

***Manuscript prepared for publication in  
SPQ, 17(4), Winter 2002.***

**Evidence-Based Interventions in School  
Psychology: Conceptual Foundations of the  
Procedural and Coding Manual of Division 16 and  
the Society for the Study of School Psychology  
Task Force**

Thomas R. Kratochwill  
*University of Wisconsin–Madison*

Karen Callan Stoiber  
*University of Wisconsin–Milwaukee*

---

The contribution of the two authors to this article is equal. Requests for reprints should be sent to Thomas R. Kratochwill, PhD, School Psychology Program, 1025 West Johnson Street, University of Wisconsin–Madison, Madison, WI 53706; E-mail: tomkat@education.wisc.edu; or Karen Callan Stoiber, PhD, Department of Educational Psychology, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin; E-mail: kstoiber@uwm.edu

The *Procedural and Coding Manual* includes feedback from our respected colleagues, Steve Quintana and Bruce Wampold of the APA Division 17 (Counseling Psychology) Task Force, who met with us in Madison, Wisconsin, to review the Division 12 *Procedural and Coding Manual for Identification of Evidence-Based Treatments* and made various recommendations featured in this draft. In addition, we benefited from thoughtful input from our own Task Force members who did not serve on the EBI manual subcommittee, including Dick Abidin, George Batsche, George Bear, Virginia Berninger, Cindy Carlson, Craig Frisby, Dirk Hightower, Jan Hughes, Colette Ingraham, Tim Keith, Bonnie Nastasi, Evelyn Oka, Sylvia Rosenfield, and Ed Shapiro. We also wish to express our sincere thanks to John Weisz and Kristin Hawley for their assistance with our work. Special appreciation is extended to our Madison and Milwaukee graduate assistants working on the project: Lori Bruno, Angela Eke, Gretchen Lewis–Snyder, Laurie Martin, Michelle Miller, and Elisa Shernoff. Finally, we express sincere appreciation to Karen O’Connell and Lois Triemstra for their work on drafts of this document. We also express appreciation to Cathy Loeb for her most helpful editorial comments on the manuscript.

Copyright © 2002 by Guilford Publications, Inc. This manuscript is posted with permission of Guilford Publications, Inc. and is subject to copyright law and restricted from further use. No part of this manuscript may be reproduced, stored in a retrieval system or transmitted in any form or by any means (electronic, photocopying, recording or otherwise) without prior written permission of the publisher. To obtain permission please contact Guilford Publications, Inc. or e-mail: permissions@guilford.com. Reprinted with permission from *School Psychology Quarterly*.

We present the conceptual, philosophical, and methodological basis for the *Procedural and Coding Manual for Review of Evidence-Based Interventions* (hereafter called the *Procedural and Coding Manual*), portions of which are discussed in the subsequent two case study application articles in this issue and which is available on the world wide web at <http://www.sp-ebi.org>. First, we discuss some key conceptual issues and areas of potential controversy surrounding the content and organization of the *Procedural and Coding Manual*. Second, we discuss our research framework for coding evidence-based interventions (EBIs), taking into account the dimensional classification approach adopted by the Task Force on Evidence-Based Interventions in School Psychology. We contrast this coding scheme to the approach embraced by the Committee on Science and Practice of the Society of Clinical Psychology, Division 12, American Psychological Association (APA), in their *Procedural and Coding Manual for Identification of Evidence-Based Treatments* (Weisz & Hawley, 2002). Third, we present our methodological framework for reviewing EBIs, including quantitative group-based and single-participant designs, qualitative research designs, and theory-guided confirmatory program evaluation models. The use of diverse methodologies to provide evidence for the efficacy of EBIs raises both conceptual and methodological issues for this “work in progress.” Finally, we introduce the concept of a coding system to be implemented by practitioners to develop a knowledge base on what works in practice and help bridge the gap between research and practice. Implications for future work on classifying EBIs are presented within the context of the research-practice gap.

Over the past three years, we have worked with a group of talented leaders in the area of prevention and intervention to construct a knowledge base on evidence-based interventions (EBIs) that have application for the field of School Psychology (known as the Task Force on Evidence-Based Interventions in School Psychology). The Task Force on Evidence-Based Intervention in School Psychology (hereafter called the Task Force) is supported by the American Psychological Association (APA) Division 16 and the Society for the Study of School Psychology (SSSP). A specially formed subcommittee of the Task Force (hereafter referred to as the Manual Subcommittee) was commissioned with the task of developing a procedural and coding manual to review and document the evidence for effective prevention and intervention in, or useful for, the field of school psychology. Table 1 provides an overview of the Task Force and its organizational framework.

Since 1999 the EBI Task Force and the Manual Subcommittee have met to move forward the agenda of defining concepts, terms, and issues that should be considered in formulating a realistic and meaningful framework for identifying effective prevention and intervention programs (see Kratochwill & Stoiber, 2000c; Stoiber & Kratochwill, 2000). In previous articles, we discussed the need for school psychology to develop an

TABLE 1. Task Force Committees and Domains

---

Chair: Thomas R. Kratochwill<sup>1</sup>

<b>Conceptual and Methodological Issues</b> Five-person committee	<b>Qualitative Methods</b> Two-person committee
<b>Multicultural Issues</b> Two-person committee	<b>Confirmatory Program Evaluation</b> One-person committee
<b>School-based Prevention Programs</b> Three-person committee	<b>Comprehensive School Health Care</b> Two-person committee
<b>School-based Intervention Programs for Social Behavior Problems</b> Two-person committee	<b>Liaison Committee</b> (Division 53) (NASP) (Division 12)
<b>Academic Intervention Programs</b> Two-person committee	(Cultural Diversity Division 17) (Division 17) (Division 12)
<b>Family Intervention Programs</b> Two-person committee	
<b>School-wide and Classroom-based Programs</b> Two-person committee	

---

<sup>1</sup>Karen Callan Stoiber served as Co-Chair for the Task Force from its inception through April 1, 2002.

organized coding structure that permits reason-based judgments regarding the amount or quality of support found for a prevention or intervention program and/or approach. As our Task Force endorsed the idea that reviewing and documenting the evidence in support of interventions is an ongoing, ever-evolving activity, we adopted the term “evidence-based” to denote this work. We chose the term “evidence-based” because it places an emphasis on our goal of evaluating the level of evidence in support of an intervention (Stoiber & Kratochwill, 2001).

We believe that the criteria and coding structure established in the *Procedural and Coding Manual for Review of Evidence-Based Interventions* (hereafter called the *Procedural and Coding Manual* and available on the www at <http://www.sp-ebi.org>) hold important ramifications for the field of school psychology. Although the movement for constructing an evidence-based knowledge for implementing interventions in schools

and other applied settings is grounded in several conceptual and practical considerations discussed later in this article, the agenda of the Task Force centered on one especially critical intention. The efforts of the Task Force in identifying evidence-based prevention and intervention approaches are intended to narrow the research-to-practice and practice-to-research gap. Such a gap or "hiatus" between research and practice was described by Kazdin, Kratochwill, and VandenBos (1986) more than 15 years ago. The research-practice gap in psychology is thought to be due to infrequent and sporadic knowledge transfer that does not penetrate the day-to-day functions of researchers and practitioners in a meaningful way. The *Procedural and Coding Manual* criteria should set guidelines relevant for the field of school psychology that are useful both for researchers and practitioners in evaluating and conducting intervention research. Thus, the potential impact of the *Procedural and Coding Manual* on intervention practices in the field suggests the need for a systematic "dialogue" with our school psychology community. In this regard, we conceptualize the *Procedural and Coding Manual* as a "work in progress" and invite commentary from our colleagues, the field of psychology, and education.

The purpose of this article is to initiate such a dialogue by sharing conceptualizations, efforts, and products of the Manual Subcommittee and larger Task Force. Thus, we review the conceptual, philosophical, and methodological foundations of the *Procedural and Coding Manual*, which was developed by the Manual Subcommittee of the Task Force on Evidence-Based Interventions in School Psychology. Specifically, we discuss domains of prevention and intervention research within the context of our approach to classifying and establishing evidence regarding the efficacy of interventions. In the next two articles in this mini-series, we illustrate applications of the *Procedural and Coding Manual* to two investigations, one featuring group design (Conduct Problems Prevention Research Group, 1999a, 1999b; see Lewis-Snyder, Stoiber, & Kratochwill, this issue) and the other single-participant designs (Kelley & McCain, 1995; see Shernoff, Kratochwill, & Stoiber, this issue). The Editor, Terry Gutkin, has invited members of our Manual Subcommittee, as well as others outside our Task Force, to offer their perspectives on our efforts. An invitation to our subcommittee colleagues is reflective of the fact that we have not always secured complete agreement on the content and structure of the *Procedural and Coding Manual* among our own group. The reactions to the manual and this article provides another opportunity for input and dialogue on critical issues we face in the future.

In this article, we first introduce some key conceptual issues and potential areas of controversy surrounding the content and organization of the *Procedural and Coding Manual*. In addition, we present the rationale

for the coding scheme adopted by the Task Force and provide some contrasting perspectives to the approach taken by the Committee on Science and Practice in the Procedural and Coding Manual for Identification of Evidence-Based Treatments (Weisz & Hawley, 2002). We present our vision for the use of four methodological research domains as the basis for coding evidence-based interventions (EBIs). Finally, we explain how our coding framework can be used to gain information on interventions as they are implemented in various school psychology practice settings. The implications of our work for future EBIs in the school psychology field are elucidated within the context of school psychology research and practice agendas.

### KEY CONCEPTUAL ISSUES AND POTENTIAL AREAS OF CONTROVERSY: GETTING FROM THERE TO HERE

When our Task Force began developing the *Procedural and Coding Manual*, we assumed we should adopt the coding criteria presented by the Division 12 Task Force.<sup>1</sup> This issue was given serious consideration within the context of mounting criticism of Division 12 criteria, as illustrated by the following summary comments of Waehler, Kalodner, Wampold, and Lichtenberg (2000):

...they are described as too lenient, providing distraction from more important research findings, relying on flawed randomized clinical trial designs, stifling therapy research, assuming diagnostic specificity, dehumanizing clients based on DSM diagnosis, and relying on adherence to treatment manuals which can hinder effectiveness by turning therapists into technicians. (p. 661)

As can be seen from this critique, Division 12's attempt to identify a list of interventions that were "empirically supported or evidence-based" was targeted as problematic due to poor generalization of interventions from research clinic to practice settings, failure to address issues of comorbidity that occur in clinical settings, the rigidity of "manualized" interventions, and an over reliance on randomized clinical trials as representing sound research (see also Wampold, 2001). Despite some penetrating criticisms of the Division 12 criteria (see Garfield, 1998; Goldfried & Wolfe, 1998), the division's work is quite comprehensive, and their criteria are based on several years of effort by a talented group of researchers and scholars. Adopting the Division 12 manual would have had the

---

1. We prepared input into earlier versions of the Division 12 manual.

advantage of allowing us to code our domains with the same format and present results to the field with consistent criteria. Nonetheless, after considerable discussion and spirited debate among our Task Force Manual Subcommittee members, we made a decision not to adopt the Division 12 manual for several reasons described below.

As we developed the *Procedural and Coding Manual* and interacted with colleagues on this task, a variety of conceptual issues emerged that are both controversial and difficult to resolve. Some of the issues we faced are not new and have been confronted by other task forces as they have identified and disseminated EBIs (see Chambless & Ollendick, 2001, for a detailed overview of these issues in the clinical psychology field). As we struggled with many of the issues confronted by our clinical psychology colleagues and tried to address some of their controversies, we confronted our own set of challenging issues, including the following: (a) constructing a scheme for evaluating the evidence that supports an intervention, (b) identifying the focus of interventions, (c) determining the content of prevention/intervention domains, (d) coding/recoding existing literature reviews, and (e) reviewing interventions designed outside the framework of our EBI coding systems. Each of these issues is discussed below, along with our recommendations for how to address them. (Some issues, such as use of intervention manuals, have been discussed previously—see Kratochwill & Stoiber, 2000c; Stoiber & Kratochwill, 2000—and will be revisited here briefly.)

#### Scheme for Coding Evidence

Prior to discussing the issues related to using the *Procedural and Coding Manual*, we present a brief overview of the structure contained in the manual. It is hoped that this overview will provide readers with a template for understanding the organizing features of the *Manual*. The *Procedural and Coding Manual* has four sections that correspond with various design qualities: (a) group-based design, (b) single-participant design, (c) qualitative research methodology, and (d) confirmatory program evaluation. Each of the four sections of the manual contains a similar schema for coding the evidence base of prevention/intervention studies. Three strands of evaluation features of a study are examined to determine whether an intervention produced intended effects: (a) general characteristics of the study, including empirical/theoretical basis, design, and statistical qualities; (b) key evidence components; and (c) other features to consider.

The first set of coding considerations used in evaluating a study relates to the empirical/theoretical basis, general design qualities, and statistical treatment of the prevention/intervention application under investi-

gation. These coding criteria address whether a strong theoretical or empirical basis for conducting the study has been established, the overall quality of design methodology, and the use of an appropriate outcome evaluation (i.e., statistical procedures). These first set of criteria provide a context for understanding what was done and why. For example, an examination of the theoretical and/or empirical basis for a study may indicate theoretical evidence such as the role of cognitive-behavioral theory in having application in intervention procedures used for treating childhood depression. At a more specific level, empirical support may exist indicating the efficacy of an intervention as it is applied to a specific problem, with children having specific demographic characteristics, and in a specific context. As noted earlier, a premium is placed on work that has been done in schools or has clear application to schools or related settings. Thus, details on specific empirical studies that support the intervention under study and that establish under what conditions outcomes were found are viewed as stronger evidence than research studies having rationales only based on theoretical underpinnings. Determining the theoretical and/or empirical basis of an intervention study is an important feature to consider across various research designs, including group-based, single-participant, qualitative, and confirmatory program evaluation. Whenever possible, the intent of the Manual Subcommittee was to be inclusive and to reflect the diversity in the psychology and education fields by examining the features of intervention studies that correspond to important evidence considerations, regardless of the type of research methodology used.

