or she receives to be 'accurate'" (p. 247). Goodyear's argument underscores the possibility that PATI could operate as an "assessment placebo," lending clients a sense of comprehension, control, and meaning that may in some cases be illusory.

Given that the acquisition of a sense of control and self-efficacy has long been regarded as a key nonspecific factor in psychotherapy (e.g., Bandura, 1977), this possibility must be considered seriously. Indeed, in his classic writings on common factors in psychotherapy, Frank (1971) proposed that "provision of new information concerning the nature and sources of the patient's problems and possible alternative ways of dealing with them" (p. 309) is one of the core conditions for therapeutic improvement. Interestingly, in their discussion of the mechanisms underlying the ostensible positive effects of therapeutic assessment, Finn and Tonsanger (1997) similarly argued that "a collaborative approach enhances the sense of efficacy and self-discovery that can be derived from an assessment in that clients, with the aid of the assessor, find their own new words for and new understandings of problems in living" (p. 382).

To our knowledge, only one study has examined whether Barnum feedback enhances treatment outcomes (see also Sakamoto, Miura, Sakamoto, & Mori, 2000, for evidence that bogus personality feedback can influence participants' behavior toward confederates and reported well -being). In a study not mentioned by Poston and Hanson (2010); Halperin and Snyder (1979) asked 48 women with a snake phobia to complete a personality questionnaire and then randomly assigned them to three conditions: (a) a control group that received no treatment, (b) a group that received treatment only (systematic desensitization), and (c) a group that received both a Barnum description (which they were falsely led to believe was genuine and based on their questionnaire responses) and treatment. The Barnum description, which was identical for all participants in the third group, was intended to provide them with a sense that they would respond well to treatment and included such statements as "You have inner resources enabling you to learn effective means of adapting to the environment" and "You have a great deal of unused potential that you have not yet turned to your advantage" (p. 142). On the two outcome indicators, a questionnaire measure of snake fear and a behavioral avoidance test of snake fear, participants in the third (Barnum plus treatment) group evidenced significantly greater improvement than did individuals in the other two groups. Calculation of effect sizes from the means and standard deviations provided in Halperin and Snyder (1979, p. 144) reveals a Cohen's d of .55 for the comparison of the Barnum plus treatment group versus the treatment only group on the behavioral avoidance test, a value slightly higher than that reported by Poston and Hanson for the mean effect of PATI (on the questionnaire measure of snake fear, the pretest scores for the three groups differed significantly, but Halperin & Snyder, 1979, did not

provide the adjusted means).² Halperin and Snyder's findings are consistent with the possibility that Barnum personality feedback may be effective and, perhaps, just as effective as actual personality feedback in enhancing psychotherapy outcomes.

How plausible is it that Barnum effects could account for the positive effects of PATI reported by Poston and Hanson (2010)? On the one hand, most research suggests that although most individuals find Barnum descriptors persuasive, they can pick out accurate feedback based on their personality test (e.g., California Psychological Inventory; Gough, 1987) results from Barnum feedback at better than chance levels (Greene, Harris, & Macon, 1979; Harris & Greene, 1984; but see Dies, 1972, and Sundberg, 1955, for different conclusions). Moreover, after asking participants to complete the Eysenck Personality Inventory (Eysenck & Eysenck, 1964), Layne and Ally (1980) found that participants were more likely to believe personality feedback when it was genuine than when it was bogus. These results suggest that PATI based on Barnum feedback may be less persuasive to clients than PATI based on actual personality feedback (Furnham & Schofield, 1987).

On the other hand, several characteristics of both the modal therapeutic setting and the modal therapy client raise the distinct possibility that individuals in psychotherapy may be especially receptive to Barnum feedback. For example, individuals tend to find Barnum descriptors especially persuasive when they believe these descriptors are tailored specifically for them (as opposed to people with whom they share some characteristics or to people in general; Dickson & Kelly, 1985; Snyder & Larsen, 1972). Some research further suggests that Barnum statements are more persuasive when delivered by assessors with high levels of prestige, as opposed to low levels of prestige (Collins, Dmitruk, & Ranney, 1977; Halperin et al., 1976; but see Dmitruk, Collins, & Clinger, 1973). Halperin, Snyder, Shenkel, and Houston (1976) reported a significant interaction as negative (as opposed to positive) Barnum feedback- delivered after taking a Rorschach Inkblot Test-was more accepted by clients if it derived from a diagnostician with moderate or high status than a diagnostician with low status. Because (a) most clients presumably perceive their psychotherapists as being of high status and (b) the personality feedback delivered to clients in psychotherapy is usually at least partially negative, these results suggest that clients may find Barnum feedback particularly believable when delivered by therapist.

