Psychological Assessment
2010, Vol. 22, No. 2, 203–212

1040-3590/10/$12.00 DOI: 10.1037/a0018679

Meta-Analysis of Psychological Assessment as a Therapeutic Intervention

John M. Poston and William E. Hanson
Purdue University

This study entails the use of meta-analytic techniques to calculate and analyze 18 independent and 52 nonindependent effect sizes across 17 published studies of psychological assessment as a therapeutic intervention. In this sample of studies, which involves 1,496 participants, a significant overall Cohen's *d* effect size of 0.423 (95% CI [0.321, 0.525]) was found, whereby 66% of treatment group means fell above the control and comparison group means. When categorical variables were taken into account, significant treatment group effects were found for therapy process variables ($d = 1.117$, [0.679, 1.555]), therapy outcomes ($d = 0.367$, [0.256, 0.478]), and combined process/outcome variables ($d = 0.547$, [0.193, 0.901]). These findings appear to be robust on the basis of fail-safe *N* calculations. Taken together, they suggest 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. They also have important implications for assessment practice, training, and policy making, as well as future research, which are discussed in the conclusion of the article.

Keywords: assessment utility, treatment validity, meta-analysis, therapeutic assessment, collaborative assessment

Assessment research that moves beyond scale development and psychometrics has long been called for in the social sciences (Haynes, Nelson, & Jarrett, 1987; Meyer et al., 2001). There is a particularly strong need for research on the treatment validity, or clinical utility, of assessment procedures in therapy. The emergence of therapeutic models of assessment, such as *therapeutic assessment* (Finn, 1996, 2007; Finn & Tonsager, 1997), provides a conceptual framework and research methodology for studying this issue directly. These models view assessment-related processes and procedures as "interventions" in their own right, as opposed to precursors or adjuncts to treatment. Although a number of independent studies have been conducted on these models, the results have not yet been scrutinized systematically or, for that matter, empirically vis-a-vis meta-analysis. The purpose of this study is to do just that, namely, to identify studies published on the topic, calculate and analyze associated Cohen's *d* effect sizes, and in turn determine the practical significance of identified differences.

Development of Therapeutic Models of Assessment

Gelso and Fretz (2001) asserted, "Our research in the coming decades will help decide whether psychological assessment is our dodo bird or a phoenix rising from the 'ashes' of the critiques of recent decades" (p. 400). Accordingly, experts agree that psychological assessment is at a pivotal point in its history. Although the ability to develop, administer, and interpret psychological tests is a distinguishing characteristic of psychologists' professional identity, its perceived usefulness has waxed and waned over the years. For example, it was frowned upon, by some, during the humanistic movement of the 1950s and 1960s (Gelso & Fretz, 2001; Goldman, 1972). Popular theoretical orientations at the time downplayed testing and diagnosis. Later, in the early 1970s, Goldman (1972) referred to testing and therapy as "the marriage that failed" (p. 213). More recently, managed care curtailed the use of tests in clinical work (Eisman et al., 2000).

Ironically, therapeutic models of assessment grew out of the humanistic movement directly, the very movement that lamented and questioned testing (Finn & Tonsager, 1997, 2002). A different spin was put on testing (and the feedback process), with some psychologists viewing it as a potentially therapeutic relational experience rather than a sterile reductionist practice (Riddle, Byers, & Grimesey, 2002). The long-held belief that sharing test results with clients was harmful was thus called into question (e.g., Fischer, 1972). In the early 1990s, ethical guidelines were revised, requiring test results to be shared with clients, with, of course, a few notable exceptions (e.g., forensic evaluations; American Psychological Association, 2002). Although uniform adherence to these guidelines is lacking (Curry & Hanson, in press; Smith, Wiggins, & Gorske, 2007), such changes constituted a major shift, paradigm-wise, in contemporary assessment-related attitudes, values, and practices, as well as research foci.

From this movement, various therapeutic models of assessment emerged (Finn, 1996, 2007; Fischer, 1994; Gorske & Smith, 2008). Though they have different names, most commonly they are called either *therapeutic assessment* (Ackerman, Hilsenroth, Baity, & Blagys, 2000; Callahan, Price, & Hilsenroth, 2003; Finn, 1996, 2003; Finn & Tonsager, 1992, 1997, 2002; Hilsenroth, Peters, & Ackerman, 2004; Michel, 2002; Newman & Greenway, 1997; Tharinger, Finn, Wilkinson, & Schaber, 2007; Wygant &

John M. Poston and William E. Hanson, Department of Educational Studies, Purdue University.

Correspondence concerning this article should be addressed to John M. Poston, Department of Educational Studies, Purdue University, 5108 BRNG, 100 North University Street, West Lafayette, IN 47907-2098. E-mail: jmposton@purdue.edu

Fleming, 2008) or *collaborative/individualized assessment* (Fischer, 2000, 2006; Purves, 2002; Riddle et al., 2002). Less prominent, though clearly related, models include *collaborative consultation to psychotherapy* (Engelman & Frankel, 2002) and *brief personalized assessment feedback* (Wild, Cunningham, & Roberts, 2007). The most prevalent contemporary model, therapeutic assessment, is a brief, highly structured, theoretically and empirically based approach to assessment and testing. It was developed by Stephen Finn and his colleagues, most notably Mary Tonsager, and was influenced greatly by humanistic and self psychology and by the early writings of Harry Stack Sullivan, Connie Fischer, and Richard Dana. Key aspects of this approach include (a) helping clients generate questions they would like answered/addressed by the assessment and testing, (b) collecting background information related to their questions, (c) exploring past assessment- and/or testing-based hurts, (d) involving clients collaboratively in discussing (and making sense of) the results, and (e) answering, as much as possible, clients' initial questions. For additional, specific details about this approach, see Finn (1996, 2007).

Although other therapeutic models of assessment differ from Finn's (1996, 2007) model in subtle and nuanced ways, they have certain unifying commonalities. Finn and Tonsager (1997) summarized and highlighted three of them, including "(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" (p. 378). Thus, these models complement traditional *information-gathering* models, adding a new and potentially therapeutic piece to the assessment and treatment puzzle.