The second type of criteria, called *key evidence components*, focuses both on internal validity criteria as well as features considered important for school- or field-based implementation. To determine whether interventions are effective, the Manual Subcommittee recommended that several key evidence components ideally should be incorporated into a study, including: (a) outcome measurement procedures that are valid, reliable, multi-method, and multi-source; (b) a comparison group demonstrating the same target problem to test outcome differences when a group design is used (i.e., group equivalence established); (c) key outcomes that are statistically significant; (d) equivalent mortality for participants; (e) evidence of durability of effects, (f) identifiable components indicating what aspects of the intervention produced which outcomes; (g) evidence of intervention fidelity/integrity, and (h) information regarding replication. Each of the eight key evidence components is evaluated using an approach that can be viewed as analogous to a "benchmark" structure. A rating of a "3" is given for strong evidence/support, denoting the highest research standard on a scale of 0 to 3; "2" for promising evidence; "1" for weak evidence/support; and "0" for no evidence/sup-

port. This “benchmarking structure” should permit an “evidence-based” framework for evaluating the effects and efficacy of school psychology or field-based interventions. It is hoped that by evaluating this set of eight key evidence components, researchers and practitioners will have a standard template for examining the quality of intervention research. Although all eight key evidence components may not be feasible or necessary to determine the “evidence-base” of a study, they are nonetheless useful for this purpose.

The third type of information addressed within the EBI criteria refers to *other dimensions or considerations* that consumers may want to consider when evaluating the appropriateness of an intervention or for determining whether it is well matched to the specific needs, qualities, or expectations of a school or other applied setting. Within the third set of criteria is an examination of external validity indicators, including descriptive information on participant characteristics and the context within which the intervention occurred. These research qualities are included within the third set because it is important to evaluate specific demographic details about what population served as participants, how the intervention was implemented, and what were the patterns of change associated with the intervention. These important details that relate to what population, what conditions and context, and what outcomes often cannot be readily incorporated into standard statistical analysis or quantitative ratings of level of evidence in support of a study. Thus, the third type of criteria describes other important features of a study that do not fit easily into a rating scale structure but are nonetheless useful for judging the beneficial aspects or dimensions of a particular intervention or program. Types of supplemental dimensions or other considerations that are evaluated include (a) demographic information on the intervention participants, (b) descriptions of the intervention program implementers (e.g., research staff, classroom teacher, parent), (c) particular implementation specifications and requirements (e.g., staff training required, program materials needed), (d) receptivity or acceptance by target participant population, (e) consumer satisfaction information, and (f) length of intervention.

In evaluating criteria in the “other dimensions or considerations” section, coders first determine a rating of evidence for external validity indicators (0 = no evidence to 3 = strong evidence) for (a) replication, and (b) school- or field-based site for implementation. When information is provided for categories permitting descriptions of the research, this information is noted specifically. For example, coding of the external validity category “participant selection and description” requests descriptions of the following child characteristics if available: grade/age, gender, ethnicity, SES, family structure, locale, disability, and functional

descriptors. Judgments also are available about the specificity of information available. Most frequently, participant descriptors are provided at a general or through cross-tabulated information [e.g., there were 317 grade 9 and 10 students, approximately equally split between males (52%) and females (48%), with 63% of the sample being African-American, 21% Hispanic, 7% White, and 8% of other minority status]. Yet, more specific demographic information is required to answer questions, such as: Of the participants, how many 9th graders were African American males? What percentage of 10th graders were girls and Hispanic? Evaluations about these other considerations are especially useful for replication work and for determining particular features that fit practice needs or specifications.

In summary, these three types of criteria should provide a schema outlining the kinds of intervention qualities that are examined and coded. Together the criteria represent a comprehensive, but not exhaustive, set of standards for designing and evaluating “evidence-based interventions.” We now turn to specific considerations that occurred in constructing the framework, focus, and intended functions of the *Procedural and Coding Manual*.

#### Focus of Interventions

For Division 12, the conceptual framework for linking interventions to childhood problems has been the *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (DSM-IV) (Weisz & Hawley, 2001). Adoption of the DSM framework clearly has a number of advantages, including, for example, (a) a common language system compatible with a large body of research literature in clinical and other applied areas of psychology that is organized around disorders (e.g., intervention texts such Carr, 2000; Christophersen & Mortweet, 2001; Dobson & Craig, 1998; Kendall, 2000; Morris & Kratochwill, 1998; Nathan & Gorman, 1998; Stoiber & Kratochwill, 1998), (b) wide-scale use in other mental health associations such as the American Psychiatric Association, and (c) compatibility with psychopharmacology intervention research linked to DSM-diagnosed childhood mental disorders and frequently compared to psychosocial interventions in clinical trial outcome research. Despite the compelling advantages of this approach, our Task Force wanted to expand our foci beyond *disorders* to embrace a broader range of issues that confront psychologists and other professionals working in educational settings. Our decision was based on the following considerations.

First, we considered how consumers would use our Task Force products and information, and, specifically, how school professionals would

use our intervention summary findings. Although our Task Force is supported by psychological organizations (i.e., APA Division 16 and the SSSP), our consumer audience is a diverse group of education professionals who serve students with special needs in schools (special and regular education teachers, speech and language clinicians, social workers, education counselors, and administrators, to name a few). These individuals do not always embrace the DSM, and their knowledge base on interventions is not organized around this framework. Second, we did not want to be limited by the DSM in addressing the wide range of student problems and issues that are the focus of interventions in schools. As our domain content suggests, we focus on literacy problems, math disabilities, health care behavior, management of classrooms, and instructional activities of teachers, among many other concerns in education. The DSM cannot capture this diversity of intervention foci in schools. Third, the DSM's pathological or medical model is limiting as a framework for organizing interventions in the school psychology field<sup>2</sup>. We were interested in reviewing prevention programs in each of the domains, and a pathological focus would limit this effort (see also Waehler et al., 2000). Our goal is to review a wide range of prevention programs that can be implemented in schools to *prevent* the disorders often focused on in the child area (see discussion below).

In summary, our Task Force is interested in reviewing disorders linked to the DSM (and will code studies with this focus with our criteria); however, we did not want the DSM to serve as the only guiding structure. We are fortunate that Division 12 has made so much progress in identifying and analyzing interventions for major childhood disorders and problems. However, we also hope to extend our review beyond this disorder-based model to consider (a) a broader range of consumers who do not embrace or use a disorder-based model, in particular, researchers and practitioners involved in school-based interventions; (b) a broader range of intervention foci; and (c) a nonpathological focus that can encompass prevention programs across our domains.

#### Content and Organization of the Prevention /Intervention Domains

Part of the work of our Task Force was to further refine and clarify our intervention domains for purposes of organizing literature reviews and making summary statements about the efficacy of intervention research. Table 2 provides the format for the organization of the *Procedural and*

---

2. It is beyond the scope of this article to discuss the limitations of the DSM model. For a penetrating perspective on the issue, see Albee (2000).

TABLE 2. Organization of the Task Force *Procedural and Coding Manual* by Content and Methodological Domains

Methodological Focus	Intervention/Prevention Content Focus				
	School- and Community- Based Intervention Program for Social and Behavioral Problems	Academic Intervention Programs	Family and Parent Intervention Programs	School-Wide and Classroom-Based Intervention Programs	Comprehensive and Coordinated School Health Services Intervention Programs
Between-Group Research					
Single-Participant Research					
Qualitative Research Procedures					
Confirmatory Program Evaluation					

*Coding Manual* by content and methodological domains (further discussion of the methodological domains is presented later in this article). Currently, we have five content domains: (a) School- and Community-Based Intervention Programs for Social and Behavioral Problems; (b) Academic Intervention Programs; (c) Family and Parent Intervention Programs; (d) School-Wide and Classroom-Based Intervention Programs; and (e) Comprehensive and Coordinated School Health Service Interventions. Reviews of these domains have been presented previously and will not be featured here (see Kratochwill & Stoiber, 2000b, and Stoiber & Kratochwill, 2001, for an overview).

A first issue we confronted was how to address prevention programs. Our Task Force is highly committed to reviewing prevention programs as well as interventions undertaken once a problem or disorder already exists. We are especially committed to prevention because schools are attractive settings for conducting prevention programs. As school psychologists become involved in prevention initiatives, they must prioritize their efforts based on specific situational factors, available resources, and budget constraints that may affect program implementation (see Zins, Elias, Greenberg, & Pruetz, 2000). Prevention programs focus primarily on the promotion of social, emotional, and behavioral competencies, are designed to reduce the incidence of mental health problems, and are implemented before a problem becomes salient or is diagnosed as a disorder. Although effective prevention programs have been around for some time, school-based prevention programs for substance abuse, violence, or other risk behaviors have become more prevalent in recent years (e.g., Minke & Bear, 2000; Durlak, 1997; Greenberg, Domitrovich, & Bumbarger, 2000; Reppucci, Woolard, & Fried, 1999).

One key Task Force concern has been how to define the population of participants in prevention research. As one moves away from a pathological model that embraces a categorical definition of participants (e.g., DSM-IV; American Psychiatric Association, 1994), the task seems enormous. We have decided to identify programs that target prevention efforts explicitly in the research article, while adhering to the Institute of Medicine's (1994) classification of prevention programs as *universal*, *selective*, and *indicated* (see Greenberg et al., 2000).

A related consideration was whether to review all prevention programs within a separate prevention domain or whether to review such programs within each content domain. We decided to go with the latter tactic because the boundaries of prevention and intervention are often blurred, especially at the *indicated* level. Although we resolved the issue of how to deal with prevention programs by including them within each of the content domains, it is very clear that as individuals begin the task of reviewing specific literatures, there will be redundancy. For example,

some programs may be coded as intervention programs and some as prevention programs.

A second issue the Task Force confronted was how to handle overlap among domains. We have decided to maintain the integrity of our five content domains and are requesting that our domain co-chairs specify clear criteria for how the intervention is being defined within the context of the domain focus. It is certainly possible that some of the interventions assigned to one domain will be featured in literature reviews in another domain. For example, some family intervention programs (e.g., Webster-Stratton, 1996) could be featured both in the School- and Community-Based Intervention Programs for Social and Behavioral Problems and in the Family and Parent Intervention Programs domain. Such redundancy should not pose a problem inasmuch as the purpose of the reviews is to give an overall picture of an intervention's efficacy within the context of a particular well-defined area of research. In short, an intervention study or line of intervention research may have relevance and will need to be reviewed across multiple domains specified by the Task Force.

#### Coding/Recoding Existing Literature Reviews

A primary challenge the Task Force has confronted is recoding research included in published literature reviews that use classification criteria at variance with our own. One example is the classification of "empirically supported treatments" for children and adolescents by the Division 12 Task Force (see Lonigan, Elbert, & Bennett-Johnson, 1998, for summary reviews of the literature on child and adolescent treatment and, more recently, Chambless & Ollendick, 2001 and, the special series in the *Journal of Pediatric Psychology* on empirically supported treatments in pediatric psychology; see Drotor & Lemanek, 2001). The literature reviewed by these authors is directly relevant to school psychology research and practice. As an example, the majority of EBIs for oppositional defiant and conduct disorder have been implemented in school settings, and this literature has been subjected to a comprehensive literature review (Brestan & Eyberg, 1998).<sup>3</sup> Although we believed initially that we should

---

3. The subtitle of the Brestan and Eyberg (1998) review – "29 years, 82 studies, and 5,272 kids" – gives some sense of the scope of work that has already gone into a review of the literature on conduct-disordered children and adolescents.

4. In the design of our coding system, we will be able to feature interventions for which there exists little or no empirical support. Intervention research with negative results will also be considered in our review efforts (see Kratochwill, Stoiber, & Gutkin, 2001).

TABLE 3. Workgroup Criteria for Identification of Empirically Supported Therapies

**Group A**

Division 12 Task Force

(Chambless et al. 1998) criteria

Well-established treatments

I. At least two good between-group design experiments must demonstrate efficacy in one or more of the following ways:

A. Superiority to pill or psychotherapy placebo, or to other treatment

B. Equivalence to already established treatment with adequate sample sizes

OR

II. A large series of single-case design experiments must demonstrate efficacy with

A. Use of good experimental design, and

B. Comparison of intervention to another treatment

AND

III. Experiments must be conducted with treatment manuals or equivalent clear description of treatment

AND

IV. Characteristics of samples must be specified

AND

V. Effects must be demonstrated by at least two different investigators or teams

Probably efficacious treatments

I. Two experiments must show that the treatment is superior to waiting-list control group

OR

II. One or more experiments must meet well-established criteria IA or IB, III, and IV above but V is not met

OR

III. A small series of single-case design experiments must meet well-established-treatment criteria

Experimental treatments

Treatment not yet tested in trials meeting task force criteria for methodology

**Group B**

Special section of Journal of Pediatric Psychology

(Spirito 1999) criteria

Well-established treatments

Same as Chambless et al. (1998)

Probably efficacious treatments

Same as Chambless et al. (1998)

Promising interventions

I.

OR

II. There must be positive support from a small number of single-case design experiments

OR

III. There must be positive support from two or more well controlled studies by the same investigator

**Group C**

Special section of Journal of Clinical Child Psychology

(1998, vol. 27, no. 2) criteria

Well-established treatments

Same as Chambless et al. (1998)

Probably efficacious treatments

Same as Chambless et al. (1998) except: there must be at least two, rather than one, group design studies meeting criteria for well-established treatments if conducted by the same investigator

#### **Group D**

Special section of Journal of Consulting and Clinical Psychology

(Kendall & Chambless, 1998) criteria

Efficacious and specific

Same as Chambless et al. (1998) for well-established treatments

Possibly efficacious and specific treatments

Same as efficacious and specific criteria above except: treatment was found superior to wait-list group in one study and superior to rival treatment in another study by a different team

Efficacious treatments

Same as Chambless et al. (1998) for well-established treatments except: treatment must be demonstrated to be better than no treatment but not been shown to be better than nonspecific intervention, placebo, or rival intervention

Possibly Efficacious Treatments

Same as Chambless et al. (1998) for probably efficacious treatments

#### **Group E**

What Works for Whom?

(Roth & Fonagy, 1996) criteria

Clearly effective treatments

I. There must be a replicated demonstration of superiority to a control condition or another treatment condition

OR

II. There must be a single high-quality randomized control trial in which:

A. Therapists followed a clearly described therapeutic method useable as the basis for training

B. There is a clearly described patient group

Promising limited-support treatments

Treatment must be innovative and a promising line of intervention

OR

Treatment is a widely practiced method with only limited support for effectiveness

#### **Group F**

A Guide to Treatments That Work

(Nathan & Gorman, 1998) criteria

Type 1 studies

I. Study must include a randomized prospective clinical trial

II. Study must include comparison groups with random assignment, blind assessments, clear inclusion and exclusion criteria, state-of-the-art diagnostic methods, and adequate sample size for power

III. There must be clearly described statistical methods

Type 2 studies

Clinical trials must be performed, but some traits of type-1 study were missing (e.g., trial with no double blind or group assignment not randomized)

Type 3 studies

I. There are open treatment studies that are aimed at obtaining pilot data

OR

II. These are case control studies in which treatment information was obtained retrospectively

**Group G**

Treatments for older adults

(Gatz et al., 1998) criteria

Same as Chambless et al. (1998) criteria

**Group H**

Treatments for chronic pain

(Wilson & Gil, 1996) criteria

Same as Chambless et al. (1998) criteria

Adapted and reproduced with permission.