In addition, individuals with an external (as opposed to an internal) locus of control, which characterizes most individuals with who are in psychological treatment, including those with depression (Benassi, Sweeney, & Dufour, 1988) and anxiety (Archer, 1979), appear to be especially open to accepting Barnum feedback (Orpen & Jamotte, 1975; Snyder & Larsen, 1972; but see Fichten & Sunerton, 1983). Individuals with high levels of emotional insecurity, a trait that again describes many individuals in psychotherapy, are also especially receptive to Barnum feedback (Snyder & Clair, 1979).

Given the findings of Halperin and Snyder (1979) and research on the predictors of acceptance of Barnum descriptions, which suggest that many of the characteristics of individuals who tend to accept Barnum descriptions dovetail with those of individuals in therapy (see Dickson & Kelly, 1985, and Furnham & Schofield, 1987, for reviews), the onus of proof falls on Poston and Hanson (2010) to demonstrate that Barnum effects do not account for their

findings. Putting it a bit differently, it is up to the proponents of PATI to show that the effects of this technique exceed those of a Barnum or "placebo" intervention; if they do not, it might suggest that the effects of PATI have little or nothing to do with accurate assessment feedback. Such a comparison might be thought of as roughly analogous to the comparison of a form of psychotherapy, such as Beck's cognitive-behavioral therapy (CBT), with an "attention-placebo" control condition, which controls at least partly for the nonspecific effects of interpretation and support from an insightful and caring individual. CBT has passed this hurdle in a number of studies (e.g., see Taylor, 1996), so it is not unreasonable to expect PATI to do so as well.

Discussion

We concur with Poston and Hanson (2010) that therapists should routinely share the results of psychological assessments with clients in an open and collaborative fashion (e.g., Finn, 2007; Fischer, 1972) and that PATI is a promising technique that warrants further empirical examination. We hope that our commentary will not be interpreted as implying that PATI is ineffective, although we believe that Poston and Hanson's conclusion that it is effective is premature. Specifically, by not separating the effects of PATI from those of psychological treatments more broadly and by omitting a large number (17, by our count) of nonsignificant results, Poston and Hanson appear to have drawn unjustified inferences regarding the treatment utility of PATI. Moreover, Poston and Hanson neglect to consider the possibility that Barnum effects account for the effectiveness of PATI, raising the possibility that PATI may sometimes "work" by lending clients an illusory sense of understanding.

In our view, it will be difficult to establish the specific efficacy (i.e., efficacy above and beyond nonspecific factors; see Chambless & Hollón, 1998) of PATI without a Barnum feedback control condition. Some authors have raised legitimate ethical concerns regarding the use of Barnum feedback in research studies, especially those involving actual clients (Dana & Graham, 1976; Finn & Tonsager, 1992; see also Beins, 1993). We are sympathetic to these concerns and do not intend to address, let alone resolve, this complex issue here. Nevertheless, if studies using Barnum control conditions cannot be performed for ethical reasons, it is incumbent on proponents of PATI to temper their conclusions regarding the mechanisms underlying its therapeutic effects.

One partial, albeit imperfect, solution to this problem is to evaluate the treatment utility of psychological assessment with

² One investigation involved an examination of whether Barnum descriptions increase individuals' faith in psychological testing and assessor skill. Handelsman and McLain (1988) recruited 24 pairs (24 men, 24 women), half of whom were intimate couples and half of whom were strangers. One member of each pair (the "participant") completed the Rorschach Inkblot Test and then received written Barnum feedback (which they were falsely led to believe was based on their Rorschach responses) in the presence of the other member of the pair (the "observer"). Following the feedback, participants reported significantly higher levels of faith in psychological tests and reported that the assessor had significantly higher levels of skill. These results suggest that accurate personality feedback may not be needed to boost individuals' belief in the validity of psychological tests.

