

# Discredited Psychological Treatments and Tests: A Delphi Poll

John C. Norcross  
University of Scranton

Gerald P. Koocher  
Simmons College

Ariele Garofalo  
University of Scranton

In the context of intense interest in evidence-based practice (EBP), the authors sought to establish consensus on discredited psychological treatments and assessments using Delphi methodology. A panel of 101 experts participated in a 2-stage survey, reporting familiarity with 59 treatments and 30 assessment techniques and rating these on a continuum from *not at all discredited* to *certainly discredited*. The authors report their composite findings as well as significant differences that occurred as a function of the experts' gender and theoretical orientation. The results should be interpreted carefully and humbly, but they do offer a cogent first step in consensually identifying a continuum of discredited procedures in modern mental health practice.

*Keywords:* Delphi Poll, discredited technique, evidence-based practice, psychological assessment, psychotherapy, quackery

Which psychotherapies are effective? Psychologists have been inundated with lists of treatment guidelines, empirically supported therapies, practice guidelines, and reimbursable therapies. But what about demonstrably *ineffective* treatments and tests?

The burgeoning evidence-based practice (EBP) movement in mental health attempts to identify, implement, and disseminate treatments that have been proven demonstrably effective according to the empirical evidence. This movement has provoked enormous controversy within organized psychology, and, with the exception of the general conviction that psychological practice should rely on empirical research, little consensus currently exists among the various stakeholders on either the decision rules to determine effectiveness or the treatments designated as “evidence-based” (Norcross, Beutler, & Levant, 2005).

We believe that it might prove to be as useful and probably easier to establish what does *not* work—discredited psychological treatments and tests. Far less research and clinical attention have been devoted to establishing a consensus on ineffective procedures as compared to effective procedures.

Recently, several authors have attempted to identify pseudoscientific, unvalidated, or “quack” psychotherapies (e.g., Carroll, 2003; Della Sala, 1999; Eisner, 2000; Lilienfeld, Lynn, & Lohr, 2003; Singer & Lalich, 1996). Parallel efforts are underway to identify assessment measures of questionable validity on psychometric grounds (e.g., Hunsley, Crabb, & Mash, 2004; Hunsley & Mash, 2005).

These pioneering efforts suffer from at least two prominent limitations. First, none of the efforts systematically relied on expert consensus to determine their contents. Instead, the authors assumed that a professional consensus already existed, or they selected entries on the basis of their own opinions. Second, these authors provided little differentiation between credible and uncredible treatments and between unvalidated and validated tests. This demarcation problem (Gardner, 2000)—the challenge of formulating sharp distinctions between validated and unvalidated—plagued earlier efforts, leading to rather crude and dichotomous judgments.

Thus, we conducted a poll of leading mental health professionals to help secure a consensus and to establish more refined characterizations of treatments and tests ranging from *not at all discredited* to *certainly discredited*.

## Delphi Poll

We searched broadly and collected nominations for discredited mental health treatments and tests via literature searches, electronic mailing list requests, and peer consultations. The inclusion criterion stated that the treatment or test was used professionally during the past 100 years in the United States or Western Europe. Note that, for inclusion, the treatment or test

---

JOHN C. NORCROSS earned his PhD in clinical psychology from the University of Rhode Island and is professor of psychology and distinguished university fellow at the University of Scranton. His areas of research include psychotherapy, clinical training, and self-initiated behavior change. GERALD P. KOOCHEER earned his PhD in clinical psychology from the University of Missouri at Columbia and is dean of the School for Health Studies at Simmons College. His research areas include adaptation to chronic illness, coping with bereavement, and ethical issues in psychology. ARIELE GAROFALO earned her BS in psychology at the University of Scranton and is now a doctoral candidate in clinical psychology at Nova Southeastern University. Her research areas include psychoanalytic treatments, projective testing, and personality disorders.

PRELIMINARY RESULTS OF THIS STUDY were presented at the 113th Annual Convention of the American Psychological Association, Washington, DC, August 2005.

WE GRATEFULLY ACKNOWLEDGE the participation of the 101 experts and the contributions of Scott O. Lilienfeld, Roger Greene, and John Hunsley to this study.

CORRESPONDENCE CONCERNING THIS ARTICLE should be addressed to John C. Norcross, Department of Psychology, University of Scranton, Scranton, PA 18510-4596. E-mail: norcross@scranton.edu

was used professionally for mental health purposes but was not necessarily used by psychologists themselves. Exclusion criteria were controversial theories of psychology that do not directly involve mental health (e.g., subliminal perception or sleep learning), broader unusual phenomena regarding human behavior (e.g., fire-walking, channeling, or extra sensory perception) that have not yielded pertinent treatments, treatments or assessments that have never found advocacy in mental health (e.g., astrology and numerology), medications or biochemical substances (including conventional, herbal, naturopathic, or homeopathic preparations), and practices used primarily outside the United States and Western Europe.

Using these criteria, we listed separately 59 treatments and 30 assessment techniques on a questionnaire. Items were presented alphabetically and with reference to a particular purpose. For example, "acupuncture for the treatment of mental/behavioral disorders" and "age-regression methods for treating adults who may have been sexually abused as children" were listed under the treatment section, and "Adult Children of Alcoholic (ACOA) checklists" and "anatomically detailed dolls or puppets to determine if a child was sexually abused" were listed under the test section.