---

adopt the Division 12 evidence about efficacious programs, our coding criteria differ from theirs in some significant ways; consequently, some Task Force members thought we should reevaluate interventions already reviewed by Division 12 with the hope of identifying additional directions for research and practice. Some Task Force members also argued that recoding some of the research reviewed by Division 12 would provide a unique opportunity to establish the reliability of the review process across different coding systems (see further discussion of this issue later in this article).

Various work groups within clinical psychology established criteria for EBIs (see summary by Chambless and Ollendick, 2001, and Table 3, which is reproduced from their review). When Chambless and Ollendick applied these criteria to the empirically validated treatments for children and adolescents (see Table 3 in the Chambless and Ollendick review), they found that when two or more groups reviewed the same treatment (e.g., behavior modification in the classroom for attention deficit hyperactivity disorder, rapid exposure for school phobia), results, on the whole, were generally consistent in indicating the efficacy of the intervention. However, Chambless and Ollendick note that conclusions about the consistency or inconsistency of coding must take into account the fact that (a) the review groups did not determine which interventions were not efficacious,<sup>4</sup> (b) the review groups did not review all the psychotherapy literature, and (c) all groups did not review all interventions. The current Division 12 Task Force criteria (Weisz & Hawley, 2001) are similar to our own in a number of ways and will become the benchmark for our identification of EBIs in the near future. Therefore, we envision some consensus between our conclusions pertaining to EBIs and those of the Division 12 Task Force.

In other cases, major reviews have used coding criteria that are differ-

ent from those of our Task Force and those of Division 12. For example, Greenberg et al. (2000) provided a comprehensive review of prevention programs. However, they did not make the research distinctions our Task Force makes regarding methodological, statistical, and measurement criteria. Thus, it is highly likely that the conclusions drawn from their review of various prevention programs will vary considerably from our own. For example, Greenberg et al. (2000) feature the Families and Schools Together (FAST) program developed by McDonald and her associates (e.g., McDonald, Billingham, Conrad, Morgan, Payton & Payton, 1997; McDonald & Sayger, 1998) as a “promising” intervention. However, close examination of the empirical literature evaluating FAST shows that some studies do not fare well when reviewed using the design criteria in the *Procedural and Coding Manual*.

To take another example, the National Reading Panel (2000), in response to a congressional charge to review reading interventions and implications for reading instruction, conducted an “evidence-based assessment of the scientific literature on reading and its implications for reading instruction” using 10 coding criteria. Table 4 lists the reading panel criteria and our own Task Force criteria for both group and single-participant designs. It is clear that the reading panel took into account multiple design criteria in considering evidence. Nevertheless, the panel’s criteria are at some variance from those our own Task Force.

The review efforts of the Division 12 Task Force have been and will likely continue to be extraordinary. Likewise, the National Reading Panel’s (2000) assessment was a Herculean effort. A very practical question is whether our Task Force should adopt these groups’ findings or begin the arduous task of reviewing and recoding the vast literature these groups have already assessed. We see two middle-ground approaches. First, in the spirit of cooperation and collaboration, we should establish formal ties with other groups involved in the EBI effort. For example, our domain co-chairs could establish cooperative working relationships with individuals on the Division 12 Task Force to collaborate in reviewing a body of research evidence. As Task Force co-chairs, we strongly encourage this collaboration. Second, in cases in which a review has already been conducted, our Task Force could take a *sampling* of investigations in the intervention area and subject them to our coding criteria. If we established a high level of agreement (e.g., a high kappa coefficient) on the sample coding trial, we would not review the entire literature again. Although the details of these two approaches still need to be worked out (e.g., establishing priority areas, sampling considerations), they may offer efficient and practical solutions to a challenging problem.

TABLE 4. Comparison Coding Criteria From the National Reading Panel and the Task Force Within the Domain of Design Methodology

---

NATIONAL READING PANEL

- \* Random assignment of participants to treatments (randomized experiment)
  - With vs. without a pretest
- \* Nonequivalent control group design (quasi-experiment) (Example: existing groups assigned to treatment or control conditions, no random assignment)
  - With vs. without matching or statistical control to address nonequivalence issue
- \* One-group repeated measure design (i.e., one group receives multiple treatments, considered a quasi-experiment)
  - Treatment components administered in a fixed order vs. order counterbalanced across subgroups of participants
- \* Multiple baseline (quasi-experiment)
  - Single-subject design
  - Aggregated-subjects design

TASK FORCE

(Group and Single-Participant Design)

General Design Characteristics

Type of Between Group Design

Randomized Designs

- Completely randomized design
- Randomized block design (between-subjects variation)
- Randomized block design (within-subjects variation)

Nonrandomized Designs

- Nonrandomized design
- Nonrandomized block design (between-subjects variation)
- Nonrandomized block design (within-subjects variation)

General Design Characteristics  
(Classify studies according to the type of design)

Type of Single-Participant Design

- Within-series design
  - Simple phase change
  - Complex phase change
  
  - Between-series design
  - Comparing two interventions
  - Comparing interventions with no interventions
  
  - Combined-series design
  - Multiple baseline across participants
  - Multiple baseline across behaviors
  - Multiple baseline across settings
  - Multiple probe design
  
  - Combined single-participant and group design or,
  
  - Combined single-participant design
  
  - Within series design
  - Simple phase change
  - Complex phase change
  
  - Between-series design
  - Comparing two interventions
  - Comparing interventions with no interventions
  
  - Combined-series design
  - Multiple baseline across participants
  - Multiple baseline across behaviors
  - Multiple baseline across settings
  - Multiple probe design
- 

Interventions Designed Through Models Outside Task Force Frameworks

The goal of the Task Force is to identify interventions within a domain, review these interventions using established criteria, and render statements (ratings) about the state of evidence to support the interventions.

Ultimately, the purpose of this task is to influence practitioners' selection and implementation of interventions in school and other applied settings. Nevertheless, what appears at first glance to be a straightforward task is complicated by a variety of conceptual and theoretical considerations. Several different approaches to organizing and developing interventions have been presented in the professional literature (see Elliott, Witt, Kratochwill & Stoiber, in press; Kratochwill & McGivern, 1996). For example, interventions can be conceptualized according to alternative models of assessment such as a diagnostic approach, the response class/keystone approach, the empirical or case formulation model, and the functional assessment approach, among others (see also Braden & Kratochwill, 1997). In this context, it may be helpful to consider our Task Force approach (as well as those of other task forces) and contrast it with alternatives that may be prominently represented in research and practice.

The *Procedural and Coding Manual* represents what we regard as an empirical guide to classifying interventions and recommending their implementation in practice. Using this approach, some population of intervention research is identified, interventions are reviewed, and summary statements are made about the match between the problem or disorder and the conclusions of the empirical literature about efficacious interventions. An example will further illustrate this process and suggest how an EBI might be selected in school psychology practice. In practice, assessment could involve understanding the topography of a problem or disorder as operationalized through empirically based rating scales or diagnostic and classification systems (Achenbach & McConaughy, 1996). Based on such assessment, a school psychologist might select an EBI parent-training program consisting of multiple components to deal with antisocial and conduct problem behavior of youth (e.g., Brestan & Eyberg, 1998; Webster-Stratton, 1996). Similarly, a cognitive behavior therapy program might be selected from among the range of EBI alternatives for implementation with children experiencing anxiety disorders (e.g., Kendall, 2000; Ollendick & King, 1998). Thus, a practitioner reviewing the literature in these areas will find specific information related to *matching* the specific problem with certain *package* intervention procedures. Yet, critics may suggest that interventions identified in an EBI Task Force framework without an individualized assessment linked to an empirically established model of intervention at the individual case level are doomed to failure in practice. Functional assessment and analysis provide one illustration of this alternative view of intervention development (see Gresham, Watson, & Skinner, 2001; Watson, Gresham, & Skinner, 2001).

Although the terms *functional assessment* and *functional analysis* are sometimes used interchangeably, they should be distinguished on both conceptual and methodological grounds (Cone, 1997). *Functional assessment* refers to a process that involves the specification of hypotheses about variables that may control behavior and typically occurs during the early phases of a multistep assessment process (O'Neill, Horner, Albin, Storey, & Sprague, 1997; Schill, Kratochwill, & Gardner, 1996). *Functional analysis* refers to an "experimental process" in which various hypotheses are examined at the individual level, often through single-participant designs (McComas & Mace, 2000). The distinction between the two levels or types of assessment is important in the context of the use of EBIs. Interventions developed through a functional assessment process are similar to empirical interventions affiliated with the EBI process, especially when functional assessment-generated hypotheses are matched to a priori functionally derived categories of interventions (e.g., escape, skill deficits). An example will help illustrate our point. In development of the *Functional Assessment and Intervention System* (Stoiber & Kratochwill, in press), we embraced an assessment model in which interventions are *matched* to the specific hypothesized functions of behavior. The interventions we offer are categorized in domains similar to those found in reviews of the empirical literature on functional assessment (see Ervin, Radford, Bertsch, Piper, Ehrhardt, & Poling, 2001). Interventions developed around the hypothesized functions of behavior may or may not be similar to those established through other methods of empirical documentation. These functional assessment-derived interventions may have "treatment validity" (Kratochwill & Plunge, 1992) and can be subjected to standards advanced for evaluating the validity of the "functional behavior assessment" process (Shriver, Anderson, & Proctor, 2001). Yet, some EBI critics who embrace a functional conceptual or theoretical model to interventions may be less enthusiastic about matching some problem to an EBI or hypothesized function intervention that has not been specifically subjected to an individual functional *analysis*.

Consider further, however, that identifying and implementing EBIs may not always run counter to functional assessment and analysis. Indeed, practitioners can review the EBI evidence in a particular area based on the nature of a problem and still conduct a functional analysis to complement EBI approaches. Interventions derived from functional assessment and EBI empirically based approaches also can be subjected to further experimental test, including analysis of variables in single-participant designs. An example of this option is illustrated in the multicomponent intervention procedures recommended for attention deficit hyperactivity disorder (ADHD) (see DuPaul & Stoner, 1994). A

package intervention for ADHD may consist of contingency management in the classroom, parent training, self-instructional procedures, and cognitive behavior therapy, among other psychosocial and biological interventions (see Rapport, Chung, Shore, & Isaacs, 2001). However, functional assessment and analysis can also be implemented with ADHD to understand unique aspects of the problem and design specific intervention tactics linked to the functions of behavior (DuPaul & Ervin, 1996). Thus, our Task Force plan is to offer the field additional tools and strategies that may facilitate implementation of an intervention. In no way are the activities of the Task Force designed to replace those strategies based on functional assessment and analysis or other unique assessment technologies that might be invoked in research and practice.

### **EXPANDED METHODOLOGICAL CRITERIA FOR CODING STUDIES**

In the tradition of quantitative research, Division 12 adopted methodological criteria that embrace true experimental investigations (e.g., random assignment, group design comparison of alternative conditions). Single-participant research design is also featured as a quantitative methodology for determining efficacy in outcome research, although the conditions of true experiments cannot always be met using this methodology. Our Task Force has clearly embraced an experimental research paradigm for reviewing intervention efficacy evidence, but it has also considered (a) methodological criteria that take into account limitations of traditional experimental research, and (b) alternative methodologies (e.g., qualitative research methodology and confirmatory program evaluation).

#### **Methodological Criteria**

Historically, true experimental design has been the hallmark for establishing evidence of the efficacy of an intervention, whether group (Kazdin, 1998) or single-participant (Kratochwill & Levin, 1992). With the extensive application of experimental design in applied settings over many years (and especially in schools), awareness of the limitations of this methodology has grown. These limitations include (a) problems using random assignment, (b) uncontrolled sources of variability following randomization to conditions, and (c) a paucity of information generated from outcomes of the study, whether positive or negative (Lipsey & Cordray, 2000). Recognition of these issues has prompted some investigators to abandon the experimental paradigm in favor of other methodologies (e.g., qualitative methodology; but see below).

Such a move is premature and actually limits the development of knowledge in various intervention domains (Kratochwill & Stoiber, 2000d; Levin & O'Donnell, 1999; Levin, O'Donnell, & Kratochwill, in press).

Our Task Force has addressed these concerns in a number of ways. To begin with, we have decided to rely on the internal and external validity criteria associated with experimental research but to use a dimensional framework in coding studies. Our "Consumer Report" approach is linked to a likert scale format wherein various dimensions of methodological and statistical criteria are applied to a study. Table 5 illustrates our dimensional coding criteria as applied to the "comparison group" issue in the group design research coding manual. It can be observed that the comparison group dimension is rated on a four-point scale (i.e., 3 to 0). (Further illustrations of the coding criteria applied to intervention research are featured in the two subsequent articles in this mini-series.)

Our Task Force approach to coding has been used previously in some literature reviews (see Troia, 1999, for an example in the reading area) and has several advantages over other possible approaches. First, studies that do not meet traditional randomized true experimental criteria can be coded as evidence for an intervention while noting the limitations of the investigations in an area (i.e., through designation in the rating system). Second, our coding framework provides consumers with information to make decisions on selection of interventions. Individuals reviewing the evidence on a particular intervention can observe areas of difficulty or lack of support (i.e., credible design and/or measurement) and draw their own conclusions based on the evidence provided. Placing responsibility on consumers (researchers, practitioners, policymakers) for selection of EBIs is in accord with best ethical practice for psychological services (Beutler, 2000). Third, we consider support for an intervention to be dynamic and in constant need of updating, as new evidence is forthcoming. Thus, some intervention programs will "move up or down the scale" as new evidence attesting to their efficacy is produced.

#### Coding by What Methodological Criteria

Because the *Procedural and Coding Manual* is organized into four major methodological domains (see Table 2), questions have emerged about which approach to use in coding specific studies. Thus, in contrast to the Division 12 approach, in which a single study is coded with a single manual, we have four options for this task. As can be seen in Table 1, we focus on group design research, single-participant research, qualitative investigations, and confirmatory program evaluation approaches (further discussion of the qualitative and confirmatory program evaluation

TABLE 5. Illustration of Dimensional Coding Criteria Featured in the Procedural and Coding Manual for Evidence-Based Interventions in School Psychology

B. Comparison Group. There must be a comparison group of some type against which to compare the intervention in judging the merits of an intervention. Single group pre-post designs and case studies are not considered. The intervention in question could have been shown to be superior to a no intervention or wait list control, an attention or placebo control, an alternate intervention, or pharmacotherapy. Or the intervention in question was compared to an alternate intervention and to a control condition in one study. It may be shown superior to the control condition but not the alternate intervention. Evidence ratings will consider all available information related to the comparison group findings.