However, the question remains: Are these models truly therapeutic? To date, this question has not been answered empirically, at least not satisfactorily. There is no consensus on the matter, even among programmatic researchers. Although many believe they are indeed therapeutic, there are those who do not (cf. Tinsley & Chu, 1999). As noted earlier, a number of studies have been conducted and published on these models. However, results of individual studies have not been aggregated across studies; thus, the overall therapeutic value of these models remains largely unknown.

In 1988, Oliver and Spokane conducted a meta-analysis on career-counseling interventions, including interventions related to assessment and test feedback. Whiston, Brecheisen, and Stephens (2003) followed up this study with a subsequent meta-analysis analyzing a similar literature base. However, only a handful of studies in these two meta-analyses related specifically to the issue at hand. Additional empirical reviews of this literature have also been published (cf. Meyer et al., 2001; Whiston, Sexton, & Lasoff, 1998), though most are conceptual or narrative in nature (Claiborn, Goodyear, & Horner, 2001; Finn & Tonsager, 1997; Goodyear, 1990; Riddle et al., 2002; Tinsley & Chu, 1999). In 2001, Meyer et al. examined the assessment and testing literature empirically. These authors conducted a meta-analysis on assessment and the validity of psychological testing as part of the American Psychological Association's assessment workgroup. They concluded, on the basis of their analysis of 69 studies, that assessment and testing practices are valid, producing positive and meaningful effects comparable to those of medical tests. Here again, as with Oliver and Spokane (1988) and Whiston et al. (2003), only a couple studies related specifically to assessment and test feedback, so a major gap in the literature still exists. Therefore, a comprehensive and inclusive meta-analysis of studies on this topic is warranted at this time. A meta-analysis of this sort would (a) address the general lack of treatment validity, or clinical utility, studies on assessment and testing practices; (b) satisfy an ethical imperative (Rosenthal, 2008); and (c) provide much-needed data regarding the practical and clinical significance of this growing (and potentially promising) body of research.

The Present Study

The present study examines the efficacy of psychological assessment as an intervention—in all its nuanced varieties—via meta-analytic techniques. Meyer et al. (2001) described the primary purposes of assessment as (a) describing client functioning, (b) refining clinical impressions of a client, (c) identifying therapeutic needs, (d) aiding in differential diagnosis, (e) monitoring treatment process, (f) minimizing legal liability, and (g) using the assessment and testing process as an intervention. So, to be clear, the last issue is the focus of this meta-analysis. In as much, it addresses two principal research questions:

1. Does psychological testing, when combined with personalized, collaborative feedback of some sort, affect treatment processes and outcomes and, ultimately, benefit clients?

2. What variables account for significant variances across studies?

We hypothesize that, on the basis of findings from multiple individual studies, as well as theoretical/conceptual writings on the topic:

1. Psychological testing, when conducted in this manner, will significantly affect treatment-related processes and outcomes, in practical and clinically meaningful ways.

2. Relative to control and comparison group means, treatment group means will be significantly higher.

3. Certain variables, such as type of dependent variable (e.g., process vs. outcome), will account for a significant percentage of variance across studies.

Method

Sample of Studies and Effect Sizes

Seventeen published studies and 18 independent and 52 nonindependent Cohen's *d* effect sizes were included in this meta-analysis. The studies were published between 1954 and 2007 in the following journals: *Addiction* ($n = 1$), *Journal of Career Assessment* ($n = 2$), *Journal of Consulting and Clinical Psychology* ($n = 1$), *Journal of Counseling & Development* ($n = 1$), *Journal of Counseling Psychology* ($n = 6$), *Journal of Mental Health Counseling* ($n = 1$), *Journal of Personality Assessment* ($n = 2$), *Psychological Assessment* ($n = 2$), and *Suicide & Life-Threatening Behavior* ($n = 1$).

Operational Definitions and Inclusion Criteria

Given the complexities of searching a large, disparate literature base, a number of steps were taken to ensure the search was systematic. First, operational definitions were used to clarify, and pinpoint, constructs of interest. Specifically, *psychological assessment as a therapeutic intervention* was defined broadly as the process of completing any formal psychological test/measure and receiving feedback on the results. Thus, for this meta-analysis, studies that included only the "bare bones" of assessment as an intervention (e.g., testing and some form/version of test feedback [with therapeutic intent]) were also included—not just studies that examined, for example, the efficacy of therapeutic assessment. Additionally, *therapeutic benefit* was defined as any dependent variable designed to demonstrate potential client improvement or enhanced therapy process.

Second, five inclusion criteria were used to identify studies. In order to be included, a study had to (a) address one of the research questions; (b) be published in English in a peer-reviewed journal; (c) utilize an experimental design suitable for calculating one or more Cohen's *d* effect sizes; (d) measure some aspect of therapeutic benefit, either process- or outcome-related; and (e) utilize authentic test results/data (i.e., no Barnum-type results). Studies not meeting these criteria were excluded, including conference presentations, book chapters, and the like. Dissertations were also excluded, because it was not possible to locate all of them and determine their inclusion eligibility.

Literature Search Procedures

Between September 2008 and February 2009, the social sciences literature was searched multiple times for potential studies. Specifically, *assessment utility*, *therapeutic assessment*, *collaborative assessment*, *test feedback*, *assessment feedback*, and *test interpretation* were entered as search terms in the PsychINFO database. These terms resulted in 1,394 matched articles. The title and abstract of each article was reviewed, using the aforementioned inclusion criteria as a guide. On the basis of this initial step, 12 studies were identified and included. As a second step, reference lists of these studies were back-checked for potentially undiscovered studies, as were those of previously published literature reviews on the topic. This step resulted in the identification of four more studies for inclusion. As a final step, another study was added on the basis of John M. Poston's and William E. Hanson's familiarity with the literature. As a result of these three steps, 17 studies were identified, coded, and included here.

The 17 studies were coded prior to calculating Cohen's *d*. Specifically, John M. Poston and William E. Hanson discussed coding variables and categories. Although the variables and categories are relatively clear cut and straightforward, all ambiguities were discussed and clarified until consensus was reached on the constructs of interest. Then the date of publication; author; total sample size; sample sizes of treatment and control/comparison groups; independent and dependent variable(s); and effect size(s), either reported or calculated post hoc, were coded. Categorical variables of effect size calculation method, research design, and type of dependent variable (e.g., process-oriented, outcome-oriented, process/outcome-oriented) were also coded. *Process* studies focused on within-session, face-to-face client/therapist interactions. *Outcome* studies focused on the effects of treatment, or a specific intervention, and/or treatment-associated changes. *Process/outcome* studies focused on aspects of both, that is, client/therapist interactions and the effects of treatment. These definitions were adopted from Hill, Nutt, and Jackson (1994). Table 1 summarizes the various codings, and Table 2 summarizes the sample characteristics.