The instructions to the respondents read as follows:

For the purpose of this Delphi poll of experts, we operationally define *discredited* as those unable to consistently generate treatment outcomes (treatments) or valid assessment data (tests) beyond that obtained by the passage of time alone, expectancy, base rates, or credible placebo. *Discredited* subsumes ineffective and detrimental interventions but forms a broader and more inclusive characterization. We are interested in identifying disproven practices.

Please rate the extent to which you view the treatment or test as discredited from *not at all discredited* to *certainly discredited*. A treatment or assessment tool can be discredited according to several types of evidence: peer-reviewed controlled research, clinical practice, and/or professional consensus. Please think in terms of the criteria for expert opinions as delineated in the *Daubert and Kumho Tire Co* (1999), legal standards. For example, in *Daubert* (1993) the Supreme Court cited factors, such as testing, peer review, error rates, and "acceptability" in the relevant scientific community, some or all of which might prove helpful in determining the validity of a particular scientific "theory or technique."

If you cannot make a rating because you are unfamiliar with the treatment or test, then kindly circle *NFT*. If you are unfamiliar with the treatment/test's research or clinical use, then kindly circle *NFR*. You may also circle both.

The response options were structured as a 5-point, Likert-type format in which 1 = *not at all discredited*, 2 = *unlikely discredited*, 3 = *possibly discredited*, 4 = *probably discredited*, and 5 = *certainly discredited*. Respondents could also circle *NFT* = *not familiar with this treatment/test* and *NFR* = *not familiar with its research or clinical use*.

The expert panelists were informed that:

Upon completion of the enclosed Delphi poll, your replies will be pooled with those of the other experts. Subsequently, you will be asked to complete a slightly modified form of the Delphi poll containing the preliminary findings. In this manner, the panel of experts will exchange opinions and arrive at a general consensus. Your individual responses will *not* be identified.

## Expert Panel

In October 2004, we mailed the five-page questionnaire to 290 doctoral-level mental health professionals: 100 randomly selected fellows of American Psychological Association (APA) Division 12 (Clinical Psychology), 45 randomly selected fellows of APA Division 17 (Counseling Psychology), 23 randomly selected fellows of APA Division 16 (School Psychology), 46 randomly selected fellows of the American Psychological Society (APS) with a major field in clinical psychology, 57 current and former editors of scholarly journals in mental health, 14 members of the APA Presidential Task Force on Evidence-Based Practice, and 5 chairs or editors of the *Diagnostic and Statistical Manual of Mental Disorders*. Following the initial mailing and a subsequent reminder, we received 138 responses for a total return rate of 48%. However, 37 of these were unusable for a multitude of reasons, principally retirement ( $n = 26$ ), which left 101 experts for a usable response rate of 35%.

We did not detect any systematic bias relating to gender or profession between persons who returned the questionnaire and those who did not. However, we did not have a reliable means of determining the theoretical orientations or professional activities of the nonrespondents, so it remains an open question if these factors influenced participation in the Delphi study.

Of the responding experts,<sup>1</sup> 93% had trained as psychologists, 82% identified themselves as male, and 9% reported belonging to a racial/ethnic minority group. Table 1 summarizes the professional characteristics of this clinically experienced and theoretically diverse panel of experts.

## Second Round

The experts' responses to the first questionnaire (Round 1) were pooled and analyzed. The same instrument was then redistributed to the 101 panelists in February 2005 along with feedback on the responses of the panel as a whole. Feedback was provided for each item in terms of means and standard deviations, depicted both numerically and graphically. Following two mailings and an e-mail prompt, 85 of the original 101 (84%) panelists responded to the second round. The demographic and professional characteristics of the experts responding to the second round mirrored those of the first round; statistical analyses did not detect any significant differences in these characteristics between Round 1 respondents and Round 2 respondents.

The second round questionnaire was identical to that of the first round in structure, directions, and items with three exceptions.

<sup>1</sup> Members of the expert panel who allowed us to report their names were: Marvin W. Acklin, George J. Allen, Elizabeth Altmaier, Juan Arellano-Lopez, David H. Barlow, Larry E. Beutler, David L. Blustein, Jeff Braden, Jeffrey M. Brandsma, Charles D. Claiborn, Elaine Clark, Karina W. Davidson, George DuPaul, Darwin Dorr, Colin D. Elliott, Albert Ellis, Al Finch, Michael First, Gerald B. Fuller, Charles J. Gelso, Roger L. Greene, Carol Goodheart, David Haaga, Steve Hollon, Bertram P. Karon, Bill Kinder, David Lachar, O. B. Leibman, Barbara McCrady, A. Scott McGowan, Julian Meltzoff, Greg Neimeyer, Mary Ann Norfleet, Ed Nottingham, Thomas H. Ollendick, Eugene G. Peniston, Mick Power, Michael C. Roberts, John Robinson, Wendy Silverman, Derald Wing Sue, Laura C. Toomey, Bruce Wampold, Drew Westen, and Danny Wedding.