alternative paradigms. One elegant method of demonstrating the potential therapeutic value of assessment is the manipulated assessment design (Hayes et al., 1987). Initially proposed by Meehl (1959), this design treats therapists as "participants," randomly assigning them to receive either certain assessment information (e.g., the results of an MMPI-2 or Rorschach) or no assessment information. The finding that therapists assigned accurate assessment information obtain superior therapeutic outcomes relative to therapists who were not would tentatively support the treatment utility of assessment. Surprisingly, only one published study has involved the use of a manipulated assessment design to examine the therapeutic value of clinical assessment. Lima et al. (2005) randomly assigned 134 adult patients seeking treatment in a university training clinic to two conditions: one in which their therapists were granted access to their MMPI-2 results and another in which their therapists were not. Across all outcome indicators, including measures of illness severity and symptom improvement, access to MMPI-2 information did not lead to enhanced treatment outcomes. Lima et al.'s results do not exclude the possibility that the MMPI-2 possesses treatment utility, but they suggest that further investigation of the conditions under which the MMPI-2 might enhance treatment outcomes is necessary. Positive results for manipulated assessment designs like those used by Lima et al. might still be accounted for by Barnum effects, but such results, in conjunction with positive results for PATI, would buttress the contention that psychological assessment can be therapeutically useful. In science, conclusions tend to be most robust when they derive from diverse, albeit imperfect, sources of evidence (Shadish, Cook, & Campbell, 2002).

Another important avenue of research would involve dismantling designs (Kazdin, 1994) aimed at decomposing some of the potentially effective components of PATI. Most forms of PATI comprise multiple components, including initial discussions of the goals of assessment with clients, test taking, sharing test results and interpretations with clients, responding to client questions and reactions to the test results, providing written feedback to clients, and so on. Moreover, in the form of PATI advocated by Finn (1996), namely, therapeutic assessment, therapists are encouraged to begin with interpretations that are more consistent with clients' self-views before moving on to those that challenge clients' self-views

Nevertheless, it is not known which, if any, of these components are essential components of a full PATI package. Nor is it known how much the presumed effects of PATI are attributable to the use of psychological assessment per se as, opposed to interaction with a warm, understanding, and perceptive individual who can provide clients with a coherent narrative for conceptualizing their life problems (e.g., Frank, 1971). Indeed, it remains to be seen whether PATI requires the inclusion of a formal assessment procedure (e.g., MMPI-2, Rorschach) at all, as many of its apparent benefits may derive from the construction of helpful life narratives and the clarification of clients' self-concepts in conjunction with a therapist (see Finn & Tonsager, 1997). Teasing apart the diverse components of PATI should help to address the crucial but still largely unresolved question of when psychological assessment does—and does not-enhance treatment outcomes.