To be rated a 3 for strong evidence, at least one type of "active" comparison group (i.e., active control, alternate intervention, or pharmacotherapy) must be used. It must be demonstrated that change occurred in the desired direction in addition to demonstrating that the intervention under study demonstrated better outcomes than the comparison group. The only way for this to be done is through the use of a pre/post test or multiple measures design in addition to a control group design (e.g., a between by within design). Group equivalency must be established, either a priori using a randomization procedure, or conduct post hoc tests for group differences. There must be evidence that change agents were counterbalanced, as well as the study must meet the criteria for equivalent mortality and low attrition at post-test, and if applicable, at follow-up.

To receive a rating of 2 for promising evidence, at least a "no intervention group" type of comparison must have occurred, and there must be equivalent mortality. If equivalent mortality is not demonstrated, an intent-to-intervene analysis must have been conducted, resulting in a finding of no significant group differences between the control and treatment groups.

A rating of 1 would require a comparison group (e.g., a waiting list), however; no group equivalence procedures were used.

A rating of 0 occurs for a pre-post design assessment only.

approaches is presented below). The different coding approaches are not mutually exclusive; on the contrary, they may be used in combination to code research. For example, it is possible that the quantitative criteria for group and single-participant research could be supplemented with the coding criteria for qualitative investigations. Domain co-chairs will need to make decisions about which particular coding structure (or combination) is applied to a particular intervention research area. Although some guidelines have been provided for this process, coding research across multiple methodological domains may prove challenging. The details of this task need to be addressed.

A second challenge pertains to studies that use more than one methodological approach. For example, some studies include both group and single-participant research methodologies in combination. In such cases, our domain co-chairs will be advised to code the investigation with one primary methodology and code those dimensions of the investigation relevant to the secondary methodology only if necessary.

### Alternative Intervention Research Methodologies

*Scope of the Issues.* Early in our deliberations as a Task Force, we considered whether to expand our methodological focus beyond quantitative research (group and single-participant design). We decided to embrace two alternative methodologies: qualitative research procedures and most recently, confirmatory program evaluation.

To facilitate the application of qualitative procedures to the review process, the Task Force created a subcommittee on qualitative methods. Co-chaired by Bonnie Nastasi and Stephen Schensul, the subcommittee was asked to develop procedures for coding studies on qualitative indicators during the intervention review process. Draft coding criteria are now available, and the *Journal of School Psychology* will feature a special mini-series on this topic in the future.

*Qualitative Research Methodology.* Especially within the past decade, interest in qualitative research methodology has burgeoned. The increasing interest is due in part to the growing list of limitations of traditional experimental research (as noted above) and a host of theoretical and conceptual issues surrounding methods of acquiring knowledge and understanding research context (Levin et al., in press)<sup>5</sup>. Generally, qualitative research refers to investigation of a broad range of phenomena that involves narrative accounts, description, interpretation, context, and meaning in the research process (Kazdin, 1998). These dimensions can be applied to intervention programs to expand knowledge about the efficacy of these interventions and their contextual application in educational and applied settings (Lipsey & Cordray, 2000; Nastasi & Berg, 1999).

The use of qualitative research methodology in applied research is controversial, and we anticipate that many researchers who embrace quantitative experimental approaches will be critical of our venture into this domain. However, there are some compelling reasons for including this approach in our coding EBI efforts. First, qualitative intervention research exists and may actually be on the increase as graduate programs offer more training in this methodology. We need to consider this research option and what it contributes to the knowledge base of intervention efficacy. Thus, we have embraced the task of developing coding criteria for this evidence, although the effort is in the early stages of

---

5. It is beyond the scope of this article to review qualitative methodology and issues relating its application to intervention research. For general sources on this topic, see Lincoln & Guba (1985); Miles & Huberman (1994); and Schensul et al. (1999); see also Chapter 10 in Kazdin (1998).

development<sup>6</sup>. Second, we are interested in supplementing quantitative methods with qualitative procedures to learn about contextual issues surrounding implementation of an intervention. Ethnographic methods can be especially useful in this regard (see Nastasi & Berg, 1999). Third, for those circumstances in which intervention research does not meet traditional internal validity standards (e.g., random assignment, appropriate control groups), we want a methodology to assist in gathering evidence to help support or refute the efficacy of the intervention. Qualitative methods represent one tool in this process and allow an examination of multiple data to understand the parameters of an intervention procedure. Fourth, we want to code investigations in our review that rely primarily, or even exclusively, on qualitative procedures to offer research directions for various intervention domains. Our rationale here supports one focus of the Task Force, which is to present future directions for research, including experimental methods that may be applied to document the efficacy of an intervention (see Kratochwill & Stoiber, 2000d). We realize that adopting the position that qualitative research provides a “stepping-stone” to more quantitative and experimental research may relegate qualitative investigation to a lesser position as a research method. Such a position runs counter to the view that qualitative methodology should be put on an equal footing with quantitative experimental methodology. We do not share this view (see Kratochwill & Stoiber [2000d] for a discussion of our developmental research model illustrated in Figure 1 in that article).

*Confirmatory Program Evaluation.* Confirmatory program evaluation, a methodological approach that may be less familiar to intervention researchers, focuses on detection of intervention program effects while taking into account theory-driven program design variables (Reynolds, 1998). Causal inferences for an intervention are addressed through consideration of size, specificity, consistency, and coherence of intervention-outcome relationships and various theory-based causal mechanisms related to intervention outcomes. Confirmatory program evaluation embraces a theory-driven model in which the researcher must explicate and assess a priori theories to determine efficacy (Bickman, 1987).

Reynolds (1998) noted that theory-based evaluations can be more “confirmatory” than other evaluation approaches because multidimensional issues can be addressed, including, for example, the following: (a) program effect size, (b) small and large program effects, (c) consistency

---

6. The coding criteria for qualitative procedures are being refined at the time of writing.

of effects (e.g., across subgroups, models), (d) causal mechanisms, and (e) factors that influence selection into the program and quality of implementation (pp. 204–205). Essentially, confirmatory program evaluation allows causal inference to be strong if the results of the analyses are consistent with the program theory and hypotheses advanced by researchers.

Confirmatory program evaluation is not designed to replace other methodological or statistical approaches but can be applied to experimental, quasi-experimental, and nonexperimental data. Within the context of challenges in carrying out traditional randomized designs in some school settings (Lipsey & Cordray, 2000), confirmatory program evaluation may be especially helpful with understanding causal inference in nonexperimental research. Table 6 outlines six criteria that have been advanced to enhance interpretation of findings in this approach. Bickman (1996) and Reynolds (1998) provide illustrations of confirmatory program evaluation.

Confirmatory program evaluation provides an additional methodological approach that many researchers are now considering in the area of prevention and intervention research. Recognizing the growing awareness of the practical and logistical problems inherent in randomized experimental research and the growing interest in quantitative alternatives, we have included this alternative methodology for coding research in our Task Force efforts. Confirmatory program evaluation, with its strategies for drawing causal inferences from program outcome data, bears some similarity to the methods and strategies used to draw valid inferences from case studies (see Kazdin, 1981; Kratochwill, 1985). Despite the advantages of this type of framework for interpretation of intervention data, consumers of our Task Force findings must be alert to the limitations of nonexperimental and nonrandomized designs in drawing conclusions from a research database.

### **THE CHALLENGE OF EVIDENCE-BASED INTERVENTIONS AND CULTURAL DIVERSITY**

Among the many important issues that the Task Force must address is the efficacy of EBIs within the context of various diversity issues (e.g., ethnic minority and cultural groups, sexual orientation, gender, disability, religious affiliation). Although past landmark events such as the Austin Conference (1975; D. W. Sue & Sue, 1999) and the National Multicultural Conference and Summit (1999) have done much to raise consciousness in psychology about culturally and ethnically different populations, the field of school psychology has been less involved, and perhaps prepared, to deal with the diversity of our society. For example,

TABLE 6. The Six Cumulative Criteria for Conducting a Confirmatory Program Evaluation

- 
1. Temporality of program exposure: At the most basic level of causal inference, the causal variable (i.e., program participation) must occur prior to the measurement of the response to the program or outcome.
  2. Strength of association: The larger the association between program participation and intended outcome (or size of the estimated program effect), the more likely the association represents a real causal effect. Other factors being equal, a program that yields an effect size of one to two standard deviations, for example, is likely to have a meaningful effect on participants even if treatment and comparison groups are not randomly assigned.
  3. Gradient effect (dosage/response): A causal inference is more warranted if, other factors being equal, a monotonic relationship exists between program exposure and the program outcome. That is, causal inference is strengthened if the outcome condition improves as an increasing function of the amount and duration of program participation.
  4. Specificity: Specificity of association refers to the situation in which the program–outcome relationship is linked to certain domains of behavior or outcome conditions.
  5. Consistency: Consistency of association between program exposure and outcome indicates whether the estimated program effect is similar across sample populations similar at different times and places, under different types of analyses, and for similar program theories. The greater the consistency of findings favoring positive effects, the more likely the observed effects are real.
  6. Coherence: At the highest level of causal interpretation is the extent to which the evaluation findings show a clear pattern of effects relative to the causes of behaviors the program is attempting to impact, the target population, the program theory, and the program implementation.
- 

Source: Reynolds, A. J. (1998). Confirmatory program evaluation: A method for strengthening causal inference. *American Journal of Evaluation*, 19, 203–221. Reproduced with permission.

the National Multicultural Conference and Summit was hosted by Divisions 17 (Counseling Psychology), 35 (Society for the Psychology of Women), and 45 (Society for the Psychological Study of Ethnic Minority Issues). The need for our profession of psychologists, and specifically the Task Force and Manual Subcommittee, to consider issues of race, culture, and ethnicity is validated by approximately 45% of students attending public schools being racial or ethnic minorities (Sue, Bingham, Porche–Burke, & Vasquez, 1999).

The challenges embedded in figuring what goals, objectives, criteria, and procedures that the Task Force needs to adopt to reflect a vision that is sensitive to multicultural values and practice is complex for several

reasons. First, traditionally many intervention researchers have operated from the assumption of the universality of intervention effectiveness, which is the belief that interventions are equally effective across diverse cultural groups (Doyle, 1998). Because it is likely that standards of practice, codes of ethics, and policies adopted in the school psychology profession reflect the biases of the larger society, our beliefs about when, what, and how to intervene may be culture bound. Such culture bound beliefs about the nature of intervention may take on invisible forms of unfairness, for example, in biasing researchers via a failure to address cultural issues in EBI research agendas.

Second, there is a paucity of intervention research in school psychology that focuses on the diverse populations present in schools (Ramirez, Lepage, Kratochwill, & Duffy, 1998). Although there are signs of attempts to address this situation, the inclusion of multiculturalism and ethnic diversity may entail using new or novel paradigms for understanding the dynamics and mechanisms influencing intervention outcomes. Many intervention concepts, theories, and approaches have been developed from a predominantly Euro–American perspective and context, which may have limited or partial application to the emerging ethnic and culturally diverse populations of our schools. Currently, it is unknown if interventions are equally effective across diverse groups at the empirical level. Thus, understanding the impact of culture and ethnicity must be one of our top priorities.

Third, the range of cultural diversity definitions has traditionally been narrow with templates established from the high frequency minority groups, government definitions as related to issues of discrimination, and a lack of consideration to the complexity of cultural makeup. As a result, the issue of which groups represent what cultures is unclear and requires unenviable decisions and perspectives. Given this context, the question is what the Task Force can do to address these issues in our future work on EBIs. Following is a glimpse of what we have done and propose to do to reflect the heterogeneous and diverse nature of intervention work.

### Cultural Awareness

Our first goal of the Task Force to facilitate multicultural awareness has been to establish a “strand” within the organization of our group to focus on cultural diversity issues and capture the spirit of a multicultural perspective. Subsequently, recognizing that cultural concerns could not be separated from the content domains, the Task Force decided to review cultural factors within each content domain. In deciding to infuse cultural concerns within the content domains, Multicultural Issues as a

separate content domain was eliminated, similar to the way in which the Prevention domain evolved. Recently, however, we recognized the need to re-establish an explicit focus on cultural issues to provide guidance to the domain chairs and to address the many complex issues in taking into account cultural factors. Thus, we have re-established the Multicultural Issues domain as a separate stand and renamed it Cultural Considerations. Similar to the domains, the multicultural strand has co-chairs with ethnic minority representation. As we advance our work in review of literatures it is expected that our multicultural strand co-chairs will help recognize and resolve issues involving multicultural values and biases. In addition, we expect the strand co-chairs to advocate for multicultural and cultural competence in the processes and procedures stemming from the *Procedural and Coding Manual*. Specifically, aside from the coding of studies and literature review process we have commissioned a working paper from our co-chairs that will focus Task Force attention on cultural diversity issues. Although the precise nature of dialogue framed in such a paper has yet to be developed, possible issues include: (a) defining culturally diverse groups, (b) determining principles and practices of adapting interventions for diverse cultural groups, (c) facilitating difficult intervention decisions and considerations related to race and ethnicity, and (d) establishing priorities for developing multicultural perspectives in school psychology research and practice.

In addition, we anticipate the need to structure the future training agenda (see below) of the Task Force on issues of cultural diversity. Our entire domain co-chairs and the to-be-selected reviewers of intervention studies must have the advice and direction from multicultural experts as well as the research and scholarly literature on ethnic minority and cultural issues. It is clear that this agenda will require heightened cultural competence and access to dialogue at personal, professional, organizational (Division 16, NASP, SSSP), and societal levels on topics aimed at accommodating diversity in prevention and intervention activities.

#### Considering a Cultural Specificity Hypothesis

In contrast to the hypothesis that EBIs are equally effective across all cultural groups, we want the Task Force to consider alternative hypotheses such as cultural specificity. This hypothesis is that intervention effectiveness differs across different cultural groups depending on the unique cultural features of the group (Tharp, 1991). As Doyle (1998) has noted, data bearing on which of the two hypotheses is supportable are not available. Nevertheless, our position will remain conservative in

that we will operate from the cultural specificity position and consider empirical evidence to the contrary. Embracing a cultural specificity hypothesis is consistent with the focus of the Task Force on context variables. It is also consistent with data suggesting that even biological responses to psychotropic medication for mental health issues such as depression and anxiety can vary due to difference in race and ethnicity (Kazdin, 1999; Lin, Poland, & Nakasaski, 1993). By focusing on the role of context in examining intervention effectiveness, issues of race, culture, and ethnicity cannot be viewed as irrelevant or hidden. The intent of putting cultural specificity on the table as a relevant topic to consider should help promote viable and productive dialogue about any potentially detrimental interpretations for any “victims of discrimination” (Fine, Weiss, Powell, & Wong, 1997).