Calculation of Cohen's *d* Effect Sizes

Issue of independence. For many of the studies, numerous Cohen's *d* effect sizes could be calculated from the data. Although results are reported individually, only one effect size per study was used in calculating the aggregate effect size. In these instances, the effect size used in calculating the aggregate effect size was the mean of the reported effect sizes, within each respective study. Table 1 shows the actual distribution of effect sizes (in the Dependent [*d*] column), only two of which are skewed. This "mean effect size" approach was used to maintain the assumption of independent data points and avoid distorting the data, as a function of inflated sample size (Lipsey & Wilson, 2001).

Primary calculations. In order to compare quantitative results between experimental and control/comparison groups, a standardized mean difference effect size was used. The standardized effect size was chosen due to the differing manner in which studies operationalized dependent variables of therapeutic benefit (e.g., symptom reduction, increase in self-understanding, strengthened working alliance). Accordingly, Cohen's *d* was used as a common calculation of standardized mean differences. The formula for this calculation is

$$d = \frac{M_E - M_C}{SD_{Pooled}}, \quad (1)$$

where M_E is the mean of the experimental group, M_C is the mean of the control group, and SD_{Pooled} is the pooled standard deviation of the two groups.

Missing data. For 11 studies, the means and standard deviations of the experimental and control groups were reported, thus allowing for calculation of Cohen's *d*. However, for the other six, these data were not reported; instead, only data that could be converted to *d* were reported. In instances of missing data, calculation procedures recommended by Lipsey and Wilson (2001) were followed to obtain comparable values for *d*. When studies reported an independent *t* test and sample sizes for the experimental and control groups ($n = 2$), the following algebraic equivalent to Cohen's *d* was calculated:

$$d = t\sqrt{\frac{n_1 + n_2}{n_1 n_2}}, \quad (2)$$

where *t* is the reported *t* score, and n_1 and n_2 are the sample sizes of the two groups being compared in the *t* test. Where an *F* ratio of a two-group analysis of variance (ANOVA) was reported ($n = 1$), the *F* value took the place of the *t* value in Equation 2 and was moved inside the square root and multiplied by the numerator to find the absolute value of *d*.

Table 1
Meta-Analytic Findings of Psychological Assessment as a Therapeutic Intervention

Study	*N*	*d*[a]	Variable(s)		
			Independent	Dependent (*d*)	Categorical
Ackerman et al. (2000)	128	0.420	Therapeutic assessment (TA) vs. information gathering (IG)	Premature termination (0.420)	3, 2, 2
Allen et al. (2003)	83	0.850	Feedback (FB) vs. no FB	Positive evaluation (0.730) Negative evaluation (0.648) Feedback satisfaction (0.750) Self-verification (1.514) Self-esteem (0.511) Self-liking (0.693) Self-understanding (1.499) Self-competence (0.456)	2, 2, 3
Finn & Tonsager (1992)	60	0.847	Assessment & FB vs. attention only	Symptomatology (0.764) Self-esteem (1.006) Hope (0.771)	2, 1, 2
Folds & Gazda (1966)	44	0.406	Individual vs. written FB	FB value for self-understanding (0.455) FB value for goals (0.357)	1, 2, 2
Hanson & Claiborn (2006)[b]	22	–0.211	Mixed delivered FB vs. mixed interactive FB	Session depth (–0.409) Counselor influence (–0.393) Feedback acceptance (0.078) Feedback helpfulness (–0.121)	1, 2, 3
Hanson & Claiborn (2006)[b]	24	0.258	Positive delivered FB vs. positive interactive FB	Session depth (0.293) Counselor influence (0.271) Feedback acceptance (0.203) Feedback helpfulness (0.263)	1, 2, 3
Hanson et al. (1997)	26	1.394	Interactive vs. delivered FB	Session depth (1.568) Counselor influence (1.219)	1, 2, 1
Hilsenroth et al. (2004)	42	1.022	TA model vs. IG model	Working alliance (1.022)	1, 2, 1
Holmes (1964)	78	0.633	Client-directed FB vs. written FB	Value of information (0.633)	1, 2, 2
Jobes et al. (2005)	37	0.266	Collaborative suicide assessment vs. treatment as usual	Sessions to resolution of suicidality (0.685) No. of hospitalizations (–0.030) Days hospitalized (0.473) No. of suicide attempts (–0.063)	1, 2, 2
Katz et al. (1999)	205	0.266	Assessment & FB vs. no assessment or FB	Career-related change (0.364) Change of career goal (0.177) Specificity of career goal (0.262) Certainty of career goal (0.261)	4, 1, 2
Luzzo & Day (1999)	77	0.615	Assessment & FB vs. no assessment or FB	Career decision making self-efficacy (0.619) Career beliefs: control (0.563) Career beliefs: responsibility (0.623) Career beliefs: working hard (0.655)	1, 1, 2
Miller et al. (1993)	42	0.543	Alcohol assessment & FB vs. no assessment or FB	Weekly consumption (0.542) Peak intoxication (0.560) Days drinking per week (0.527)	1, 1, 2
Newman & Greenway (1997)	60	0.354	Assessment & FB vs. assessment only	Symptomatology (0.441) Self-esteem (0.267)	1, 2, 2
Rogers (1954)	94	0.072	Self-evaluative FB vs. test-centered FB	Self-understanding (0.072)	1, 2, 2
Wild et al. (2007)	306	0.214	Alcohol assessment & FB vs. assessment only	Binge drinking in problem drinkers (0.214)	3, 2, 2
Worthington et al. (1995)	48	0.143	Face-to-face FB vs. written FB with couples	Dyadic consensus (0.000) Affectional expression (0.000) Dyadic satisfaction (0.297) Dyadic cohesion (–0.008)	1, 2, 2
Wright (1963)	120	0.677	Assessment & FB vs. assessment only	Self-rating accuracy (0.677)	2, 2, 2
Total	1,496				
M		0.423			

Note. Higher positive *d* values indicate greater therapeutic benefit from assessment. Categorical variables and codes: Cohen's *d* calculation method (numbered as reported in the Method section), research design (1 = *true control group*, 2 = *comparison group*), and type of dependent variable (1 = *process variable[s]*, 2 = *outcome variable[s]*, 3 = *both*).