References

- Ackerman, S. J., Hilsenroth, M. J., Baity, M. R., & Blagys, M. D. (2000). Interaction of therapeutic process and alliance during psychological assessment. *Journal of Personality Assessment*, 75, 82-109. doi: 10.1207/S15327752JPA7501_7
- Allen, A., Montgomery, M., Tubman, J., Frazier, L., & Escovar, L. (2003).
 The effects of assessment feedback on rapport-building and self-enhancement processes. *Journal of Mental Health Counseling*, 25, 165–182.
- Archer, R. P. (1979). Relationship between locus of control and anxiety. Journal of Personality Assessment, 43, 617-626. doi:10.1207/s15327752jpa4306_10
- Arkowitz, H., & Miller, W. R. (2008). Learning, applying, and extending motivational interviewing. In H. Arkowitz, H. A. Westra, W. R. Miller,
 & S. Rollnick (Eds.), Motivational interviewing in the treatment of psychological problems (pp. 1-25). New York, NY: Guilford Press.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215. doi:10.1037/0033-295X.84.2.191
- Beins, B. C. (1993). Using the Barnum effect to teach about ethics and deception in research. *Teaching of Psychology*, 20, 33-35. doi:10.1207/ s15328023top2001_6
- Benassi, V. A., Sweeney, P. D., & Dufour, C. L. (1988). Is there a relationship between locus of control orientation and depression? *Journal of Abnormal Psychology*, 97, 357-367. doi:10.1037/0021-843X.97.3.357
- Boutron, I., Dutton, S., Ravaud, P., & Altman, D. G. (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA: Journal of the American Medical Association*, 303, 2058–2064. doi:10.1001/jama.2010.651
- Butcher, J. N., & Rouse, S. V. (1996). Personality: Individual differences and clinical assessment. Annual Review of Psychology, 47, 87-111. doi:10.1146/annurev.psych.47.1.87
- Chambless, D. L., & Hollon, S. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66, 7-18. doi:10.1037/0022-006X.66.1.7
- Cohen, J. (1969). Statistical power analysis for the behavioral sciences. New York, NY: Academic Press.
- Cohen, J. (1988). Statistical power for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.
- Collins, R. W., Dmitruk, V. M., & Ranney, J. T. (1977). Personal validation: Some empirical and ethical considerations. *Journal of Consulting and Clinical Psychology*, 45, 70-77.
- Dana, R. H., & Graham, D. (1976). Feedback of client-relevant information and clinical practice. *Journal of Personality Assessment*, 40, 464-469. doi:10.1207/s15327752jpa4005_3
- Dickson, D. H., & Kelly, I. W. (1985). The "Barnum effect" in personality assessment: A review of the literature. *Psychological Reports*, 57, 367–382.
- Dies, R. R. (1972). Personal gullibility or pseudodiagnosis: A further test of the "fallacy of personal validation." *Journal of Clinical Psychology*, 28, 47–50. doi:10.1002/1097-4679(197201)28:1<47::AID-JCLP2270280116>3.0.CO;2-K
- Dmitruk, V. M., Collins, R. W., & Clinger, D. L. (1973). The "Barnum effect" and acceptance of negative personal evaluation. *Journal of Consulting and Clinical Psychology*, 41, 192–194. doi:10.1037/h0035106
- Dutton, D. L. (1988). The cold reading technique. Experientia, 44, 326–332. doi:10.1007/BF01961271
- Emery, C., & Lilienfeld, S. O. (2004). The validity of child sexual abuse survivor checklists in the popular psychology literature: A Barnum effect? Professional Psychology: Science and Practice, 35, 268-274. doi:10.1037/0735-7028.35.3.268
- Eysenck, H. J., & Eysenck, S. B. G. (1964). Manual of the Eysenck Personality Inventory. London, England: University of London.