#### Structuring Research on Culture and Interventions

A major impact that the Task Force can have on the future of the relationship of cultural diversity to EBIs is to promote research on this issue. The agenda becomes one of promoting research on the differential effectiveness of interventions across ethnic and cultural groups. The agenda is not restricted to understanding differential effectiveness but one of assuring that stereotyping and discrimination does not occur, however unintentional or naïve in nature. Intervention research, as recommended by Kazdin, will need “to sample broadly, to evaluate the moderating role of sample differences, and to pursue mechanisms through which moderating factors may operate” (1999, p. 20). Such a perspective will likely require considerable initiative and resources. For example, when the APA Division 12 Task Force on Psychological Interventions (Chambless et al., 1996) took on the charge of determining which psychotherapeutic interventions had met their criteria for empirically validated treatment, no rigorous study had been with members of ethnic minority groups. The problem of generalizability of findings to ethnic minority groups was accommodated by shifting the construct from “empirically validated treatment” to “empirically supported treatment” (Chambless & Hollon, 1998).

A research perspective that examines the role of cultural diversity to EBIs also will likely require a rethinking of research design and interpretation. S. Sue (1999) has proposed several solutions to increasing the quality and quantity of research to incorporate ethnic minority populations that have relevance for the Task Force, including: (a) minorities must be included in all human participant research; (b) cost should not be considered a viable excuse for excluding ethnic minorities, and (c) outreach efforts should be supported to include these groups in clinical

research. As S. Sue pointed out, scientific principles also need to be rethought with increased attention to external validity (generalization and relevance to the realities of school life and interventions for ethnic minorities) rather than the traditional focus on internal validity (methodological precision and purity). Research across our domains will extend the sensitivity in practice already established by the APA in 1993 through the Guidelines for Providers for Psychological Services to Ethnic, Linguistic and Culturally Diverse Populations to empirical guidance for selection and implementation of EBIs (see Doyle, 1998).

#### Structuring the Literature Review and Future Research on EBIs

To facilitate the integration of a cultural diversity agenda into the Task Force activities we need to address some important issues in our review and future research efforts. First, research reviews of various interventions must be coded for information that leads toward an understanding of cultural contextual variables. In this effort, we must identify the core areas or issues related to potential multicultural bias as well as those that have application for diverse individuals. We are currently working on this aspect of the *Procedural and Coding Manual* and hope to unveil a draft of this component in the near future. Second, in our review efforts we are directing our Task Force co-chairs to identify culturally diverse groups in the research investigations and where possible, calculate separate effect size statistics for these groups. Over time, single-participant and group design meta-analysis may allow us to make some differential conclusions about EBIs when taking into account cultural issues. Third, consistent with the perspective of S. Sue (1999), D.W. Sue et al. (1999), and others who have addressed multicultural issues, we need the financial commitment of our professional organizations and funding agencies to support research on cultural diversity and EBIs.

#### **USING EMPIRICAL EVIDENCE TO CLOSE THE GAP: PRACTITIONER CODING OF EVIDENCE-BASED INTERVENTIONS**

As noted earlier, the notion that the knowledge base on effective intervention remains mostly unknown to practicing psychologists is not a new or novel thought (Kazdin et al., 1986; Nathan & Gorman, 1998). There are at least three related concerns that surround the impetus for making scientifically based evidence on interventions more accessible to clinicians. The first concern is that practitioners tend to be more influenced by clinical judgment than research on treatment effectiveness in designing and evaluating their own interventions (Wilson, 1996a,

1996b). A second concern is that many psychologists endorse a “one size fits all” approach (often referred to as the equivalence of therapies hypothesis) or the belief that doing something is better than doing nothing at all (DuPaul, Eckert, & McGoey, 1997; Nathan & Gorman, 1998). Due to these eclectic practice tendencies where psychosocial interventions are viewed as equally effective, psychologists may fail to conduct a careful analysis of their intervention plans and decisions. Third, the integration of an empirical basis into practice does not match the day-to-day demands of practitioners’ lives. Thus, even when psychologists are aware of the empirical evidence, they fail to infuse this evidence into practice because it requires more work or resources. In response to these concerns about intervention practices the perspective of the Task Force is straightforward: We need to move in directions that will yield functional and meaningful information for practicing school psychologists. To accomplish this objective the Task Force has examined and proposed several activities to narrow the research–practice gap.

Within this dialogue, we have grappled with various questions related to intervention effectiveness where researchers and practitioners might contribute to the knowledge base. These questions are not limited to what knowledge is needed (e.g., What therapist characteristics are most important to evaluate in determining intervention effects? How does one’s theoretical orientation influence intervention decisions and outcomes? What period or duration of time should be applied to determine “long-term” outcomes?), but also with what knowledge is known (e.g., When should a problem such as reading delay, anxiety, aggression, or family dysfunction be considered “clinically significant” by school personnel and to necessitate psychosocial intervention? What psychosocial interventions work for which characteristics of ADHD? When should a small group intervention occur rather than individual-based intervention?). One critical distinction often made in the literature on EBIs that bears on how these questions are addressed is to differentiate *effectiveness* and *efficacy* (Chambless et al., 1998; Nathan & Gorman, 1998; see also American Psychological Association, 2000).

Efficacy is the standard for evaluating interventions in controlled research, whereas effectiveness is the standard for evaluating interventions in a practice context. Efficacy studies are generally conducted in laboratories or clinical research facilities and use well-designed and precise methodology. Conversely, effectiveness studies are mostly focused on issues of generalizability of the intervention, and consider such issues as the intervention’s feasibility, use, and acceptability across different settings and clients. In our view, both efficacy and effectiveness studies are sorely needed by the school psychology profession and should be evaluated. The criteria presented by our Task Force are designed to ad-

vance broadly the knowledge base on evidence-based practices. Although the explicit sensitivity of the Task Force to school context considerations may suggest a strong endorsement of effectiveness methods, efficacy research is critical and will be evaluated.

Recall that an important goal of the Task Force is to help school psychologists and individuals under their charge make decisions about the quality of research evidence and empirical support for various prevention and intervention programs implemented in school and related educational settings. Specifically, we are interested in providing state-of-the-art efficacy evidence with dimensional criteria pertaining to research support for various intervention programs across the five content domains identified in Table 2. Although this information will be of interest and relevance to researchers in our field, we emphasize strongly that the primary purpose of identifying EBIs is to enable practitioners in schools and other applied settings to use these procedures and programs to serve the educational and mental health needs of children and families. Thus, we also need effectiveness data on interventions. On the surface, this seems like a clear mission. Yet generalization from research to practice settings is not a straightforward process and often psychologists encounter a variety of barriers (Kazdin, Kratochwill, & VandenBos, 1986; Kratochwill & Stoiber, 2000c; Stoiber & Kratochwill, 2000; Weisz, 2000).

The Task Force is considering several tactics for addressing the lacunae between research and practice and facilitating improved generalization (Drotar & Lemanek, 2001; Hayes, Barlow, & Nelson-Gray, 1999; Kazdin et al., 1986; Soldz & McCullough, 2000; Weisz, 2000). Some of the strategies to help generalize the findings from research to practice are in the hands of researchers, some are in the hands of practitioners, and some will require ongoing collaboration. We especially advocate a shared responsibility for ensuring the effectiveness of interventions. To begin with, standardization of the assessment and intervention process is often associated with research on EBIs (a responsibility of researchers). We also recommend that practitioners use manuals or other procedural guidelines to facilitate the process of implementing interventions in practice settings, as long as these permit flexibility and adaptation of the intervention to the specific problem or issue addressed by the practitioner (Beutler, 2000). Although there are controversies regarding the use of manuals in practice (see Kratochwill & Stoiber, 2000c, for an overview and further discussion below), preliminary survey information from practicing psychotherapists indicates that they consider manuals helpful, use them extensively, and have few concerns about their use despite their known limitations at least in clinical settings (Najavits, Weiss, Shaw, & Dierberger, 2000). One of the major problems with manualized

interventions is that they are not available for many school-wide prevention programs or classroom-based interventions. Such interventions incorporate rather complex implementation strategies that simply may not be feasible for long-term or extensive intervention/prevention programs. Nonetheless, intervention manuals can hold an important role in intervention integrity and we are in the process of examining their use in school psychology training and practice.

Second, we recommend the creation of opportunities for training to reduce the gulf between research and school practice activities. One of the Task Force's major goals is to create a database that will be useful to training programs and professional organizations in teaching practitioners, graduate students, university trainers, and other interested individuals to facilitate the dissemination of effective interventions to practice settings. Currently, much of the knowledge base on EBIs has a clinical flavor, and thus, little is known about classroom-based, small-group, or home-school interventions (Stoiber & Reinemann, 2001). School psychology encompasses a diverse context and diverse set of interventions (Christenson & Sheridan, 2001; Nastasi, 2000). Information on EBIs that have application for school psychology must reflect the needs of students, teachers, and families; classrooms and schools; academic, social, and health-related concerns. Our goal is to offer professional continuing education for school psychology researchers and practitioners at the state, regional, and national level. If we are serious about narrowing the hiatus between research and practice, it seems important that attention should be given to how the manual criteria are applied and used to identify EBIs. Specifically, the descriptive features and experimental procedures that are associated with strong evidence should be explicated for researchers and practitioners. Practitioners would likely benefit from information related to decisions on how to proceed when implementing interventions and what should be weighed when confronting intervention situations that are not clearly identified as matching evidence-based practices. Researchers would benefit from exchanges with practitioners on how intervention decisions are determined and what manual criteria and procedures are regarded as most useful. Obviously, practitioners and researchers alike would benefit from the collaborative spirit that occurs through shared sense making on how to apply and enhance the products associated with the work of the Task Force. Both need the most up-to-date and best sources of information available on EBIs. Like Division 12 and 53, we have created a Web site that will provide updated information on our Task Force activities and EBIs (see [www.sp-ebi.com](http://www.sp-ebi.com)).

A third tactic to strengthen the connection between research and practice is to recommend that practitioners embrace a scientist-practitioner

model in their professional work (Soldz & McCullough, 2000). To aid in the decisions that practitioners will need to make to implement an intervention in their practice setting, we hope to provide extensive information on intervention program implementation context. Traditionally, research on child interventions and especially on child psychotherapy has not been very context sensitive (Weisz, 1998). By context sensitive we mean that the research should include the context in which the child functions (e.g., with parents, teachers, peers) as part of understanding efficacy of an intervention. Various ecological or contextual factors will need to be understood in assessment, intervention planning and implementation, and outcome assessment. Our Task Force will hopefully be able to provide context information by considering this information across studies in a particular intervention domain. Yet, despite our best efforts to present context and various issues relating to implementation, it is quite apparent that no series of research investigations (both efficacy and effectiveness) can take into account all contextual and related factors that influence whether an intervention is likely to be beneficial in a particular context such as a school or classroom. Therefore, it is essential that practitioners engage in self-evaluation and invoke methods of empirical clinical practice to facilitate an understanding of how beneficial a previously supported intervention will be when applied in a new school context or different situation.

As one illustration of empirical clinical practice, we hope that the process of clinical replication (Kazdin et al., 1986) will be invoked when practitioners conduct effectiveness studies of interventions under typical practice conditions. Single-participant design, case study investigations, and related methods might be used to facilitate not only the evaluation of an intervention within a particular context (see Morgan & Morgan, 2001), but also information sharing among researchers about how to revise or redesign intervention programs to make them more effective (Hayes et al., 1999). The Task Force hopes to promote the use of research findings in practice, and during the later stages of our activities, we hope to make this goal part of our dissemination agenda. The next section sets forth some strategies for advancing this agenda.

#### Specific Task Force Strategies

*Scope of the Issues.* To address the research-to-practice gap, our Task Force is pursuing some strategies that we hope will facilitate the integration of EBIs in practice and help researchers modify interventions to make them more effective in practice settings. The three strategies have their conceptual underpinnings in establishing a link between research and practice to better understand intervention effectiveness

(Kratochwill & Stoiber, 2000d; Soldz & McCullough, 2000). The strategies are (a) developing a practice research network in school psychology, and (b) establishing guidelines that school psychology practitioners can use in evaluating EBIs in practice.

*Practice Research Network.* A practice research network is a group of practitioners who engage in the evaluation of clinical replication outcome programs of research (Hayes et al., 1999). The clinicians typically provide information to a research team on a wide range of variables including, for example, diagnosis, evaluation procedures, intervention, and cost of services. The network is often put online to facilitate data collection and management. Practice research networks have been established by the American Psychiatric Association and the Practice Directorate of the APA (see Hayes et al., 1999, pp. 259–260, for discussion of these efforts). The research practice network is one way to facilitate the important agenda of involving practitioners in the research process (Keith, 2000; Kratochwill & Stoiber, 2000d).

Over the next several years, the Task Force will be working on the development of a practice research network in school psychology. The purpose of the network will parallel that of networks in other fields and will engage practitioners in schools in evaluating various EBIs. To accomplish this goal, the Task Force will need to secure funding, solicit volunteer participants from various schools and educational settings, and train these individuals in intervention evaluation protocols. We hope our national organizations (APA Division 16, the National Association of School Psychologists, and the Society for the Study of School Psychology) will endorse and support this effort. The protocol, which remains to be developed, could include some of the components described in the next section. Ultimately, the goal of this effort is to strengthen the connection between research and practice and create more efficacious interventions.

*Guidelines for Practitioner Evaluation of Interventions.* Practitioners can be considered the primary consumers of EBIs. Unfortunately, what we have learned is that most clinical training programs and internship sites are not involved in teaching EBIs to future psychologists in clinical practice (see Crits-Christoph, Chambless, Frank, & Brody, 1995). Although we have less information on what is happening in school psychology training programs, and especially in school-based practice, we have begun to gather information on these practices in a Task Force-sponsored survey of training programs (see Shernoff, Kratochwill, & Stoiber, 2002).

Practitioners are interested in how a particular intervention can be generalized to their settings—that is, they engage in “logical generalization” of research findings to their practice (Hayes et al., 1999). Hayes et

al. (1999) presented guidelines practitioners can use when they consume intervention research, and we have adapted those guidelines in a checklist format that could be used in practice evaluation of EBIs. We plan to use this framework with the network of practitioners who are willing to evaluate interventions in practice settings and provide feedback regarding their utility using Task Force criteria. This information will then be added to the knowledge base of available data on a particular intervention as it is disseminated by the Task Force. Table 7 is adapted from the work of Hayes et al., (1999) and will eventually be integrated into the coding format for the Task Force manual. We believe that "informed school practice" on interventions can only occur when practitioners have provided input on how an intervention can be generalized to particular school contexts (Kratochwill & Stoiber, 2000d; Stoiber & Kratochwill, 2000)<sup>7</sup>.