[a] The reported value of *d* for each study is the average value of calculated effect sizes reported in the dependent variable(s) column. [b] The research design utilized in this study allowed for the calculation of two aggregate effect sizes for two independent samples.

Table 2
Sample Characteristics of 17 Studies

Study	*N*	Mean age (years)	Ethnicity (%)	Gender (%)	Client/nonclient
Ackerman et al. (2000)	128	28.0	NR	M = 42 F = 58	Client
Allen et al. (2003)	83	22.6	AA = 10 A/PI = 10 EA = 12 L = 69	M = 14 F = 86	Nonclient
Finn & Tonsager (1992)	60	23.3	NR	M = 30 F = 70	Client
Folds & Gazda (1966)	44	NR	NR	NR	Nonclient
Hanson & Claiborn (2006)	46	26.2	AA = 9 AI = 4 EA = 65 L = 15 NR = 7	M = 30 F = 70	Nonclient
Hanson et al. (1997)	26	NR	A/PI = 8 EA = 77 L = 15	M = 23 F = 77	Client
Hilsenroth et al. (2004)	42	30.6	NR	M = 33 F = 67	Client
Holmes (1964)	78	NR	NR	NR	Nonclient
Jobes et al. (2005)[a]	37	29.1	AA = 3 A/PI = 4 BR = 4 EA = 84 L = 6	M = 65 F = 35	Client
Katz et al. (1999)[a]	205	22	NR	M = 28 F = 72	Nonclient
Luzzo & Day (1999)[a]	77	18.4	AA = 5 AI = 2 EA = 54 L = 39	M = 35 F = 65	Nonclient
Miller et al. (1993)	42	40	NR	M = 57 F = 43	Client
Newman & Greenway (1997)	60	30	NR	M = 23 F = 77	Client
Rogers (1954)	94	NR	NR	NR	Nonclient
Wild et al. (2007)[a]	306	40.7	NR	M = 53 F = 47	Nonclient
Worthington et al. (1995)	48	NR	NR	NR	Nonclient
Wright (1963)	120	NR	NR	M = 43 F = 57	Nonclient
Total	1,496				

Note. NR = not reported; M = male; F = female; AA = African American; A/PI = Asian/Pacific Islander; EA = European American/White; L = Latino(a); AI = American Indian; BR = Biracial.
[a] Demographic information reported for the entire sample; however, these demographics may not coincide precisely with the composition of subsamples included in the present meta-analysis.

When studies reported a chi-square test with one degree of freedom ($n = 2$), without the means and standard deviations of the experimental and control groups, the following formula was used to approximate Cohen's *d*:

$$|d| = 2\sqrt{\frac{\chi^2}{N - \chi^2}}, \quad (3)$$

where *N* is the total sample size.

Lastly, when a study reported success rate percentages but not means and standard deviations for the experimental and control groups ($n = 1$), the following formula, based on arcsine transformations, was used to approximate Cohen's *d*:

$$d = \text{arcsine}(p_1) - \text{arcsine}(p_2), \quad (4)$$

where p_1 and p_2 are the respective success rates of groups examined in the study.

Data transformations and corrections. To enhance validity of the calculated effect sizes, a number of transformations and corrections were made to the data. Because larger sample sizes have a reduced rate of sampling error, the inverse of the sampling error variance was calculated for all effect sizes. This calculation ensures that smaller samples are not weighted, effect-size-wise, the same as larger samples are (Lipsey & Wilson, 2001). The resultant Cohen's *d* effect sizes were tested for homogeneity (through the use of *Q* statistics), and confidence intervals (CIs) for mean effect sizes were calculated as the square root of the sum of the inverse variance weights (Hedges & Olkin, 1985). Given the importance of representative sampling of studies for meta-analyses, and given

that dissertations were excluded, robustness of aggregate effect sizes were also calculated through the use of fail-safe *N* (Orwin, 1983).

Results

Initial analysis of the 17 identified studies resulted in 18 independent and 52 nonindependent effect sizes of psychological assessment as a therapeutic intervention. As shown in Table 1, the analysis of included studies produced an overall effect size of $d = 0.423$ (CI [0.321, 0.525]), on the basis of a total sample of 1,496 participants. Sample characteristics may be seen in Table 2. This effect size is significant at the .01 level ($z = 8.135$). Thus, in this sample of studies, approximately 66% of participants who engaged in psychological assessment as an intervention fell above the control/comparison group mean.

Analysis of Homogeneity and Categorical Variables

The 18 independent effect sizes were assessed for homogeneity to determine whether variability across Cohen's *d* was greater than expected from sampling error alone. Initial analyses found the data to be homogenous ($Q[17] = 30.72$, $p \geq .01$). However, because a significant *Q* rejects homogeneity, this result is a liberal estimate. Accordingly, following the procedure recommended by Lipsey and Wilson (2001), independent effect sizes were analyzed in a manner analogous to that of ANOVA to test systematic variance in Cohen's *d* due to categorical variables, such as research design and type of dependent variable (e.g., process study, outcome study, process/outcome study).

The first model of systematic variance was tested with the categorical variable of research design. Because studies varied on their use of control/comparison groups, this categorical variable was tested to see whether it accounted for the heterogeneity across values for Cohen's *d*. This model demonstrated within-group homogeneity comparable to that in the previous model ($Q_W[16] = 30.650$, $p \geq .01$), and between-groups differences were nonsignificant ($Q_B[1] = 0.070$, $p \leq .05$). Thus, this particular model did not provide a more homogenous model of the data.