Table 1  
Professional Characteristics of the 101 Experts

Characteristic	N	%	M	Mdn	SD
Highest degree					
PhD	94	93			
MD	3	3			
Other	4	4			
Year of highest degree			1971	1972	11.0
Years of mental health experience			33.2	33.0	10.6
Percent of time devoted to					
Clinical work			20.5	10.0	24.5
Teaching			20.1	20.0	16.2
Research/writing			31.7	30.0	24.6
Supervision			8.6	10.0	8.7
Administration			12.5	5.0	18.6
Primary employment setting					
University department	60	62			
Private practice	14	15			
Medical school	11	12			
Other	11	12			
Primary theoretical orientation					
Behavioral	24	25			
Cognitive	18	19			
Eclectic/integrative	21	22			
Humanistic/existential	8	9			
Interpersonal (IPT)	5	5			
Psychoanalytic/psychodynamic	7	7			
Systems/family systems	3	3			
Other	9	10			

First, as noted previously, the second round questionnaire presented the pooled responses from the first round, which is the standard procedure in Delphi polls. Second, we eliminated the four treatments (Goggle therapy, holotropic breath work, Pesso Boyden Psychomotor System, and the Sedona method) and five tests (Manson Evaluation, MAPS, Mira Myokinetic Psychodiagnostic Test, Pigem Test, and Zamboni Test) on the second round questionnaire that were rated by fewer than 25% of the experts during the first round. Third, precipitated by our difficulty in determining the cut off for the percentage of experts rating any particular item, we inserted a final question on the second round questionnaire concerning the percentage of responding experts that would constitute a respectable minority (see later discussion for details).

As expected, the variability among the expert ratings decreased from Round 1 to Round 2. In other words, the pooled feedback from the initial round led to greater consensus among the panel. Specifically, the standard deviations decreased on 50 of the 54 treatment ratings. The mean difference was  $-.20$ . Similarly, the standard deviations for 23 of the 25 tests decreased, with an average difference of  $-.15$  from the first to second round. The mean ratings evidenced less change: there was a net increase (i.e., in the direction of more discredited) of  $.11$  for the treatments and  $.04$  for the tests.

### Discredited Psychological Treatments

Table 2 presents, in ranked order, the experts' mean ratings of 55 potentially discredited treatments. Table 2 presents the mean, standard deviations, and percentage of experts not familiar with (and thus not rating) each item for both rounds of data collection. For those treatments rated by at least 25% of experts, considerable

convergence existed on treatments consensually viewed as *certainly discredited* (mean rating of 4.50 or higher on the 5-point scale). For the specific purpose listed, experts considered as certainly discredited 14 psychological treatments: angel therapy, use of pyramid structures, orgone therapy, crystal healing, past lives therapy, future lives therapy, treatments for post-traumatic stress disorder (PTSD) caused by alien abduction, rebirthing therapies, color therapy, primal scream, chiropractic manipulation, thought field therapy, standard prefrontal lobotomy, and aroma therapy. Another 11 treatments were consensually designated as *probably discredited* (mean rating of 4.0 or greater), including Erhard Seminar Training (EST) and age-regression methods, for various specific purposes.

On the other end of the continuum, a handful of treatments were consensually judged to be unlikely discredited. Treatments receiving mean scores below 3.0 were behavior therapy for sex offenders, thought stopping procedures for ruminations, psychosocial (i.e., nonbehavioral) therapies for attention deficit and hyperactivity disorder (ADHD), laughter or humor therapy for depression, Bion's method of group analysis, Moreno's psychodrama, and eye movement desensitization and reprogramming (EMDR) as a treatment for trauma.

### Discredited Psychological Tests

Table 3 displays the mean ratings of 25 potentially discredited tests in ranked order. As with the treatments, Table 3 presents the item mean, standard deviation, and the percentage of the experts lacking familiarity with that test for both rounds. No test received a mean rating above 4.50. Five tests rated by at least 25% of the experts in terms of being discredited for a specific purpose received mean scores of 4.0 or higher: Lüscher Color Test, Szondi Test, handwriting analysis (graphology), Bender Visual Motor Gestalt Test (for assessment of neuropsychological impairment), enagrams, and Lowenfeld Mosaic Test.

Several instruments were consensually viewed as unlikely discredited, at least for the specific purpose listed on the questionnaire. Those tests receiving mean ratings below 3.0 were office-based cognitive task assessments for ADHD, use of IQ scores and discrepancy formulas to identify specific learning disabilities, Thematic Apperception Test (TAT) for personality assessment, Myers-Briggs Type Indicator for assessment of personality, sentence completion tests for personality assessment, Adult Children of Alcoholic (ACOA) checklists, and the Rorschach Inkblot Test (Exner's comprehensive system) to diagnose specific disorders.

### Defining a Respectable Minority

As noted earlier, we struggled with the question of what percentage of experts would need to have familiarity with a given psychological treatment or test to qualify it for inclusion in our results. We decided to ask the expert panelists on the second round questionnaire.

We would ask for your help with one additional question. We are attempting to set a threshold of familiarity among experts by which to include a particular approach in our study. That is, what percentage of experts needs to be familiar with a given treatment or test for continued inclusion or discussion about discredited approaches? We feel stymied in reaching a reasoned criterion.