- Fichten, C. S., & Sunerton, B. (1983). Popular horoscopes and the "Barnum effect." *Journal of Psychology*, 114, 123–134. doi:10.1080/00223980.1983.9915405
- Finn, S. E. (1996). Manual for using the MMPI-2 as a therapeutic intervention. Minneapolis, MN: University of Minnesota Press.
- Finn, S. E. (2007). In our clients' shoes: Theory and techniques of therapeutic assessment. Mahwah, NJ: Erlbaum.
- Finn, S. E., & Tonsager, M. E. (1992). Therapeutic effects of providing MMPI-2 test feedback to college students awaiting therapy. *Psychological Assessment*, 4, 278–287. doi:10.1037/1040-3590.4.3.278
- Finn, S. E., & Tonsager, M. E. (1997). Information-gathering and therapeutic models of assessment: Complementary paradigms. *Psychological Assessment*, 9, 374–385. doi:10.1037/1040-3590.9.4.374
- Fischer, C. T. (1972). Paradigm changes which allow sharing of results. *Professional Psychology*, 3, 364–369. doi:10.1037/h0033991
- Folds, J. H., & Gazda, G. M. (1966). A comparison of the effectiveness and efficiency of three methods of test interpretation. *Journal of Counseling Psychology*, 13, 318–324. doi:10.1037/h0023708
- Forer, B. R. (1949). The fallacy of personal validation: A classroom demonstration of gullibility. *Journal of Abnormal and Social Psychol*ogy, 44, 118-123. doi:10.1037/h0059240
- Frank, J. D. (1971). Therapeutic factors in psychotherapy. American Journal of Psychotherapy, 25, 350-361.
- Furnham, A., & Schofield, S. (1987). Accepting personality test feedback: A review of the Barnum effect. Current Psychological Research & Reviews, 6, 162-178. doi:10.1007/BF02686623
- Garb, H. N., Lilienfeld, S. O., Nezworski, M. T., Wood, J. M., & O'Donohue, W. T. (2009). Can quality improvement processes help psychological assessment meet the demands of evidence-based practice? Scientific Review of Mental Health Practice, 7, 17–25.
- Goodyear, R. K. (1990). Research on the effects of test interpretation: A review. The Counseling Psychologist, 18, 240-257. doi:10.1177/ 0011000090182006
- Gough, H. G. (1987). California Psychological Inventory administrator's guide. Palo Alto, CA: Consulting Psychologists Press.
- Greene, R. L., Harris, M. E., & Macon, R. S. (1979). Another look at personal validation. *Journal of Personality Assessment*, 43, 419-423. doi:10.1207/s15327752jpa4304_16
- Halperin, K., Snyder, C. R., Shenkel, R. J., & Houston, B. K. (1976).
 Effects of source status and message favorability on acceptance of personality feedback. *Journal of Applied Psychology*, 61, 85–88. doi: 10.1037/0021-9010.61.1.85
- Halperin, K. M., & Snyder, C. R. (1979). Effects of enhanced psychological test feedback on treatment outcomes: Therapeutic implications of the Barnum effect. *Journal of Consulting and Clinical Psychology*, 47, 140-146. doi:10.1037/0022-006X.47.1.140
- Handelsman, M. M., & McLain, J. (1988). The Barnum effect in couples: Effects of intimacy, involvement, and sex on acceptance of generalized personality feedback. *Journal of Clinical Psychology*, 44, 430–444. doi:10.1002/1097-4679(198805)44:3<430::AID-JCLP2270440319>3.0 .CO;2-V
- Hanson, W. E., Claiborn, C. D., & Kerr, B. (1997). Differential effects of two test-interpretation styles in counseling: A field study. *Journal of Counseling Psychology*, 44, 400-405. doi:10.1037/0022-0167.44.4.400
- Harkness, A. R., & Lilienfeld, S. O. (1997). Individual differences science for treatment planning: Personality traits. *Psychological Assessment*, 9, 349-360. doi:10.1037/1040-3590.9.4.349
- Harris, M. E., & Greene, R. L. (1984). Students' perception of actual, trivial, and inaccurate personality feedback. *Journal of Personality As*sessment, 48, 179-184. doi:10.1207/s15327752jpa4802_13
- Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment. American Psychologist, 42, 963–974. doi:10.1037/0003-066X.42.11.963
- Hewitt, C. E., Mitchell, N., & Torgerson, D. J. (2008). Heed the data when