The second model of systematic variance was tested with the categorical variable of type of dependent variable. Because studies varied in how they operationalized therapeutic benefit (e.g., in terms of process, outcome, process/outcome), this categorical variable was also tested to see whether it accounted for the heterogeneity across effect sizes. This model demonstrated superior within-group homogeneity ($Q_W[15] = 19.623$, $p \leq .10$) and between-groups differences ($Q_B[2] = 11.097$, $p \geq .01$) than did earlier models, thus providing a better model of overall homogeneity. As shown in Figure 1, the average effect size for the process category was $d = 1.117$; for the outcome category, $d = 0.367$; and for the process/outcome category, $d = 0.547$. Moreover, CIs

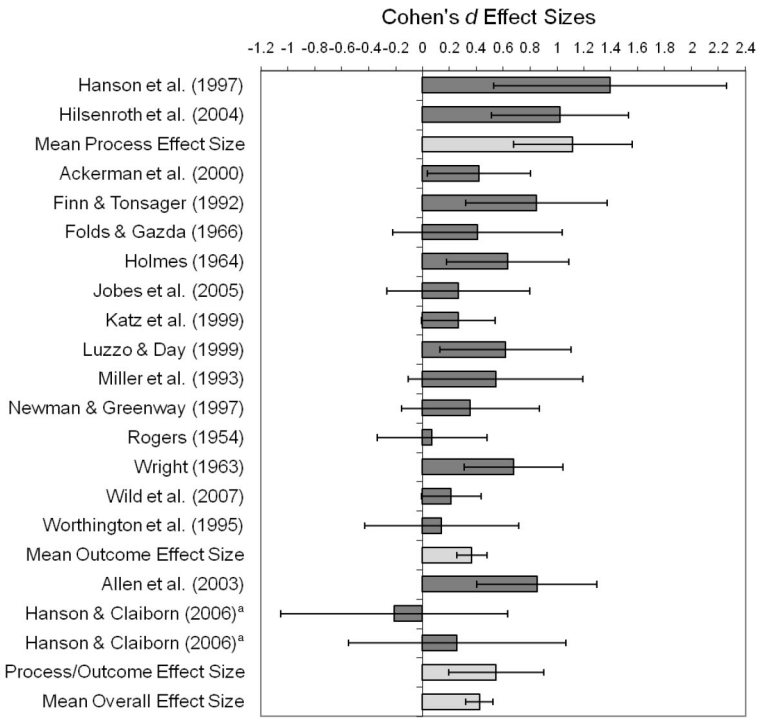

Figure 1. Effect sizes and 95% confidence intervals determined on the basis of a meta-analysis of psychological assessment as a therapeutic intervention. The effect size reported for each study is the average value of the calculated Cohen's *d* effect sizes within that study. The studies are grouped by whether Cohen's *d* was calculated on the basis of process, outcome, or both process and outcome dependent variables (a statistically significant categorical variable at the .01 level). Individual study effect sizes are presented in dark gray, whereas aggregate effect sizes are reported in light gray. [a] Hanson and Claiborn's (2006) research design included two independent variable levels and thus generated two independent effect sizes.

indicate that, upon repeated sampling, the CI of 95% of sampled effect sizes would include the population mean effect size. The CIs for the present effect sizes are 0.679 and 1.555 for the process category, 0.256 and 0.478 for the outcome category, and 0.193 and 0.901 for the combined category. Overlap of intervals may be seen in Figure 1. The collective distribution of effect sizes—for the outcome category—may be seen in Figure 2. Standard scores calculated as the absolute value of the mean effect size divided by its standard error (Lipsey & Wilson, 2001; Thompson, 2006) indicate mean effect sizes are significant at the .01 level for each category: process ($z = 5.009$), outcome ($z = 6.493$), and combined ($z = 3.025$).

Analysis of Overall Effect and Robustness

Regarding the research questions and overall benefit, success proportions were calculated on the basis of Cohen's *d* in the derived homogenous model. In the process category, approximately 86% of participants in the treatment groups fell above the control group means. The percentage over the control group means was 66% and 73% for the outcome and combined categories, respectively. Importantly, we calculated Orwin's (1983) fail-safe *N* to test the robustness of the calculated effect sizes. This analysis determines how many new/unidentified studies—with an assumed effect size of zero—would be needed to reduce the effect sizes reported herein to a criterion value specified by the researchers. We selected an effect size of 0.200 as the criterion value in this study to examine the robustness of the results against being reduced to this particular effect size. The number of new studies needed—with zero effects—to reduce the reported effect sizes to 0.200 is approximately nine (process category), 11 (outcome category), and five (combined category). Thus, the values for Cohen's *d* reported here appear to be robust.

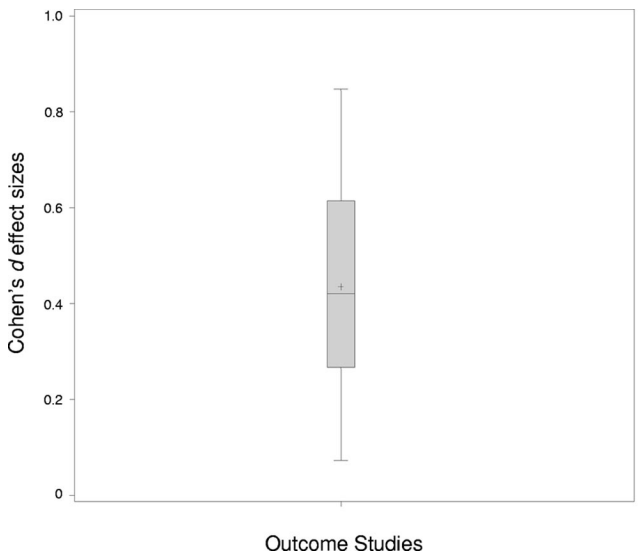

Figure 2. Distribution of average Cohen's *d* effect sizes across outcome studies of assessment as a therapeutic intervention. The top and bottom marks of the box plot indicate the maximum and minimum values of the distribution; the top and bottom of the gray area depict the 3rd and 1st quartiles, respectively; the line in the middle of the gray area depicts the median of the distribution; and the + sign depicts the mean of the distribution.

Discussion

In this study we used meta-analytic techniques to calculate and analyze the effect sizes of the results of studies related to psychological assessment as a therapeutic intervention. By synthesizing these results, we ascertained defensible data-based answers to the research questions and hypotheses.