Table 2  
*Experts' Ratings of Mental Health Treatments in Ranked Order*

Treatment	Round 1			Round 2		
	<i>M</i>	<i>SD</i>	% not familiar	<i>M</i>	<i>SD</i>	% not familiar
Angel therapy for treatment of mental/behavioral disorders	4.92	0.28	71.4	4.98	0.14	46.4
Use of pyramids for restoration of energy	4.89	0.32	31.7	4.98	0.13	28.0
Orgone therapy (use of orgone box or orgone energy accumulator) for treatment of mental/behavioral disorders	4.92	0.28	26.0	4.97	0.17	16.9
Crystal healing for treatment of mental/behavioral disorders	4.82	0.44	34.0	4.95	0.21	21.0
Past lives therapy for treatment of mental/behavioral disorders	4.81	0.44	22.8	4.92	0.27	7.2
Future lives therapy for treatment of mental/behavioral disorders	4.67	0.80	42.4	4.88	0.33	30.5
Treatments for PTSD caused by alien abduction	4.69	0.68	26.7	4.85	0.40	20.5
Rebirthing therapies for treatment of mental/behavioral disorders	4.58	0.82	10.9	4.75	0.46	4.8
Color therapy for treatment of mental/behavioral disorders	4.56	0.56	59.6	4.68	0.62	50.6
Primal scream therapy for treatment of mental/behavioral disorders	4.51	0.77	5.9	4.61	0.72	4.8
Chiropractic manipulation for mental/behavioral disorders	4.36	0.70	29.3	4.57	0.60	14.6
Thought Field Therapy for treatment of mental/behavioral disorders	4.21	1.08	56.0	4.56	0.66	46.4
Standard prefrontal lobotomy for treatment of mental/behavioral disorders	4.44	0.90	9.9	4.55	0.80	8.3
Aromatherapy for treatment of mental/behavioral disorders	4.33	0.77	32.3	4.55	0.63	20.2
Erhard Seminar Training for treatment of mental/behavioral disorders	4.29	0.78	37.4	4.47	0.85	28.7
Age-regression methods for treating adults who may have been sexually abused as children	4.12	0.97	4.1	4.41	0.77	8.3
Craniosacral therapy for treatment of anxiety and depression	4.20	0.76	67.7	4.38	0.78	58.5
Preventive intervention for "born criminals"	4.19	1.13	66.3	4.36	0.85	49.4
Sexual reorientation/ reparative therapy for homosexuality	4.25	1.00	9.1	4.29	0.91	7.1
Holding therapy for reactive attachment disorder	4.14	1.06	37.8	4.29	0.71	33.7
Treatments for mental disorders resulting from Satanic ritual abuse	3.98	1.31	34.7	4.28	1.05	22.0
Healing touch (not massage therapy) for treatment of mental/behavioral disorders	4.10	1.12	41.4	4.13	0.96	25.6
Psychological treatments of schizophrenia based on the schizophrenogenic theory of schizophrenia	3.81	1.17	18.2	4.04	0.98	13.4
Parenting therapies for treatment of mental/behavioral disorders	3.84	1.13	24.8	4.03	0.94	13.4
Bettelheim model for treatment of childhood autism	3.94	1.19	19.4	4.01	0.90	14.3
Dolphin-assisted therapy for treatment of developmental disabilities	3.73	1.34	63.6	3.97	0.96	52.4
Equine therapy for treatment of eating disorders	3.64	1.15	70.7	3.96	0.73	69.5
Neuro-Linguistic Programming (NLP) for treatment of mental/behavioral disorders	3.57	1.06	26.7	3.87	0.92	24.1
Psychosynthesis for treatment of mental/behavioral disorders	3.69	1.29	68.0	3.83	1.04	65.1
Scared Straight programs for prevention of conduct disorders and criminal behavior	3.48	1.10	23.0	3.80	0.87	20.7
Emotional Freedom Technique for tx of mental/behavioral disorders	3.77	1.15	74.7	3.80	0.77	81.2
DARE programs for prevention of substance abuse/dependence	3.65	1.23	23.2	3.79	0.99	18.3
Family therapy for schizophrenia based on the double-bind theory of schizophrenia	3.73	1.30	10.2	3.76	1.07	6.2
Jungian sand tray therapy for treatment of adolescent and adult disorders	3.49	1.34	38.4	3.76	0.94	39.8
Rage reduction therapy for depression	3.71	1.14	42.6	3.72	1.01	36.9
Freudian dream analysis for treatment of mental/behavioral disorders	3.65	1.35	2.0	3.67	1.21	3.7
Facilitated communication for treatment of autism	3.49	1.48	26.3	3.67	1.20	21.0
Bioenergetic therapy for treatment of mental/behavioral disorders	3.37	1.11	47.5	3.63	0.86	33.3
Insight-oriented psychotherapies for sex offenders	3.47	1.32	7.1	3.54	0.99	3.6
Catharsis/ventilation treatment for anger disorders	3.32	1.23	9.1	3.53	0.94	8.4
Psychotherapy for the treatment of penis envy	3.60	1.51	22.8	3.52	1.34	18.3
Marathon encounter groups for treatment of mental/behavioral disorders	3.43	1.21	12.9	3.49	1.02	12.0
Acupuncture for treatment of mental/behavioral disorders	3.53	1.08	24.5	3.48	0.94	23.8
Psychotherapy for castration anxiety	3.42	1.52	25.7	3.26	1.34	20.7