#### Enhancing Guidelines for Interventions

Once EBIs have been identified and made available to practitioners, issues emerge surrounding whether and how such strategies will be accepted and adopted in schools and other educational settings. Many concerns may prohibit the use of EBIs in practice, such as training, acceptability, theoretical allegiance, cost, and logistics, to name just a few. For example, allegiance to functional analysis in the tradition of applied behavior analysis may inhibit some school psychologists from adopting cognitive therapy, an intervention for anxiety problems in youth that has already been identified as an EBI by our child clinical colleagues (see Kendall, 2000). In addition, the use of EBI manuals in practice raises a multitude of concerns over adherence to the process and content of interventions, whether as part of psychotherapy in clinics and hospitals (Beutler, 2000) or school-based interventions (Kratochwill & Stoiber, 2000c). Given these concerns, what can be done to integrate EBIs into future practice in school psychology?

One option might be to educate graduate students and practitioners in guidelines that extend beyond the application of specific EBI approaches. These enhanced guidelines would be designed specifically to educate graduate students and school psychologists in a variety of strategies that could promote the use of EBIs, while at the same time addressing many of the concerns raised by critics of the EBI movement.<sup>8</sup> One work group in psychotherapy training, which addressed, in part, con-

---

8. Our concept of enhanced guidelines was stimulated by the work of Beutler (2000). Beutler has developed "optimal and enhancing guidelines" for treatment of depression.

TABLE 7. Example Dimensions to Consider in Use of Evidence-Based Interventions in School Psychology Practice

<b>Dimension Considered in Practice</b>		
1. Does your client appear similar to those described in the EBI?	Yes	No
2. Are you able to replicate the intervention based on the description provided in the manual and/or procedures?	Yes	No
3. Are the conditions of implementation of the EBI similar to those of your setting?	Yes	No
4. Are the measures you used to assess outcome identical to those used to establish the EBI?	Yes	No
5. Were all the measures recommended in the EBI used to evaluate the intervention?	Yes	No
6. Was ongoing evaluation (repeated assessment) of student progress conducted?	Yes	No
7. Have new outcome measures been added to the intervention evaluation?	Yes	No
8. Did you use an intervention manual or protocol that specified the intervention?	Yes	No
9. Can individual characteristics of students be identified that are related to intervention outcomes?	Yes	No
10. When group intervention data are reported, is the percentage of individuals showing the effect reported?	Yes	No
11. When individual data are reported, have the data been presented in graphic form?	Yes	No
12. Have the EBI positive effects reported in research been replicated with your student(s)?	Yes	No
13. Have you replicated the EBI more than once?	Yes	No
14. Have others in your school setting replicated the EBI?	Yes	No
15. Would you rate the effects as strong as the original EBI effects?	Yes	No
16. Would you rate the effects as clinically meaningful?	Yes	No
17. Did you and/or staff find the EBI acceptable for use in your school?	Yes	No
18. Was the EBI cost-efficient for implementation in your school?	Yes	No
19. Are there specific contextual factors in your setting that could account for the success or failure of the EBI?	Yes	No
20. Do you plan to adopt the EBI for future implementation in your setting?	Yes	No

Source: Reproduced and adapted with permission from: Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. O. (1999). *The scientist practitioner: Research and accountability in the age of managed care*.

cerns over the use of intervention manuals, arrived at several suggestions for the content of future psychotherapy training (Caspar, 2001). These suggestions, as well as those offered for facilitating empirical clinical practice (e.g., Kazdin et al., 1986; Stoiber & Kratochwill, 2002), may foster a dialogue on what the content of future school psychology preservice and inservice training might include.

*Focus on Understanding Basic Principles of Change.* In concert with

the notion that there are some generalized concepts and strategies involved in all interventions, it may be prudent to teach some basic principles and strategies of behavior change when recommending the use of EBIs. An understanding of such principles and strategies is needed to move trainers and practitioners beyond a focus on one theoretical model or technique and help them become good problem solvers able to deal with the complexity and uncertainty surrounding all interventions. Transtheoretical analysis is an example of an approach that focuses on change independent of any one theoretical model (e.g., Prochaska, 1984; Prochaska, DiClemente, & Norcross, 1992).

*Focus on Understanding Indications and Contraindications of EBIs.*

Ideally, when an intervention is identified as “evidence-based,” guidelines for when it should and should not be used should be available from the research literature. Our work with the Task Force suggests that this information is not available in most cases, although we are committed to providing data on issues such as educational context and application with minority populations, to name a few. There are many reasons for this state of affairs, including, for example, the lack of emphasis on replication research and the bias against publishing negative results research (see Kratochwill, Stoiber, & Gutkin, 2001). The primary responsibility for specifying indications and contraindications of an EBI should rest with the researchers who develop the intervention. Our Task Force may be able to share in this responsibility through the research practice network process outlined above. Nevertheless, it will be critical to establish training agendas that specify the range and scope of applications for EBIs in practice.

*Focus on Understanding the Variability in Intervention Implementation.* A concern related to the contextual conditions under which EBIs are implemented is how variable intervention implementation is likely to be under conditions of varying skill and ability of intervention agent(s). In clinical psychology, this concern has focused primarily on the skill of the psychotherapist. In school psychology, however, the intervention agents are diverse—potentially including the school psychologist but also parents, teachers, counselors, administrators, and students’ peers—a factor that has major implications for training (see Kratochwill & Pittman, in press). EBI outcomes are likely to vary when consultants and consultees receive minimal training. These concerns should prompt us to emphasize the need for competencies in EBIs, training of consultees in these EBIs, and supervision of EBI implementation in practice to promote integrity of the procedures.

*Focus on Teaching Problem-Solving Strategies for EBI Planning.* Selection and implementation of EBIs in practice will be complex at best. As our discussion of assessment processes linked to problem identifica-

tion for EBI matching illustrates, one dimension of the task complexity relates to the selection of interventions. We endorse a steady and persistent posing of the question, "What is the evidence for a selected practice?" The answer to this question is also complex, and will vary based on various concerns, circumstances, and populations of children and families. In addition, the selection process must take into account the role of the intervention agent (e.g., consultant, therapist) who basically is the problem solver in the intervention selection process. The individual responsible for making the selection of the intervention must be aware of problem-solving strategies and tactics that allow a careful matching of the nature of the problem to the EBI. The procedures for evaluation of practice can often be helpful in this regard (see below).

*Focus on Teaching Evaluation of EBIs in Practice.* No matter how much evidence is amassed for a particular intervention, the EBI will need evaluation as it is applied in practice. In some cases, evaluation protocols were constructed along with development of the EBI and may even be represented in the manual. In other cases, generic protocols for evaluating any intervention might be used to facilitate the evaluation of an EBI. As an illustration, we have developed a protocol called *Outcomes: Planning, Monitoring, Evaluating (Outcomes: PME)* that facilitates selection of problems, design and implementation of interventions, and evaluation of outcomes (Stoiber & Kratochwill, 2002). Developed around a goal attainment and progress-monitoring framework, *Outcomes: PME* offers an example protocol for training and practice in intervention planning and monitoring. It has application for practitioners and researchers in conceptualizing and conducting outcome assessment in conjunction with intervention implementation.

*Focus on Self-Evaluation Strategies for EBI Implementation Agents.* Engaging in and embracing the EBI movement is not just about providing quality interventions in school psychology in the short term. It also requires a commitment to continuing professional development and self-evaluation in the long term. Both the APA and the National Association of School Psychologists focus on this domain of conduct in their professional code of ethics, and it is relevant to the facilitation of EBIs in practice. Selection of EBIs for a specific concern given a specific set of circumstances and the continuing education required to implement them or supervise their implementation are truly about "best practices."

### FOCUS OF FUTURE EDUCATION EFFORTS

The issues and considerations that have been discussed thus far lead us to a final important agenda of the Task Force: To educate both researchers and practitioners on concepts and materials associated with identifi-

cation, review, and dissemination of EBIs. The training mission can be organized along several dimensions including (a) training on use of the *Procedural and Coding Manual*, (b) specialized training in conducting literature reviews, and (c) dissemination efforts and training. Each of these education foci is presented briefly to help readers conceptualize the purpose and content of Task Force education and dissemination efforts.

#### Training on Use of the Procedural and Coding Manual

The content of training focused on the *Procedural and Coding Manual* will target the focus and scope of the domains, purpose and use of the coding process, and special topics needed to conduct and produce effective and high quality literature reviews. Each of these issues is discussed briefly.

*Task Force Domains.* Currently the Task Force is focused on five domains with the prevention area embedded across each of the domains (see Stoiber & Kratochwill, 2001). An important training agenda will be to assist potential reviewers with a conceptual framework for what is included in each domain. The content of this training will focus on the organization of domains, what is included and excluded for potential review, and the relationship of the domain under consideration to the other domains. The domain co-chairs will likely carry out this training with the assistance of other Task Force members. In the training it will be critical for the domain co-chairs to define clearly the range and scope of their domains and coordinate the descriptions with other domain co-chairs. We envision that domain co-chairs will develop a detailed document that specifies their vision for the organization and content of their respective domain.

*The Procedural and Coding Manual.* The centerpiece of the Task Force work is the *Procedural and Coding Manual* and its application to the literature review process. The focus on training on the manual will take several directions and will have several important functions. First, individuals need to be trained on the respective codes and coding process as applied to the domain in which the literature review is being conducted. Currently, two case examples of the coding process have been developed (Lewis-Snyder et al., 2002 and Shernoff et al., 2002). Domain co-chairs are also developing several additional case study examples that illustrate various coding challenges within their respective areas. The case studies coding applications would feature, for example, applications of the coding process with mixed methodology/design studies, negative result outcomes, multicultural and diverse samples, and particularly complex investigations that may be challenging to code. Second, training in the *Procedural and Coding Manual* would be focused on establishing the accuracy and reliability of the coding process. Although

we have conducted initial reliability trials with the manual, we view the Coding Manual as an “evolving product” and ongoing accuracy and reliability checks will be necessary, especially as a broad range of studies are coded. The intent of this training is to increase both the accuracy of coding (against developed example coded investigations) and the reliability of the coding across independent coders (individuals blind to each others’ coding).

### Specialized Training in Conducting Literature Reviews

Conducting literature reviews is a complex process and can be considered a specialized area to target for Task Force education efforts. Three major areas will be targeted for training and include the literature review process, specialized statistical treatment of the data (e.g., data analysis, effect size issues), and methods of creating summaries and reporting findings.

*Conducting the Literature Review.* A rather extensive literature on how to conduct literature reviews has been produced (e.g., Cooper, 1998; Hart, 2001; Light & Pillemer, 1984; Slavin, 1986). We will draw extensively from this area and feature in training some of the central issues surrounding the best practices for the literature review process including, for example, formulating a problem, data collection, data evaluation, analysis and interpretation, and public presentation (Cooper, 1998).

*Statistical Treatment of the Data.* Although the literature review process typically includes some content devoted to data analysis, we perceive the need for specialized training on some less familiar statistical issues including, for example, group and single-participant research design meta-analysis, qualitative data analysis, and confirmatory program evaluation models. For example, although it is likely that many individuals who conduct reviews for the Task Force will be familiar and competent in the use of group design meta-analysis, the application of these procedures in single-participant research may be less familiar (e.g., Busk & Serlin, 1992). Specialized types of statistical treatment, such as computing effect sizes, are also expected to require attention. As noted earlier issues regarding external validity and qualitative design features, which frequently receives less attention in the literature, should be included in the training agenda. We are hoping to engage our methodological domain Task Force group in these training efforts as well as solicit the assistance from experts in the field of psychology and education.

*Reporting Findings of Literature Reviews.* The reporting of findings from literature review efforts will need to be a special topic of training. The issue of reporting findings from literature reviews is part of the pro-

cess of a comprehensive approach to this area (see Cooper, 1998) but special issues are associated with the Task Force activities. To begin with, our reporting will need to take into account the specialized coding framework associated with content areas in the manual. Second, we will need to develop a summary format linked to the review codes that is user friendly to both researchers and practitioners who hope to use the findings in their research and practice, respectively.

#### Dissemination Efforts and Training

As noted previously in the article, the ultimate goal of the Task Force is to disseminate findings and have a beneficial influence on a number of groups who will profit from our efforts. Thus, our goal is to offer educational and training opportunities to a number of constituencies. To begin with, we hope to offer continuing education opportunities to individuals whose primary work is in the practice of school psychology in schools. Opportunities for this target exist through state, regional, and national organizations (e.g., APA, NASP). Second, we hope to offer similar opportunities to researchers and scholars and will target similar professional and scientific groups (e.g., SSSP). Third, we hope to influence both faculty and graduate students in our school psychology training programs. To truly have a far-reaching impact, the knowledge base on EBIs must be integrated into course work. We are aware of some school psychology programs that have begun to offer coursework on EBIs, and other programs that have formulated EBIs as an area of concentration. Another approach for integrating content on EBIs into graduate training would be for universities to encourage cross-disciplinary courses (offered co-jointly by school, counseling, and clinical departments) or interdisciplinary concentration courses on EBIs. To date, the work of the Task Force has centered mostly on constructing the manual and conducting reviews of the empirical literature. However, the *Procedural and Coding Manual* may offer some helpful information to training programs in the teaching of research design, methodology, and data analysis. Finally, we hope that our efforts will influence policy makers as these professionals consider programs for funding and adoption in schools, and accreditation organizations as they formulate and apply accreditation criteria.

#### The Challenge of Developing a Knowledge Base on EBIs: Forging a Balance of Scientific Principles and Practice Realities

At the time of this writing, the ultimate impact of the Task Force's efforts at identifying EBIs in school psychology remains uncertain. A first ini-

tiative of the Task Force has been to mobilize what is known about effective interventions into future research protocols and school psychology practices through the construction and application of the *Procedural and Coding Manual*. As articulated by Hughes (2000) and Stoiber and Kratochwill (2000), the goals of an endeavor to promote a knowledge base on EBIs are ambitious, encompassing school psychology practice, training, and research. However, if the history of ongoing tension between research and practice within psychology (Kazdin et al., 1986; Nathan & Gorman, 1998; Rice, 1997) are any indication, it seems prudent to be cautionary, if not perhaps vigilant, of potential constraints and opportunities surrounding the efforts of the Task Force. As noted by Schneider, "with rare exceptions, the science and practice of psychology have failed to give meaning to each other" (1990, p. 524).

In the process of developing the *Procedural and Coding Manual*, we have assessed and reassessed an important question: What goals and procedures must our Task Force and profession adopt that will serve as a visionary template both for practitioners and researchers? We believe that an answer to this question can only be found in an open forum that invites discussion and dialogue. The answer also lies in accepting that scientists and researchers see the world from different perspectives, yet realizing that our worlds are intertwined and interdependent (Cantor, 1999). One purpose of this article is to serve as a catalyst for dialogue both from researchers and practitioners and to allow opportunities for input from diverse sources within the field of psychology and school psychology.