Regarding the first, efficacy-based question, it was hypothesized that testing and feedback—when conducted in a personalized, collaborative manner—would affect treatment processes and outcomes in clinically meaningful and measurable ways. It was also hypothesized that, relative to control and comparison groups, treatment group means would be significantly higher. Both hypotheses were supported, because the overall Cohen's *d* effect size (0.423) was statistically significant and robust. This effect may, by conventional standards, be considered medium in size (cf. Cohen, 1977). However, the use of *small*, *medium*, and *large* as benchmarks (and terminology) is controversial (Glass, McGaw, & Smith, 1981; Thompson, 2001). Even Cohen himself explained that these benchmarks were "offered as conventions because they were needed in a research climate characterized by a neglect of attention to issues of [effect size] magnitude" (Cohen, 1988, p. 532). Thus, to truly understand and interpret this finding, one must compare it to the effect sizes obtained in the psychotherapy literature, generally (Thompson, 2001, 2002, 2006). Inasmuch, the overall effect size—and associated aggregate effect sizes (0.367–1.117)—are, it appears, comparable to those found in substance abuse/dependence treatment (0.450; Dutra et al., 2008). They are also approaching those found in cognitive–behavioral therapy treatment for anxiety disorders (0.890–2.590; Stewart & Chambless, 2009) and psychotherapy in general (0.800; Wampold, 2001).

Importantly, it appears the effect sizes reported here are robust. Fifty-eight percent of the treatment group would, for example, still fall above the control group mean ($d = 0.200$), even if 11 undiscovered or newly published outcome studies demonstrated zero effect. Given that therapeutic models of assessment involve as few as three sessions (see Finn, 1996, for a manualized model of therapeutic assessment), these effect sizes are noteworthy and help justify their place in treatment, especially in brief therapies (Clair & Prendergast, 1994; Levenson & Evans, 2000).

Regarding the second, variance-based question, it was hypothesized that certain study variables (e.g., design, type/focus) would account for significant variance across studies. This hypothesis was, like the first two, also supported, because the type-of-study model better explained effect size variances across studies. In this model, outcome variables produced an effect size of 0.367, and process variables 1.117. This latter finding is particularly compelling, given its magnitude and link, theoretically and otherwise, to positive therapeutic gains (cf. Tryon, 2001). Close inspection of Table 1 shows process variables fit with, among other theories, social influence (Strong, 1968), common factors perspectives (Wampold, 2001), and self-psychology (Kohut, 1977). Consequently, it is possible these theories help explain, and shed additional light on, the underlying change mechanisms associated with psychological assessment as an intervention.

It is interesting that the research design categorical variable—that is, whether studies included a no-treatment control (absolute efficacy) or a comparison (relative efficacy) group—did not constitute a significant categorical variable and, similarly, did not explain observed variances. This is interesting, in part, because it suggests that assessment and testing as usual (e.g., an information-gathering approach) and/or testing without feedback may be as therapeutically inert—in terms of treatment processes and outcomes—as receiving no treatment at all. This may explain why some psychologists, especially those unfamiliar with therapeutic models of assessment, do not believe in the treatment validity or clinical utility of assessment procedures. This is an empirical question, however, and future research should address it directly.

It is safe to say, on the basis of the meta-analytic findings reported here, that psychological assessment/testing and feedback is, if done in a specific way, a minitreatment/intervention in its own right, thus supporting contemporary therapeutic models of assessment (Finn & Tonsager, 1997). This take-home point is consistent with those of previous narrative reviews (Goodyear, 1990; Riddle et al., 2002), as well as those of empirical reviews of slightly different, though clearly related, literatures (Meyer et al., 2001). Perhaps Anastasi (1992) was right in noting that "whether any test is an instrument of good or harm depends on how the test is used" (p. 610). If tests are used collaboratively—and if they are accompanied by personalized, highly involving feedback—then clients and treatment appear to benefit greatly. Specifically, how and why it is beneficial remains largely unknown. Although researchers have speculated about this (Claiborn & Hanson, 1999; Finn & Tonsager, 1997, 2002; Ward, 2008), it remains a critically important direction for future research.

Limitations

Although the aggregate effect sizes reported here appear to be robust, relatively few studies are included in the analyses and dissertations were excluded.[1] The sample of studies may therefore be a limitation, in terms of both its representativeness and its comprehensiveness. It is possible that some relevant studies were inadvertently missed, even though the literature was searched systematically and conscientiously and involved multiple precautionary steps along the way. Until the findings are replicated in future studies and meta-analyses, they should be considered tentatively and in the context of the definitions of the constructs of interest provided in the article. For example, variability may exist in defining *process*, *outcome*, and *process/outcome* studies. If that were the case, then findings at the categorical level may change. Here again, future studies and meta-analyses are needed to tease out this potential limitation. Relatedly, of the 1,496 sample participants, the vast majority were European American women in their 20s, which limits the generalizability of the findings.

Because mean effect sizes may be unduly weighted in meta-analyses (Lipsey & Wilson, 2001), results must be interpreted cautiously. For example, in this sample of studies, one study produced a negative effect; a few produced confidence intervals (at the study level) below zero (see Figure 1); and, for two, effect size distributions were skewed slightly. Lastly, results of meta-analyses are only as sound as the individual studies themselves. Consequently, typical issues of internal and external validity apply equally to this meta-analysis. These limitations notwithstanding, the results have important preliminary implications for practice, training, and policy making, as well as future research.

Implications for Practice, Training, and Policy Making

Regarding practice implications, the results indicate that clinicians should familiarize themselves with therapeutic models of assessment. A number of excellent readings and resources address this topic (e.g., Finn, 2007). Clinicians should also seek out continuing-education training related to these models. Those who engage in *assessment and testing as usual* may miss out, it seems, on a golden opportunity to effect client change and enhance clinically important treatment processes. Similarly, applied training programs in clinical, counseling, and school psychology should incorporate therapeutic models of assessment into their curricula, foundational didactic classes, and practica (Curry & Hanson, in press).

Regarding policy making, the results indicate that competency benchmarks and guidelines for psychological assessment practice should be revisited to make sure they include key aspects of therapeutic models of assessment. Furthermore, managed care policy makers should take these results into account, especially as they make future policy and reimbursement decisions regarding assessment and testing practices (Eisman et al., 2000).