Table 2 (continued)

Treatment	Round 1			Round 2		
	<i>M</i>	<i>SD</i>	% not familiar	<i>M</i>	<i>SD</i>	% not familiar
Critical incident stress debriefing (CISD) for acute trauma	3.02	1.43	15.5	3.25	1.06	11.0
Psychosocial therapies for treatment of pedophilia	3.30	1.26	16.8	3.23	1.04	14.5
Neurofeedback for ADHD	3.10	1.08	38.0	3.14	1.00	30.1
Classical psychoanalysis for removal of Axis I symptoms	3.22	1.31	1.0	3.10	1.07	3.7
Eye movement desensitization and reprocessing (EMDR) for the treatment of trauma	2.88	1.13	4.0	3.06	1.20	2.4
Moreno's psychodrama for treatment of mental/behavioral disorders	2.95	1.14	25.7	2.93	0.95	17.1
W. R. Bion's method of group analysis for treatment of mental/behavioral disorders	2.93	1.36	67.0	2.91	1.13	60.7
Laughter or humor therapy for treatment of depression	2.83	0.92	36.6	2.81	0.74	24.1
Psychosocial (nonbehavioral) therapies for ADHD	2.85	1.12	17.8	2.79	1.02	13.6
Thought stopping procedures for ruminations/intrusive worry	2.25	1.10	6.9	2.20	0.94	2.4
Behavior therapy for sex offenders	1.97	1.03	7.1	2.05	0.91	3.6

Note. Ratings made on a 5-point, Likert-type scale where 1 = *not at all discredited*, 2 = *unlikely discredited*, 3 = *possibly discredited*, 4 = *probably discredited*, 5 = *certainly discredited*. PTSD = post-traumatic stress disorder; tx = treatment; ADHD = attention deficit hyperactivity disorder.

We want to use the “respectable minority” standard typically applied in medical malpractice cases. This legal standard notes that not every practitioner will follow the same approach; we want to include approaches recognized by a respectable minority of practitioners. What percentage of experts would need to be familiar with an approach to consider it a “respectable minority” for this purpose?

\_\_\_\_\_ % of responding experts would constitute a respectable minority.

The responses ranged from 10% to 49% and averaged 30.4% (*Mdn* = 30.0, mode = 25, *SD* = 9.08). It is interesting to note that 10 experts were either confused by the question or disagreed with the legal standard of *respectably minority* in that they provided percentages above 50%. As a result, we ignored these 10 responses in the data analysis for this question.

#### Gender and Orientation Differences

In the context of discovery, we conducted a series of analyses to investigate differences in item ratings as a function of the experts' gender (*t* tests) and theoretical orientation (univariate analyses of variance [ANOVAs]) using a two-tailed alpha level of .025 to minimize Type I errors. We discovered nine statistically significant differences between men and women in our panel of experts among the 80 comparisons (55 treatments, 25 tests) in the second round data. Women accorded higher mean ratings (indicative of being discredited) to aromatherapy, family therapy based on the double-bind theory of schizophrenia, future lives therapy, primal scream therapy, and sexual reorientation/reparative therapy for homosexuality than did the men. Men accorded higher mean ratings to Jungian sand therapy, House-Tree-Person Technique, Kahn Test of Symbol Arrangement, and the Lüscher Color Test. These mean differences ranged from .16 to 1.17 and averaged .62. These were modest differences, but they were sufficient to alter the conclusion of whether a particular procedure warrants consideration as discredited for a specific purpose.

We conducted a parallel series of analyses on second round ratings among those experts self-identifying as (a) cognitive and

behavioral, (b) eclectic and integrative, or (c) psychodynamic and humanistic in their primary theoretical orientation. Table 4 summarizes the ANOVA results among these groups of panelists (Student Newman-Keuls post hoc tests, bidirectional alpha of .05). As seen in Table 4, we discovered 35 differences on the 80 items. Cognitive and behavioral therapists rated the procedures higher or more discredited in 34 of the 35 differences. Psychodynamic and humanistic therapists rated the procedures lower or less discredited in 29 of these 35 differences. The eclectic and integrative therapists, as a group, tended to occupy the middle ground of ratings.

#### Cautions and Caveats

Several factors demand that we exercise caution in interpreting these findings. First, many of our experts lacked familiarity with the many “fringe” therapies or “unusual” assessment techniques. Second, some might challenge the relatively modest size of our panel of experts. Third, our reliance on traditionally trained and academically vetted experts with a disproportionate number of cognitive-behavioral therapists might be too narrow. Fourth, the robust rating differences due to theoretical orientation (see Table 4) indicate that the epistemological commitments of the expert panel materially influence the results and thus the conclusions of what is discredited. Fifth, several panel members noted that a single item assessing the credibility of an omnibus assessment measure for a given purpose was insufficient. In particular, we wish that we had presented additional items on ability of the Rorschach to diagnose specific disorders (e.g., schizophrenia) and the Wechsler to identify specific learning disabilities.