- the results are not significant. *British Medical Journal*, 336, 23-25. doi:10.1136/bmj.39379.359560.AD
- Hilsenroth, M. J., Peters, E. J., & Ackerman, S. J. (2004). The development of therapeutic alliance during psychological assessment: Patient and therapist perspectives across treatment. *Journal of Personality Assess*ment, 83, 332–344. doi:10.1207/s15327752jpa8303_14
- Hunsley, J. (2003). Introduction to the special section on incremental validity and utility in clinical assessment. *Psychological Assessment, 15*, 443–445. doi:10.1037/1040-3590.15.4.443
- Jobes, D. A., & Drozd, J. F. (2004). The CAMS approach to working with suicidal patients. *Journal of Contemporary Psychotherapy*, 34, 73–85. doi:10.1023/B:JOCP.000010914.98781.6a
- Jobes, D. A., Wong, S. A., Conrad, A. K., Drozd, J. F., & Neal-Walden, T. (2005). The collaborative assessment and management of suicidality versus treatment as usual: A retrospective study. Suicide and Life-Threatening Behavior, 35, 483-497. doi:10.1521/suli.2005.35.5.483
- Kazdin, A. E. (1994). Methodology, design, and evaluation in psychotherapy research. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 19-71). New York, NY: Wiley.
- Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry*, 44, 1116–1129. doi:10.1111/1469-7610.00195
- 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. doi:10.1093/clipsy.bpg025
- Layne, C., & Ally, G. (1980). How and why people accept personality feedback. *Journal of Personality Assessment*, 44, 541-546. doi:10.1207/ s15327752jpa4405_16
- Lima, E. N., Stanley, S., Kaboski, B., Reitzel, L. R., Richey, J. 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. doi:10.1037/1040-3590.17.4.462
- Logue, M. B., Sher, K. J., & Frensch, P. A. (1992). Purported characteristics of COAs: A possible "Barnum effect." Professional Psychology: Research and Practice, 23, 226-232. doi:10.1037/0735-7028.23.3.226
- Luborsky, L., & Crits-Christoph, P. (1997). Understanding transference (2nd ed.). Washington, DC: American Psychological Association.
- Meehl, P. E. (1956). Wanted—A good cookbook. *American Psychologist*, 11, 263–272. doi:10.1037/h0044164
- Meehl, P. E. (1959). Some ruminations on the validation of clinical procedures. *Canadian Journal of Psychology, 13*, 102–128. doi:10.1037/h0083769
- Miller, W. R., Benefield, R. G., & Tonigan, J. S. (1993). Enhancing motivation for change in problem drinking: A controlled comparison of two therapist styles. *Journal of Consulting and Clinical Psychology*, 61, 455-461. doi:10.1037/0022-006X.61.3.455
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLOS Med, 6, e1000097.
- Nelson-Gray, R. O. (2003). Treatment utility of psychological assessment. *Psychological Assessment, 15,* 521–531.
- Newman, M. L., & Greenway, P. (1997). Therapeutic effects of providing MMPI-2 test feedback to clients at a university counseling service: A collaborative approach. *Psychological Assessment*, 9, 122–131.
- Orpen, C., & Jamotte, A. (1975). The acceptance of generalized personality interpretations. *Journal of Social Psychology*, 38, 147–148.
- Poston, J. M., & Hanson, W. E. (2010). Meta-analysis of psychological assessment as a therapeutic intervention. *Psychological Assessment*, 22, 203-212.
- Roback, H. B. (1972). Inexact personality interpretations and behavior change. *The Ontario Psychologist*, 4, 92–99.

- Sakamoto, A., Miura, S., Sakamoto, K., & Mori, T. (2000). Popular psychological tests and self-fulfilling prophecy: An experiment of Japanese undergraduate students. Asian Journal of Social Psychology, 3, 107-124.
- Shadish, W., Cook, T., & Campbell, D. (2002). Experimental and quasiexperimental designs for generalized causal inference. Boston, MA: Houghton Mifflin.
- Smith, J. D., Handler, L., & Nash, M. R. (2010). Therapeutic assessment for preadolescent boys with oppositional defiant disorder: A replicated single-case time-series design. *Psychological Assessment*, 22, 593-602.
- Snyder, C. R. (1974). Why horoscopes are true: The effect of specificity on acceptance of astrological interpretations. *Journal of Clinical Psychol*ogy, 30, 577-580.
- Snyder, C. R., & Clair, M. S. (1979). Does insecurity breed acceptance? Effects of trait and situational insecurity on acceptance of positive and negative diagnostic feedback. *Journal of Consulting and Clinical Psychology*, 46, 1493–1495.
- Snyder, C. R., & Larson, G. R. (1972). A further look at student acceptance of general personality interpretations. *Journal of Consulting and Clini*cal Psychology, 38, 384-388.
- Sundberg, N. D. (1955). The acceptability of "fake" versus "bona fide"

- personality test interpretations. Journal of Abnormal and Social Psychology, 50, 145-147.
- Taylor, S. (1996). Meta-analysis of cognitive-behavioral treatments for social phobia. *Journal of Behavior Therapy and Experimental Psychi*atry, 27, 1–9.
- Valentine, J. C., & Cooper, H. (2003). Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes. Washington, DC: What Works Clearinghouse.
- Wild, T. C., Cunningham, J. A., & Roberts, A. B. (2007). Controlled study of brief personalized assessment-feedback for drinkers interested in self-help. Addiction, 102, 241–250.
- Wood, J. M., Garb, H. N., Lilienfeld, S. O., & Nezworski, M. T. (2002). Clinical assessment of personality. *Annual Review of Psychology*, 53, 519-543.
- Woodworth, R. S. (1918). Dynamic psychology. New York, NY: Columbia University Press.

Received July 6, 2010
Revision received June 9, 2011
Accepted June 10, 2011

Copyright of Psychological Assessment is the property of American Psychological Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.