Future Research

Because this study demonstrates empirically the efficacy of psychological assessment as a therapeutic intervention, more research regarding the associated change mechanisms is needed. To accomplish this, dismantling, qualitative, and mixed methods studies are needed (Creswell, Hanson, Plano Clark, & Morales, 2007; Creswell & Plano Clark, 2007; Hanson, Creswell, Plano Clark, Petska, & Creswell, 2005; Kazdin, 2002). Dismantling studies could help identify (and clarify) the most important/salient aspects of therapeutic models of assessment, such as involving clients in the assessment process, having them develop personally relevant questions they would like answered by the testing, and teaching them how to optimize/maximize the process (via cognitive monitoring, self-regulated learning, and socialization of clients to the process). Qualitative and mixed methods studies and designs, such as grounded theory and sequential exploratory mixed methods designs, could help develop an overarching theory of change related to these models—a theory capable of guiding future research and hypothesis testing (cf. Creswell et al., 2007; Hanson et al., 2005).

As researchers gain insight into the change process, they will be better positioned to explain variances in observed effect sizes across studies. A number of variables warrant special empirical attention in this regard—variables including the "dose," or amount of test feedback given; the discrepancy of the results from clients' self-perceptions; the favorability, or valence, of the results (Finn, 1996; Hanson & Claiborn, 2006); and the types of tests used (e.g., objective, projective, neuropsychological). Special empirical attention should also be given to individual difference variables, in-

[1] One anonymous reviewer lamented the exclusion of dissertations in particular.

cluding age, ethnicity (e.g., African American and American Indian, in particular), and gender, as well as specific Axis I and II disorders as defined by the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; American Psychiatric Association, 2000) and other potentially relevant variables (e.g., clients' need for cognition and readiness for change). Attention to these kinds of variables may account for variability observed in CIs reported here. Additionally, given some clinicians' (and managed care agencies') negative attitudes toward psychological assessment and testing, future research should attend to this issue as well. Research of this sort could help inform—and facilitate—policy reform regarding assessment ethics, training competencies, and reimbursement infrastructures, as well as evidence-based practice. Finally, future research should calculate and report study-specific effect sizes and client and therapist demographic information.

Conclusion

Applied psychologists agree: Assessment and testing practices are at a critical point in their history. These practices will survive if, and only if, they demonstrate (empirically) their treatment validity and clinical utility. To this end, the study reported here contributes to their long-term survival. In addition to producing positive treatment outcomes, psychological assessment—as a therapeutic intervention—also, it appears, significantly enhances the treatment process. Thus, assessment and testing practices should, as a professionally distinguishing clinical activity, rise from the ashes of past critiques and criticisms, phoenixlike, and continue playing an important role in the profession generally and psychological treatment specifically.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*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.

*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,* 165–181.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychological Association. (2002). *Ethical standards of psychologists and code of conduct.* Washington, DC: Author.

Anastasi, A. (1992). What counselors should know about the use and interpretation of psychological tests. *Journal of Counseling & Development, 70,* 610–615.

Callahan, K. L., Price, J. L., & Hilsenroth, M. J. (2003). Psychological assessment of adult survivors of childhood sexual abuse within a naturalistic clinical sample. *Journal of Personality Assessment, 80,* 173–184.

Claiborn, C. D., Goodyear, R. K., & Horner, P. A. (2001). Feedback. *Psychotherapy: Theory, Research, Practice, Training, 38,* 401–405.

Claiborn, C. D., & Hanson, W. E. (1999). Test interpretation: A social influence perspective. In J. W. Lichtenberg & R. K. Goodyear (Eds.), *Scientist-practitioner perspectives on test interpretation* (pp. 151–166). Boston, MA: Allyn & Bacon.

Clair, D., & Prendergast, D. (1994). Brief psychotherapy and psychological assessments: Entering a relationship, establishing a focus, and providing feedback. *Professional Psychology: Research and Practice, 25,* 46–49.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York, NY: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Creswell, J. W., Hanson, W. E., Plano Clark, V. L., & Morales, A. (2007). Qualitative research designs: Selection and implementation. *The Counseling Psychologist, 35,* 236–264.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

Curry, K. T., & Hanson, W. E. (in press). National survey of psychologists' test feedback training, supervision, and practice: A mixed methods study. *Journal of Personality Assessment.*

Dutra, L., Stathopoulou, G., Basden, S. L., Leyro, T. M., Powers, M. B., & Otto, M. W. (2008). A meta-analytic review of psychosocial interventions for substance use disorders. *American Journal of Psychiatry, 165,* 179–187.

Eisman, E. J., Dies, R. R., Finn, S. E., Eyde, L. D., Kay, G. G., Kubiszyn, T. W., . . . Moreland, K. L. (2000). Problems and limitations in the use of psychological assessment in the contemporary health care delivery system. *Professional Psychology: Research and Practice, 31,* 131–140.

Engelman, D. H., & Frankel, S. A. (2002). The three person field: Collaborative consultation to psychotherapy. *Humanistic Psychologist, 30,* 49–62.

Finn, S. E. (1996). *Manual for using the MMPI-2 as a therapeutic intervention.* Minneapolis, MN: University of Minnesota Press.

Finn, S. E. (2003). Therapeutic assessment of a man with "ADD." *Journal of Personality Assessment, 80,* 115–129.

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.

Finn, S. E., & Tonsager, M. E. (1997). Information-gathering and therapeutic models of assessment: Complementary paradigms. *Psychological Assessment, 9,* 374–385.

Finn, S. E., & Tonsager, M. E. (2002). How therapeutic assessment became humanistic. *Humanistic Psychologist, 30,* 10–22.

Fischer, C. T. (1972). Paradigm changes which allow sharing of results. *Professional Psychology, 3,* 364–369.

Fischer, C. T. (1994). *Individualizing psychological assessment.* Mahwah, NJ: Erlbaum.

Fischer, C. T. (2000). Collaborative, individualized assessment. *Journal of Personality Assessment, 74,* 2–14.

Fischer, C. T. (2006). Phenomenology, Bruno Klopfer, and individualized/collaborative assessment. *Journal of Personality Assessment, 87,* 229–233.

*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.