We recommend interpreting these results carefully and humbly. Professional consensus does not equal an epistemic warrant; even experts can be and have been wrong. Expert opinions may be widely held because they are correct or because most experts simply share the same worldview. Test validity is conditional; clinical utility is purpose- and context-specific. A treatment or test considered discredited for one purpose might be credible for another. We should take care not to threaten innovation and

Table 3  
*Experts' Ratings of Mental Health Tests in Ranked Order*

Test	Round 1			Round 2		
	<i>M</i>	<i>SD</i>	% not familiar	<i>M</i>	<i>SD</i>	% not familiar
Lüscher Color Test for personality assessment	4.19	0.81	73.3	4.48	0.58	67.1
Szondi Test for personality assessment	4.15	1.05	44.6	4.46	0.85	36.6
Handwriting analysis (graphology) for personality assessment	4.36	0.70	10.0	4.27	1.00	8.5
Bender Visual Motor Gestalt Test for assessment of neuro-psychological impairment	4.92	0.28	9.1	4.23	1.33	14.5
Eneagrams for personality assessment	3.37	1.11	72.0	4.14	0.59	65.9
Lowenfeld Mosaic Test for personality assessment	3.48	1.27	39.0	4.05	0.74	74.4
Bender Visual Motor Gestalt Test for assessment of personality	3.82	1.20	14.0	3.93	1.04	13.3
Anatomically detailed dolls or puppets to determine if a child was sexually abused	4.12	0.97	14.0	3.90	0.95	7.1
Blacky test for assessment of children's personality and pathology	3.84	1.12	29.0	3.85	1.10	20.5
Hand Test for personality assessment	3.32	1.23	52.5	3.84	1.06	47.0
Kahn Test of Symbol Arrangement for personality assessment	3.52	1.16	67.0	3.77	0.94	63.9
Use of Wechsler IQ scale scores for personality assessment	3.52	1.37	14.9	3.42	1.18	6.1
Voice stress analysis for lie detection	3.07	1.18	45.5	3.31	0.97	34.9
Rosenzweig Picture Frustration Study for personality assessment	3.42	1.27	39.0	3.29	1.03	38.6
House-Tree-Person for personality assessment	3.22	1.31	9.0	3.23	1.18	6.0
Human Figures Drawings for personality assessment	3.13	1.37	5.9	3.19	1.18	6.1
Child Sexual Abuse Survivor checklists	3.00	1.14	60.0	3.07	1.02	44.6
Thematic Apperception Test—Thompson modification for personality assessment of ethnic/racial minorities	3.05	1.52	55.4	3.03	1.16	55.6
Rorschach (comprehensive system) to diagnose specific disorders	3.12	1.48	4.0	2.84	1.24	3.6
Adult Children of Alcoholic (ACOA) checklists	2.62	1.09	56.6	2.71	0.85	38.1
Sentence completion tests for personality assessment	2.64	1.20	7.0	2.67	0.99	1.2
Myers-Briggs Type Indicator for assessment of personality	2.60	1.25	18.8	2.63	1.09	9.6
Thematic Apperception Test for personality assessment	2.60	1.31	4.0	2.51	1.21	3.6
Use of IQ scores and discrepancy formulas to identify specific learning disabilities	2.41	1.23	15.0	2.26	0.88	3.6
Office-based cognitive task assessments for ADHD	2.35	1.09	55.0	2.23	0.89	47.0

*Note.* Ratings made on a 5-point, Likert-type scale where 1 = *not at all discredited*, 2 = *unlikely discredited*, 3 = *possibly discredited*, 4 = *probably discredited*, 5 = *certainly discredited*. PTSD = post-traumatic stress disorder; ADHD = attention deficit hyperactivity disorder.

creativity in psychological practice by branding all nonresearched procedures as discredited. We must avoid hubris by remembering that contemporary treatments and tests may become discredited 30 years from now. We must avoid false pride, as science should be ever vigilant and self-correcting.

### Research Implications

Before turning to the practical implications of our findings, we would amplify several methodological limitations of our study and point to corresponding improvements for future studies.

Our operational definition of discredited treatments and tests may have merged, for some experts, untested and demonstrably ineffective, unvalidated and invalidated. To paraphrase Justice Oliver Wendell Holmes, the absence of evidence is not the evidence of absence. Untested procedures are not necessarily ineffective procedures; they are merely untested. We left the criteria for making the discredited ratings to the experts on the basis of "several types of evidence: peer-reviewed controlled research, clinical practice, and/or professional consensus." Studies in the

future might provide separate ratings for psychological procedures that are considered untested versus those that are considered tested and found wanting or between procedures that are discredited by controlled research versus those discredited by clinical practice.

Several entries on our questionnaire were listed more than once for specific uses (e.g., Bender-Gestalt for assessment of neuropsychological impairment and again for assessment of personality), whereas other tests were listed only once as a generic use (for personality assessment). In the latter instance, the experts did not have the opportunity to render differential ratings. In the future, we advise investigators to offer multiple ratings on popular instruments for their different uses, such as the Rorschach, TAT, and sentence completion tests.