As we look to the future we hold to the belief that our efforts must be collaborative and determined to connect research and practice in meaningful ways. Many potential ideas and suggestions for the *Procedural and Coding Manual* were distilled from prior exemplary efforts within the field of psychology and psychiatry and from our Task Force and Manual Subcommittee. In some regards, the decisions that we made may be viewed at odds with other available conceptualizations and coding schemes. Thus, we want to visit choices taken within the goal of operating in a spirit that advocates on behalf of practitioners and researchers. We hope to set a path that fosters an avenue for partnering and cross-fertilization of practice and science. We also hope to reduce skepticism about the role of EBIs in school psychology by scrutinizing potential trajectories that may unfold.

The first potential trajectory stemming from the work of the Task Force may lead to barriers between school psychology and other fields of professional psychology. The structure, criteria, and dimensional coding scheme of the *Procedural and Coding Manual* are unique and distinct from the foundational classification systems developed by the APA

Division 12 Task Force. One possible result of this decision is that it will lead to confusion in the field of psychology, especially for practitioners. The distinct coding schemes embraced by different divisions, groups, and societies in psychology may produce different types of information, judgments, and conclusions about findings from research. As mentioned above, due to differences in coding systems there will likely be differences in how prevention and intervention programs are assessed. We will probably not be able to establish “universal” application or evaluation of intervention methods, strategies, techniques, or programs. Thus, in addition to a lack of clarity about what programs meet the “gold benchmark or standard,” there may exist a widening gap within clinical–school–counseling psychology. Along such a trajectory there would be missed opportunities for the various fields of professional psychology to inform each other on the evidence in support of or opposition to intervention approaches. This outcome has never been the intent of the Task Force.

Rather the Task Force argued that if a narrow set of evaluation criteria were adopted for all intervention types without regard to the realities of research and practice in school settings, then school psychology practitioners would perceive the EBI effort as unrealistic and not germane to the work they do. Similarly, if evaluating the evidence base of studies results in highly exclusive lists of “approved” interventions and programs that do not accommodate the complexity and diversity of schools and school psychology practice, then these efforts will likely be ignored as irrelevant. The intent of the Task Force is to promote reflective, professional decision making in researchers and practitioners alike. The intervention evaluation guidelines as outlined in the *Procedural and Coding Manual* and above are hoped to promote this reflective agenda.

As has been noted, another potential trajectory of the EBI efforts is that it will lead to theoretically based schisms within the profession of school psychology. It is possible that through the evaluation process, interventions linked with one theoretical approach (such as behavioral, cognitive–behavioral, psychodynamic, biological based or psychopharmacological) may be found to have greater empirical evidence than another approach. Such recognition for one approach/model over another may lead to serious schisms within the profession where one theoretically aligned group espouses to have an “evidence–base” or to use “validated” practices more than those who adhere to a different theoretical model. Along such a trajectory, it is also possible that the “endorsed” list of interventions resulting from the evaluation process will lack alignment with the “humanistic” or interpersonal orientations of many educators. As a result, consumers of school psychological services (e.g., teachers, parents, children) will likely view

the role of the school psychologist as lacking relevance at precisely the time when school psychologists are seeking greater integration into the day-to-day functioning of schools (e.g., DeAngelis, 2000; Oakland & Cunningham, 1999; Stoiber & Waas, in press).

Our decision to incorporate a dimensional coding scheme that examines internal and external validity, and where levels of evidence function as “benchmarks” for evaluating what is known was based on an important objective: To develop a creative structure for increasing communication and understanding both across psychology and within school psychology. Our thinking was that a rigid set of “endorsed” intervention options would be viewed as suggesting that interventions occur in a vacuum rather than a rich context. It may also suggest that knowledge development is unidirectional, flowing from scientist to practitioners. Unfortunately, even the best controlled efficacy studies may produce flawed conclusions because the research design may bear little resemblance to the realities of schools. We also felt that a dimensional coding scheme would facilitate rather than constrict professional decision making. The intent is that coding the evidence in support of intervention programs will extend the question from “What do I know?” to questions such as “How do I know?” “When does it apply?” and “What can be done to know more?” A third possible trajectory of the Task Force efforts, thus, is considered most probable. Following this trajectory, the coding of intervention studies provides a collaborative network where practitioners inform research agendas and research psychologists inform practice. This agenda would occur through the Task Force extending open communication lines with practitioners through ongoing feedback loops and evaluation modalities. It would also occur through providing conduits wherein Task Force members, researchers, and practitioners work together in summarizing for intervention scientists what practitioners need to further relevant research. Following this trajectory, an information rich database is co-constructed about a variety of school-based interventions and programs. This database is disseminated widely by traditional avenues such as journal articles and conference presentations and by nontraditional means such as electronic media and web sites on the Internet. It is also hoped that new avenues would be made available, such as a section of a school psychology journal or newsletter that would offer “science/practice dialogue on evidence-based practices.” Importantly, practitioners will have access to detailed ratings of interventions on a variety of important dimensions. Based on this information, practitioners are able to weigh the relative merits of an intervention and make more informed judgments about its appropriateness for the identified target objectives, characteristics of the

children and families, and the demands of the setting wherein the intervention occurs.

Thus, there is an auspicious sense that surrounds the development of the manual and the larger EBI efforts. Practicing school psychologists are expected to develop hypotheses about what and why an intervention works and to work toward recognizing evidence that has been gathered. As a result of the increased rigor in intervention practice and decision making, teachers, children, families, and other consumers are more satisfied with the precision, relevance, and effectiveness of interventions. Finally, researchers in psychology and school psychology become more productive through collaborative interactions. These interactions catalyze focused commitment to forging “unexplored territories” that future research on interventions should and must address.

### SUMMARY AND CONCLUSIONS

As school psychology begins the task of identifying EBIs it is faced with significant challenges and opportunities. One might infer from this article that the multiple constituencies, priorities, and issues considered in developing the *Procedural and Coding Manual* are so broad that it will be difficult to maintain a viable focus. Clearly, the movement toward EBIs in school psychology is bigger and more demanding than what the Task Force alone can facilitate and implement. We face this task cognizant that prior EBI efforts within other psychology, psychiatry, and professional groups have been met with extensive controversy. Yet the vision of the Task Force in school psychology goes well beyond these earlier efforts in the plans and procedures that have been established, potentially leading to greater dilemmas, confusion, and criticism.

Just the level of coding competence required to evaluate intervention programs and create a meaningful knowledge base is daunting. The ethical and professional obligation to produce a knowledge base on EBIs that correspond to the diverse needs of practitioners and researchers is even more overwhelming. However, the complex needs of children and families, the eclectic nature of school psychology service delivery, the diversity of settings in which school psychologists function, and the need for research-based strategies to create optimal social and learning environments leads us to a clear focus. We remain steadfast in the belief that constructing, organizing, and disseminating evidence to support intervention choices is a laudable and necessary endeavor. Although the EBI effort holds many challenges, it should lead the profession to heightened levels of competence and newly shared experiences between researchers and practitioners. To produce meaningful outcomes the EBI effort must facilitate initiatives that include: (a) active involvement of the do-

main co-chairs and other reviewers in conducting quality analyses and evaluations of evidence to support selected interventions; (b) formal and informal researcher and practitioner dialogues aimed at refining and improving the *Procedural and Coding Manual*; (c) practitioner support networks for implementing, testing, and evaluating evidence-based interventions; (d) faculty and graduate student preparation on EBIs, including interdisciplinary courses across psychology departments; (e) recruitment of minority representation and outline strategies for assuring a multicultural perspective in reviewing studies and conducting future investigations; (f) procedures for formal recognition and dissemination of research that demonstrates evidence on school-based interventions; and (g) commitment of personal, professional, and financial resources to establish the evidence base of interventions and facilitate their use into professional practice. The resolution to commit to the scientist-practitioner basis inherent in the manual will be essential to meet this far-reaching agenda. If attained, the EBI agenda set forth by the Task Force will contribute to the data-based, problem-solving approach to school psychology services that continues to be the hallmark of the profession.

## REFERENCES

- Achenbach, T. M., & McConaughy, S. H. (1996). Relations between DSM-IV and empirically based assessment. *School Psychology Review, 25*, 329-341.
- Albee, G. W. (2000). The Boulder Model's fatal flaw. *American Psychologist, 55*, 247-248.
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychological Association. (1993). Guidelines for providers of psychological services to ethnic, linguistic and culturally diverse populations. *American Psychologist, 48*, 45-48.
- American Psychological Association. (2000). *Template Implementation Work Group of the Board of Professional Affairs, Board of Scientific Affairs, & Committee for the Advancement of Professional Psychology. Criteria for evaluating treatment guidelines*. Washington, DC: Author.
- Beutler, L. E. (2000). David and Goliath: When empirical and clinical standards of practice meet. *American Psychologist, 55*, 997-1007.
- Bickman, L. (1987). The functions of program theory. In L. Bickman (Ed.), *Using program theory in evaluation* (New Directions for Program Evaluation No. 33) (pp. 5-18). San Francisco: Jossey-Bass.
- Bickman, L. (1996). A continuum of care: More is not always better. *American Psychologist, 51*, 689-701.
- Braden, J. P., & Kratochwill, T. R. (1997). Treatment utility of assessment: Myths and realities. *School Psychology Review, 26*, 475-485.
- Brestan, E. V., & Eyberg, S. M. (1998). Effective psychosocial treatments of conduct-disordered children and adolescents: 29 years, 82 studies, and 5,272 kids. *Journal of Clinical Child Psychology, 27*, 180-189.
- Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R.

- Kratochwill and J. R. Levin (Eds.). Single case research design and analysis (pp. 187–212). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cantor, D. W. (1999). Ensuring the future of professional psychology. *American Psychologist*, *54*, 922–930.
- Carr, A. (Ed.). (2000). *What works for whom with children and adolescents? A critical review of psychological interventions with children, adolescents, and their families*. Philadelphia, PA: Brunner–Routledge.
- Caspar, F. (2001). Introduction to the issue of therapist inner processes and their training. In F. Caspar (Ed.), *The inner processes of psychotherapists: Innovations in clinical training*. New York: Oxford University Press.
- Chambless, D. L., Baker, M. J., Baucom, D. H., Beutler, L., Calhoun, K. S., Crits–Christoph, P., et al. (1998). Update on empirically validated therapies, II. *Clinical Psychologist*, *51*, 3–16.
- Chambless, D., L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*, 7–18.
- Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, *52*, 685–716.
- Chambless, D. L., Sanderson, W. C., Shoham, V., Bennett–Johnson, S., Pope, K. S., Crits–Christoph, P., Baker, M., Johnson, B., Woody, S. R., Sue, S., Beutler, L., Williams, D. A., & McCurry, S. (1996). An update on empirically validated therapies. *Clinical Psychologist*, *49*, 5–18.
- Christenson, S. L., & Sheridan, S. M. (2001). *Schools and families: Creating essential connections for learning*. New York: Guilford.
- Christophersen, E. R., & Mortweet, S. L. (2001). *Treatments that work with children: Empirically supported strategies for managing childhood problems*. Washington, DC: American Psychological Association.
- Conduct Problems Prevention Research Group. (1999a). Initial impact of the Fast Track prevention trial for conduct problems: I. The high–risk sample. *Journal of Consulting and Clinical Psychology*, *67*, 631–647.
- Conduct Problems Prevention Research Group. (1999b). Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology*, *67*, 648–657.
- Cone, J. D. (1997). Issues in functional analysis in behavioral assessment. *Behavioral Research & Therapy*, *35*, 259–275.
- Cooper, H. M. (1998). *Integrating research: A guide for literature reviews*. (3rd ed.). Beverly Hills, CA: Sage.
- Crits–Christoph, P., Chambless, D. L., Frank, E., & Brody, C. (1995). Training in empirically validated treatments: What are clinical psychology students learning? *Professional Psychology: Research and Practice*, *26*, 514–522.
- DeAngelis, T. (2000). School psychologists: In demand and expanding their reach. *Monitor on Psychology*, *31*, 30–32.
- Dobson, K. S., & Craig, K. D. (Eds.). (1998). *Empirically supported therapies: Best practice in professional psychology*. Thousand Oaks, CA: Sage.
- Doyle, A. B. (1998). Are empirically validated treatments valid for culturally diverse populations? In K. S. Dobson and Kenneth D. Craig (Eds.). *Empirically supported therapies: Best practice in professional psychology* (pp. 93–103). Thousand Oaks, CA: SAGE Publications.
- Drotor, D., & Lemanek, K. (2001). Steps toward a clinically relevant science of interventions in pediatric settings: Introduction to the special issue. *Journal of Pediatric Psychology*, *26*, 385–394.
- DuPaul, G. J., Eckert, T. L., & McGoey, K. E. (1997). School–based interventions for children