Gelso, C., & Fretz, B. (2001). *Counseling psychology* (2nd ed.). Belmont, CA: Wadsworth.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Goldman, L. (1972). Tests and counseling: The marriage that failed. *Measurement & Evaluation in Guidance, 4,* 213–220.

Goodyear, R. K. (1990). Research on the effects of test interpretation: A review. *Counseling Psychologist, 18,* 240–257.

Gorske, T. T., & Smith, S. R. (2008). *Collaborative therapeutic neuropsychological assessment.* New York, NY: Springer.

*Hanson, W. E., & Claiborn, C. D. (2006). Effects of test interpretation style and favorability in the counseling process. *Journal of Counseling & Development, 84,* 349–357.

*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.

Hanson, W. E., Creswell, J. W., Plano Clark, V. L., Petska, K. S., & Creswell, J. D. (2005). Mixed methods research designs in counseling psychology. *Journal of Counseling Psychology, 52,* 224–235.

Haynes, 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.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hill, C. E., Nutt, E. A., & Jackson, S. (1994). Trends in psychotherapy process research: Samples, measures, researchers, and classic publications. *Journal of Counseling Psychology, 42,* 364–377.

*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 Assessment, 83,* 332–344.

*Holmes, J. E. (1964). The presentation of test information to college freshmen. *Journal of Counseling Psychology, 11,* 54–58.

*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 & Life-Threatening Behavior, 35,* 483–497.

*Katz, L., Joyner, J. W., & Seaman, N. (1999). Effects of joint interpretation of the Strong Interest Inventory and the Myers-Briggs Type Indicator in career choice. *Journal of Career Assessment, 7,* 281–297.

Kazdin, A. E. (2002). *Research design in clinical psychology* (4th ed.). Boston, MA: Allyn & Bacon.

Kohut, H. (1977). *The restoration of the self.* New York, NY: International Universities Press.

Levenson, H., & Evans, S. A. (2000). The current state of brief therapy training in American Psychological Association-accredited graduate and internship programs. *Professional Psychology: Research and Practice, 31,* 446–452.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

*Luzzo, D. A., & Day, M. A. (1999). Effects of Strong Interest Inventory feedback on career decision-making self-efficacy and social cognitive career beliefs. *Journal of Career Assessment, 7,* 1–17.

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. *American Psychologist, 56,* 128–165.

Michel, D. M. (2002). Psychological assessment as a therapeutic intervention in patients hospitalized with eating disorders. *Professional Psychology: Research and Practice, 33,* 470–477.

*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.

*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.

Oliver, L. W., & Spokane, A. R. (1988). Career-intervention outcome: What contributes to client gain? *Journal of Counseling Psychology, 35,* 447–462.

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational Statistics, 8,* 157–159.

Purves, C. (2002). Collaborative assessment with involuntary populations: Foster children and their mothers. *Humanistic Psychologist, 30,* 164–174.

Riddle, B. C., Byers, C. C., & Grimesey, J. L. (2002). Literature review of research and practice in collaborative assessment. *Humanistic Psychologist, 30,* 33–48.

*Rogers, L. B. (1954). A comparison of two kinds of test interpretation interview. *Journal of Counseling Psychology, 4,* 224–231.

Rosenthal, R. (2008). Science and ethics in conducting, analyzing, and reporting psychological research. In D. N. Bersoff (Ed.), *Ethical conflicts in psychology* (4th ed., pp. 390–394). Washington, DC: American Psychological Association.

Smith, S. R., Wiggins, C. M., & Gorske, T. T. (2007). A survey of psychological assessment feedback practices. *Assessment, 14,* 310–319.

Stewart, R. E., & Chambless, D. L. (2009). Cognitive-behavioral therapy for adult anxiety disorders in clinical practice: A meta-analysis of effectiveness studies. *Journal of Consulting and Clinical Psychology, 77,* 595–606.

Strong, S. R. (1968). Counseling: An interpersonal influence process. *Journal of Counseling Psychology, 15,* 215–224.

Tharinger, D. J., Finn, S. E., Wilkinson, A. D., & Schaber, P. M. (2007). Therapeutic assessment with a child as a family intervention: A clinical and research case study. *Psychology in the Schools, 44,* 293–309.

Thompson, B. (2001). Significance, effect sizes, stepwise methods, and other issues: Strong arguments move the field. *Journal of Experimental Education, 70,* 80–93.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31*(3), 24–31.

Thompson, B. (2006). Research synthesis: Effect sizes. In J. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 583–603). Washington, DC: American Education Research Association.

Tinsley, H. A., & Chu, S. (1999). Research on test and interest inventory interpretation outcomes. In M. L. Savickas & A. R. Spokane (Eds.), *Vocational interests: Meaning, measurement, and counseling use* (pp. 257–276). Palo Alto, CA: Davies–Black.

Tryon, G. S. (2001). *Counseling based on process research: Apply what we know.* Boston, MA: Allyn & Bacon.

Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings.* Mahwah, NJ: Erlbaum.

Ward, R. M. (2008). Assessee and assessor experiences of significant events in psychological assessment feedback. *Journal of Personality Assessment, 90,* 307–322.

Whiston, S. C., Brecheisen, B. K., & Stephens, J. (2003). Does treatment modality affect career counseling effectiveness? *Journal of Vocational Behavior, 62,* 390–410.

Whiston, S. C., Sexton, T. L., & Lasoff, D. L. (1998). Career-intervention outcome: A replication and extension of Oliver and Spokane (1998). *Journal of Counseling Psychology, 45,* 150–165.

*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.

*Worthington, E. L., Jr., McCullough, M. E., Shortz, J. L., Mindes, E. J., Sandage, S. J., & Chartrand, J. M. (1995). Can couples assessment and feedback improve relationships? Assessment as a brief relationship enrichment procedure. *Journal of Counseling Psychology, 42,* 466–475.

*Wright, E. W. (1963). A comparison of individual and multiple counseling for test interpretation interviews. *Journal of Counseling Psychology, 10,* 126–134.

Wygant, D. B., & Fleming, K. P. (2008). Clinical utility of MMPI-2 Restructured Clinical (RC) scales in therapeutic assessment: A case study. *Journal of Personality Assessment, 90,* 110–118.

Received September 8, 2009
Revision received December 4, 2009
Accepted December 4, 2009 ■