*Expert* in our study was defined by status as journal editor or association fellow, which predictably produced a disproportionate percentage of academics and cognitive-behavioral proponents. Were we to repeat the study, we would expand our definition of expert by more heavily sampling stellar practitioners of various theoretical orientations. More inclusive sampling might result in

Table 4  
*Summary of Statistically Significant Differences in Discredibility Ratings as a Function of Experts' Theoretical Orientation*

Item	Orientation differences	F	df	p
Catharsis/ventilation treatment for anger disorders	CBT > PD-HUM	4.5	2, 68	.015
Classical psychoanalysis for removal of Axis I symptoms	CBT > ECL, PD-HUM	14.0	2, 68	< .001
DARE programs for prevention of substance abuse/dependence	CBT > PD-HUM, ECL	7.8	2, 60	.001
Erhard Seminar Training for tx of behavioral/mental disorders	ECL, CBT > PD-HUM	5.6	2, 52	.006
Family therapy for schizophrenia based on double-bind theory	CBT, ECL > PD-HUM	14.7	2, 66	< .000
Freudian dream analysis for tx of mental/behavioral disorders	CBT > ECL > PD-HUM	17.1	2, 69	< .000
Healing touch for tx of mental/behavioral disorders	CBT > ECL > PD-HUM	8.9	2, 56	< .000
Holding therapy for reactive attachment disorder	CBT > ECL, PD-HUM	5.4	2, 50	.007
Insight-oriented psychotherapies for sex offenders	CBT > ECL, PD-HUM	8.8	2, 72	< .000
Jungian sand tray therapy for tx of adolescent and adult disorders	CBT > ECL	4.3	2, 44	.02
Marathon encounter groups for tx of mental/behavioral disorders	CBT, ECL > PD-HUM	10.1	2, 65	< .000
Moreno's psychodrama for tx of mental/behavioral disorders	CBT > PD-HUM	6.3	2, 60	.003
Neurofeedback for ADHD	CBT > PD-HUM, ECL	10.8	2, 52	< .000
Neuro-Linguistic Programming for tx of mental/behavioral disorders	CBT > PD-HUM	5.1	2, 57	.009
Primal scream therapy for tx of behavioral/mental disorders	CBT > PD-HUM	4.3	2, 70	.017
Psychotherapy for castration anxiety	CBT > ECL, PD-HUM	8.7	2, 59	< .000
Psychotherapy for the treatment of penis envy	CBT > ECL, PD-HUM	10.7	2, 61	< .000
Rage reduction therapy for depression	CBT > PD-HUM	6.6	2, 48	.003
Bion's group analysis for tx of mental/behavioral disorders	CBT > ECL, PD-HUM	7.9	2, 27	.002
Anatomically detailed dolls or puppets to determine if a child was sexually abused	CBT, PD-HUM > ECL	4.2	2, 69	.020
Bender Visual Motor Gestalt Test for assessment of personality	CBT > PD-HUM, ECL	6.1	2, 63	.004
Blacky test for assessment of children's personality and pathology	CBT > PD-HUM	4.3	2, 57	.018
Eneagrams for personality assessment	CBT, ECL > PD-HUM	4.8	2, 23	.018
Hand Test for personality assessment	CBT > PD-HUM	5.1	2, 39	.011
Handwriting analysis (graphology) for personality assessment	CBT, ECL > PD-HUM	12.4	2, 67	< .000
House-Tree-Person for personality assessment	CBT > ECL, PD-HUM	9.0	2, 68	< .000
Human Figures Drawings for personality assessment	CBT > ECL, PD-HUM	9.8	2, 67	< .000
Rorschach (comprehensive system) to diagnose specific disorders	CBT > ECL, PD-HUM	18.3	2, 72	< .000
Rosenzweig Picture Frustration Study for personality assessment	CBT > ECL, PD-HUM	11.4	2, 45	< .000
Sentence completion tests for personality assessment	CBT > ECL, PD-HUM	13.4	2, 71	< .000
Szondi test for personality assessment	CBT, ECL > PD-HUM	8.8	2, 44	.001
Thematic Apperception Test for personality assessment	CBT > ECL, PD-HUM	10.8	2, 70	< .000
Thematic Apperception Test—Thompson modification for personality assessment of ethnic/racial minorities	CBT > ECL, PD-HUM	5.0	2, 30	.013
Use of Wechsler IQ scale scores for personality assessment	CBT > ECL, PD-HUM	10.5	2, 67	< .000

*Note.* CBT = cognitive and behavior panelists; PD-HUM = psychodynamic and humanistic panelists; ECL = eclectic and integrative panelists; tx = treatment; ADHD = attention deficit hyperactivity disorder.

broader representation and avoid the temptation of non-CBT psychologists to simply put these results aside as being the self-fulfilling prophesy of academic clinicians who are largely of cognitive-behavioral persuasions.

### Practice Implications

The results of our Delphi poll demonstrate widespread unfamiliarity with unusual mental health procedures, theory-driven divergence in opinions, and yet substantial consensus on what qualifies as psychoquackery. In closing, we advance several practice implications emanating from these three principal results.

A surprising finding to us was the large percentage of experts unfamiliar with the listed practices. For example, in the first round, 56% were not sufficiently familiar with Thought Field Therapy to render a rating and 37% were not familiar with Erhard Seminar Training. These numbers may help us to understand one important and perhaps unappreciated reason for the relative apathy of many mental health experts toward discredited practices: many experts simply do not know much about them.