- with Attention-Deficit/Hyperactivity Disorder: One size does not fit all. *School Psychology Review*, 26, 369-381.
- DuPaul, G. J., & Ervin, R. A. (1996). Functional assessment of behaviors related to attention-deficit/hyperactivity disorder: Linking assessment to intervention design. *Behavior Therapy*, 27, 601-622.
- DuPaul, G. J., & Stoner, G. (1994). *ADHD in the schools: Assessment and intervention strategies*. New York: Guilford Press.
- Durlak, J. A. (1997). *Successful prevention programs for children and adolescents*. New York: Plenum.
- Elliott, S. N., Witt, J. C., Kratochwill, T. R., & Stoiber, K. C. (in press). Selecting and evaluating classroom interventions. In M. Shinn, G. Stoner, and H. Walker (Eds.). *Interventions for academic and behavior problems III: Preventive and remedial approaches*. Washington, DC: National Association of School Psychologists.
- Ervin, R. A., Radford, P. M., Bertsch, K., Piper, A. L., Ehrhardt, K. E., & Poling, A. (2001). A descriptive analysis and critique of the empirical literature on school-based functional assessment. *School Psychology Review*, 30, 193-210.
- Fine, M., Weiss, L., Powell, L. C., & Wong, L. M. (1997). *Off White: Readings on race, power, and society*. New York: Routledge.
- Garfield, S. L. (1998). The Future and the Scientist-Practitioner Split. *American Psychologist*, 11, 1231-1232.
- Goldfried, M. R., & Wolfe, B. E. (1998). Toward a more clinically valid approach to therapy research. *Journal of Consulting & Clinical Psychology*, 11, 1231-1232.
- Greenberg, M. T., Domitrovich, C., & Bumbarger, G. (2000). *Preventing mental disorders in school-age children: A review of the effectiveness of prevention programs*. Retrieved December 7, 2001, from Penn State University, Prevention Research Center for the Promotion of Human Development Web site: HYPERLINK "http://www.prevention.psu.edu"
- Gresham, F. M., Watson, T. S., & Skinner, C. H. (2001). Functional behavioral assessment: Principles, procedures, and future directions. *School Psychology Review*, 30, 156-172.
- Hart, C. (2001). *Doing a literature search: A comprehensive guide for the social sciences*. London: SAGE.
- Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. O. (1999). *The scientist-practitioner: Research and accountability in the age of managed care*. Boston: Allyn & Bacon.
- Hughes, J. N. (2000). The essential role of theory in the science of treating children: Beyond empirically supported treatments. *Journal of School Psychology*, 38, 301-330.
- Institute of Medicine. (1994). *Reducing risks for mental disorders: Frontiers for preventive intervention research*. Washington DC: National Academy Press.
- Kazdin, A. E. (1981). Drawing valid inferences from case studies. *Journal of Consulting and Clinical Psychology*, 49, 183-192.
- Kazdin, A. E. (1998). *Research design in clinical psychology* (3rd ed.). Boston: Allyn & Bacon.
- Kazdin, A. E. (1999). Overview of research design issues in clinical psychology. In P. C. Kendall, J. N. Butcher, & G. N. Holmbeck (Eds.), *Handbook of research methods in clinical psychology* (pp. 3-30). New York: Wiley.
- Kazdin, A. E. (2001). Bridging the enormous gaps of theory with therapy research and practice. *Journal of Clinical Child Psychology*, 30, 59-66.
- Kazdin, A. E., Kratochwill, T. R., & VandenBos, G. (1986). Beyond clinical trials: Generalizing from research to practice. *Professional Psychology: Research and Practice*, 3, 391-398.
- Keith, T. Z. (2000). Research in school psychology: What can the future hold? *School Psychology Review*, 29, 604-605.
- Kelley, M., & McCain, A. P. (1995). Promoting academic performance in inattentive chil-

- dren: The relative efficacy of school-home notes with and without response cost. *Behavior Modification*, 19, 357-375.
- Kendall, P. C. (Ed.). (2000). *Child & adolescent therapy: Cognitive-behavioral procedures*. New York: Guilford Press.
- Kratochwill, T. R. (1985). Case study research in school psychology. *School Psychology Review*, 14, 204-215.
- Kratochwill, T. R., & Levin, J. R. (1992). *Single-case research design and analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kratochwill, T. R., & McGivern, J. E. (1996). Clinical diagnosis, behavioral assessment, and functional analysis: Examining the connection between assessment and intervention. *School Psychology Review*, 25, 342-355.
- Kratochwill, T. R., & Pittman, P. (in press). Defining constructs in consultation: An important training agenda. *Journal of Educational and Psychological Consultation*.
- Kratochwill, T. R., & Plunge, M. (1992). DSM-III-R, treatment validity, and functional analysis: Further considerations for school psychologists. *School Psychology Quarterly*, 7, 227-232.
- Kratochwill, T. R., & Stoiber, K. C. (2000a). Diversifying theory and science: Expanding boundaries of empirically supported interventions in schools. *Journal of School Psychology*, 38, 349-358.
- Kratochwill, T. R., & Stoiber, K. C. (2000b). Empirically supported intervention programs: Announcing a new standing section of *School Psychology Quarterly*. *School Psychology Quarterly*, 15, 69-74.
- Kratochwill, T. R., & Stoiber, K. C. (2000c). Empirically supported interventions and school psychology: Conceptual and practical issues: Part II. *School Psychology Quarterly*, 15, 233-253.
- Kratochwill, T. R., & Stoiber, K. C. (2000d). Uncovering critical research agendas for school psychology: Conceptual dimensions and future directions. *School Psychology Review*, 29, 591-603.
- Kratochwill, T. R., Stoiber, K. C., & Gutkin, T. B. (2001). Empirically supported interventions in school psychology: The role of negative results in outcome research. *Psychology in the Schools*.
- Levin, J. R., & O'Donnell, A. M. (1999). What to do about educational research's credibility gaps? *Issues in Education; Contributions from Educational Psychology*, 5, 177-229
- Levin, J. R., O'Donnell, A. M., & Kratochwill, T. R. (in press). A case for enhancing the credibility of educational psychological intervention research. In I. Weiner (Ed.), *Comprehensive handbook of psychology* (pp. ). New York: Wiley.
- Lewis-Snyder, G., Stoiber, K. C., Kratochwill, T. R. (2002). Application of the school psychology task force criteria to group-based intervention research: A case illustration. *School Psychology Quarterly*
- Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.
- Lin, K. M., Poland, R., & Nakasaki, G. (Eds.) (1993). *Psychopharmacology and the psychobiology of ethnicity*. Washington DC: American Psychiatric Press.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Lipsey, M. W., & Cordray, D. S. (2000). Evaluation methods for social intervention. *Annual Review of Psychology*, 51, 345-375.
- Lonigan, C. J., Elbert, J. C., & Bennett-Johnson, S. (1998). Empirically supported psychosocial interventions for children: An overview. *Journal of Clinical Child Psychology*, 27, 138-145.
- McComas, J. J., & Mace, F. C. (2000). Theory and practice in conducting functional analysis.

- In E. S. Shapiro & T. R. Kratochwill (Eds.), *Behavioral assessment in schools: Theory, research, and clinical foundations* (2nd ed.) (pp. 78–103). New York: Guilford Press.
- McDonald, L., Billingham, S., Conrad, T., Morgan, A., Payton, N. O., & Payton, E. (1997). Families and schools together (FAST): Integrating community development with clinical strategies. *Families in Society, 78*, 140–155.
- McDonald, L., & Sayger, T. V. (1998). Impact of a family and school-based prevention program on protective factors for high risk youth. *Drugs & Society, 12*, 61–85.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.
- Minke, K. M., & Bear, G. C. (2000). *Preventing school problems – promoting school success*. Bethesda, MD: National Association of School Psychologists.
- Morgan, D. L., & Morgan, R. K. (2001). Single-participant research design: Bringing science to managed care. *American Psychologist, 56*, 119–127.
- Morris, R. J., & Kratochwill, T.R. (Eds.). (1998). *The practice of child therapy* (3rd ed.). Needham Heights, MA: Allyn & Bacon.
- Najavits, L. M., Weiss, R. D., Shaw, S. R., & Dierberger, A. E. (2000). Psychotherapists' view of treatment manuals. *Professional Psychology: Research and Practice, 31*, 404–408.
- Nastasi, B. (2000). School psychologists as health-care providers in the 21st Century: Conceptual framework, professional identity, and professional practice. *School Psychology Review, 29*, 540–554.
- Nastasi, B. K., & Berg, M. J. (1999). Using ethnography to strengthen and evaluate intervention programs. In J. J. Schensul, M. D. LeCompte, G. A. Hess, Jr., B. K.K. Nastasi, M. J. Berg, L. Williamson, et al. (Eds.) *Using ethnographic data: Interventions, public programming, and public policy* (pp. 1–49). Walnut Creek, CA: AltaMira Press.
- Nathan, P. E., & Gorman, J. M. (Eds.). (1998). *A guide to treatments that work*. New York: Oxford University Press.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: U.S. Department of Health & Human Services, National Institute of Child Health & Human Development.
- Oakland, T. D., & Cunningham, J. (1999). The futures of school psychology: Conceptual models for its development and examples of their applications. In C. R. Reynolds & T. B. Gutkin (Eds.), *Handbook of school psychology* (3rd ed., pp. 34–55). New York: Wiley.
- Ollendick, T. H., & King, N. J. (1998). Empirically supported treatments for children with phobic and anxiety disorders. *Journal of Clinical Child Psychology, 27*, 156–167.
- O'Neill, R., Horner, R., Albin, R., Storey, K., & Sprague, J. (1997). *Functional assessment and program development for behavior problems*. Pacific Grove, CA: Brooks/Cole.
- Prochaska, J. O. (1984). *Systems of psychotherapy: A transtheoretical analysis* (2nd ed.). Homewood, IL: Dorsey Press.
- Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist, 47*, 1102–1114.
- Ramirez, S. Z., Lepage, K. M., Kratochwill, T. R., & Duffy, J. L. (1998). Multicultural issues in school-based consultation: Conceptual and research considerations. *Journal of School Psychology, 36*, 479–509.
- Rappaport, M. D., Chung, K. M., Shore, G., & Isaacs, P. (2001). A conceptual model of child psychopathology: Implications for understanding attention deficit hyperactivity disorder and treatment efficacy. *Journal of Clinical Child Psychology, 30*, 48–58.
- Reppucci, N. D., Woolard, J. L., & Fried, C. S. (1999). Social, community, and preventive interventions. *Annual Review of Psychology, 50*, 387–418.

- Reynolds, A. J. (1998). Confirmatory program evaluation: A method for strengthening causal inference. *American Journal of Evaluation, 19*, 203–221.
- Rice, C. E. (1997). Scenarios: The scientist–practitioner split and the future of psychology. *American Psychologist, 52*, 1173–1181.
- Roth, A. D., & Fonagy, P. (1996). *What works for whom? A critical review of psychotherapy research*. New York: Guilford.
- Schensul, J. J., & LeCompte, M. D. (Eds.). (1999). *Ethnographer's toolkit* (Vols. 1–7). Walnut Creek, CA: AltaMira Press.
- Schill, M. T., Kratochwill, T. R., & Gardner, W. I. (1996). Conducting a functional analysis of behavior. In M. J. Breen & C. R. Fiedler (Eds.), *Behavioral approach to assessment of youth with emotional/behavioral disorders: A handbook for school-based practitioners* (pp. 83–180). Austin, TX: Pro-Ed.
- Schneider, S. F. (1990). Psychology at a crossroads. *American Psychologist, 45*, 521–529.
- Shernoff, E. S., Kratochwill, T. R., & Stoiber, K. C. (2001, August). *Training in evidence-based interventions: What are school psychology programs teaching?* Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA.
- Shernoff, E. S., Kratochwill, T. R., & Stoiber, K. C. (2002). Evidence-based interventions in school psychology: An illustration of Task Force coding criteria using single-participant research design. *School Psychology Quarterly, 15*, 75–115.
- Shriver, M. D., Anderson, C. M., & Proctor, B. (2001). Evaluating the validity of functional behavior assessment. *School Psychology Review, 30*, 180–192.
- Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*, 5–11.
- Soldz, S., & McCullough, L. (Eds.). (2000). *Reconciling empirical knowledge and clinical experience: The art and science of psychotherapy*. Washington, DC: American Psychological Association.
- Spirito, A. (Ed.). (1999). Empirically supported treatments in pediatric psychology. *Journal of Pediatric Psychology, 24*, 87–174.
- Stoiber, K. C., & Kratochwill, T. R. (1998). *Handbook of group intervention for children and families*. Boston: Allyn & Bacon.
- Stoiber, K. C., & Kratochwill, T. R. (2000). Empirically supported interventions in schools: Rationale and methodological issues: Part I. *School Psychology Quarterly, 15*, 75–115.
- Stoiber, K. C., & Kratochwill, T. R. (2001). Evidence-based intervention programs: Rethinking, refining, and renaming the new standing section of *School Psychology Quarterly*. *School Psychology Quarterly, 16*, 1–8.
- Stoiber, K. C., & Kratochwill, T. R. (2002). *Outcomes: Planning, monitoring, evaluating*. San Antonio, TX: The Psychological Corporation.
- Stoiber, K. C., & Kratochwill, T. R. (in press). *Functional assessment and intervention system*. San Antonio, TX: The Psychological Corporation.
- Stoiber, K. C., & Reinemann, D. H. S. (2001). All interventions were not created equal: The case against an anything goes approach to psychotherapy. *School Psychology Quarterly, 16*, 220–238.
- Stoiber, K. C., & Waas, G. (in press). A contextual and methodological perspective on the evidence-based movement within school psychology in the United States. *Educational and Child Psychology, 19*(3).
- Sue, D. W., Bingham, R. P., Porche-Burke, L., & Vasquez, M. (1999). The diversification of psychology: A multicultural revolution. *American Psychologist, 54*, 1061–1069.
- Sue, D. W., & Sue, D. (1999). *Counseling the culturally different*. New York: Wiley.
- Sue, S. (1999). Science, ethnicity, and bias: Where have we gone wrong? *American Psychologist, 54*, 1070–1077.

- Tharp, R. (1991). Cultural diversity and the treatment of children. *Journal of Consulting and Clinical Psychology, 59*, 799–812.
- Troia, G. A. (1999). Phonological awareness intervention research: A critical review of the experimental methodology. *Reading Research Quarterly, 34*, 28–52.
- Waehler, C. A., Kalodner, C. R., Wampold, B. E., & Lichtenberg, J. W. (2000). Empirically supported treatment (ESTs) in perspective: Implications for counseling psychology training. *Counseling Psychologist, 28*, 657–671.
- Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, NJ: Erlbaum.
- Watson, T. S., Gresham, F. M., & Skinner, C. H. (2001). Introduction to the mini-series: Issues and procedures for implementing functional behavior assessments in schools. *School Psychology Review, 30*, 153–155.
- Webster-Stratton, C. (1996). Early intervention with videotape modeling: Programs for families of children with oppositional defiant disorder or conduct disorder. In E. D. Hibbs & P. S. Jensen (Eds.), *Psychosocial treatments for child and adolescent disorders: Empirically-based strategies for clinical practice* (pp. 435–474). Washington, DC: American Psychological Association.
- Weisz, J. R. (1998). Empirically supported treatments for children and adolescents: Efficacy, problems, and prospects. In K. S. Dobson and K. D. Craig (Eds.) *Empirically supported therapies: Best practice in professional psychology* (pp. 66–92). Thousand Oaks, CA: SAGE.
- Weisz, J. R. (2000). Lab-clinic differences and what we can do about them: II. Linking research and practice to enhance our public impact. *Clinical Child Psychology Newsletter, 15*, 1–5, 9.
- Weisz, J. R., & Hawley, K. M. (20012). *Procedural and coding manual for identification of beneficial treatments* (Draft #4). Washington, DC: American Psychological Association, Society for Clinical Psychology Division 12 Committee on Science and Practice.
- Wilson, G. T. (1996a). Empirically validated treatments: Reality and resistance. *Clinical Psychology: Science and Practice, 3*, 241–244.
- Wilson, G. T. (1996b). Manual-based treatments: The clinical application of research findings. *Behaviour Research and Therapy, 34*, 295–314.
- Zins, J. E., Elias, M. J., Greenberg, M. T., & Pruetz, M. K. (2000). Promoting quality implementation in prevention programs. *Journal of Educational and Psychological Consultation, 11*, 173–174.