At the same time, and equally surprising, a large number of expert panelists informed us that completing the initial questionnaire prompted them to secure and read critical reviews of the treatments and tests on the questionnaire. Although many experts did not know much about these psychological interventions, they apparently had a keen interest in knowing more. We hope our study and this article facilitate such awareness among the practitioner community.

Experts of disparate theoretical orientations differed in many cases on the degree to which psychological treatments and tests were discredited. In 97% (34 of 35) of the significant differences, cognitive-behavioral panelists considered the procedure to be more discredited, perhaps reflecting their greater interest in experimental verification or their shared biases. Large orientation differences materially impact interpretation of the results. In the largest difference, on Freudian dream analysis for treatment of mental/behavioral disorders, the mean rating of cognitive-behavioral therapists was 4.4 (between probably and certainly discredited), the mean rating of eclectics was 3.4 (between possi-

bly and probably discredited), and the mean rating of psychodynamic–humanistic therapists was 2.8 (between unlikely discredited to possibly discredited). These differences underscore the robust effect of theoretical orientation in determining what constitutes discredited (and perhaps what stands as effective). Such findings remind us that panels, task forces, and commissions intent on identifying discredited (and effective) therapies should ideally include persons of various theoretical orientations, genders, ethnicities/races, sexual orientations, and so on if their goal is to create a professionwide consensus.

The recent APA Presidential Task Force on Evidence-Based Practice (2005) exemplified such diversity in its composition and provided a reasonably consensual definition: “Evidence-based practice in psychology (EBPP) is the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences.”

Our Delphi study systematically compiled clinical expertise on credibility, based perhaps in part on the best available research. Clinical consensus is only one valuable source of evidence that should be used in conjunction with empirical research and patient characteristics in assisting practitioners in determining what works and what does not work.

The consensus emerging on this Delphi poll on potentially discredited treatments and tests leaves us feeling encouraged. Multiple books, several Web sites (e.g., Quackwatch, Skeptic’s Dictionary), and a journal (*The Scientific Review of Mental Health Practice*) have dedicated themselves to publicizing psychological myths and discredited procedures. Psychological science tends to be self-correcting in that its foundation lies in empirical evidence (more than most professions, anyway). As a field, we have made progress in differentiating science from pseudoscience in the practice of psychology.

We believe that our study offers a cogent, positive first step in consensually identifying the “dark side,” “soft underbelly,” or “quack factor” of modern mental health practice and in providing a more granular analysis of the continuum of discredited procedures. Mature sciences and professions should have the ability to publicly shun discredited practices. The President’s New Freedom Commission on Mental Health (2003; www.mentalhealthcommission.gov) called attention to both the *underuse* of proven treatments and the *overuse* of treatments for which little empirical evidence exists. We can simultaneously use (inclusively defined)

EBPs to promote what *does* work and avoid (consensually identified) discredited practices to eradicate what does *not* work.

## References

- American Psychological Association Presidential Task Force on Evidence-Based Practice. (2005). *APA report on evidenced-based practice*. Retrieved October 22, 2005, from www.apa.org/practice/ebp.html
- Carroll, R. T. (2003). *The skeptic’s dictionary*. New York: Wiley.
- Daubert v. Merrell-Dow Pharmaceuticals, Inc., 509 U.S. 579, 113 S. Ct. 2786 (1993).
- Della Sala, S. (Ed.). (1999). *Mind myths: Exploring popular assumptions about the mind and brain*. New York: Wiley.
- Eisner, D. A. (2000). *The death of psychotherapy: From Freud to alien abduction*. Westport, CT: Praeger.
- Gardner, M. (2000). *Did Adam and Eve have navels? Debunking pseudoscience*. New York: Norton.
- Hines, T. (2003). *Pseudoscience and the paranormal* (2nd ed.). Amherst, NY: Prometheus.
- Hunsley, J. J., Crabb, R., & Mash, E. J. (2004). Evidence-based clinical assessment. *The Clinical Psychologist*, 57(3), 25–32.
- Hunsley, J. J., & Mash, E. J. (2005). Developing guidelines for evidence-based assessment of adult disorders. *Psychological Assessment*, 17(3), 251–323.
- Kumho Tire Co. v. Carmichael (97–1709), 526 U.S. 137 (1999), 131 F.3d 1433, reversed.
- Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (Eds.). (2003). *Science and pseudoscience in clinical psychology*. New York: Guilford.
- New Freedom Commission on Mental Health. (2003). *Achieving the promise: Transforming mental health care in America. Final report*. (DHHS Pub. No. SMA-03–3832). Rockville, MD: Department of Health and Human Services.
- Norcross, J. C., Beutler, L. E., & Levant, R. F. (Eds.). (2005). *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions*. Washington, DC: American Psychological Association.
- Park, R. L. (2000). *Voodoo science: The road from foolishness to fraud*. New York: Oxford University Press.
- Singer, M. T., & Lalich, J. (1996). *“Crazy” therapies*. San Francisco: Jossey-Bass.
- Wanijek, C. (2003). *Bad medicine: Misconceptions and misuses revealed from distance healing to vitamin o*. New York: Wiley.

Received May 24, 2005

Revision received January 27, 2006

Accepted February 7, 2006 ■