Conceptualizing Changes in Behavior in Intervention Research: The Range of Possible Changes Model

Andres De Los Reyes
Yale University

Alan E. Kazdin
Yale University School of Medicine

An international movement has focused on identifying evidence-based interventions that were developed to change psychological constructs and that are supported by controlled studies. However, inconsistent findings within individual intervention studies and among multiple studies raise critical problems in interpreting the evidence, and deciding when and whether an intervention is evidence-based. A theoretical and methodological framework (Range of Possible Changes [RPC] Model) is proposed to guide the study of change in intervention research. The authors recommend that future quantitative reviews of the research literature use the RPC Model to conceptualize, examine, and classify the available evidence for interventions. Future research should adopt the RPC Model to both develop theory-driven hypotheses and conduct examinations of the instances in which interventions may or may not change psychological constructs.

Keywords: efficacy, effectiveness, range of possible changes, intervention, treatment

The terms *efficacy* and *effectiveness* denote the study of whether interventions can successfully change specific psychological constructs or behaviors (e.g., Hoagwood, Hibbs, Brent, & Jensen, 1995; Kazdin, 2000; Lambert & Ogles, 2004).[1] An overriding goal of intervention research is to identify evidence-based interventions: Evidence from controlled experiments suggests these interventions change constructs they were developed to change. Researchers have developed classification systems through which a given intervention can be identified as evidence-based, based on prior well-controlled experimental outcome studies examining the intervention (e.g., Lonigan, Elbert, & Johnson, 1998; Nathan & Gorman, 2002; Roth & Fonagy, 2005). However, the supportive evidence raises fundamental questions that relate directly to how changes in psychological constructs ought to be conceptualized, examined, and classified in intervention research.

The purpose of this article is to identify a key issue in interpreting the evidence as support for interventions, and propose a theoretical and methodological framework to guide the study of change in intervention research. First, we briefly discuss existing systems that identify evidence-based interventions and suggest that an additional consideration be made when deciphering whether interventions change constructs. Second, we propose a theoretical and methodological framework (Range of Possible Changes [RPC] Model) to conceptualize and examine change both within and between intervention studies. From this framework, we extrapolate categories to guide the classification of intervention change within and between studies. Third, we illustrate the framework using findings from prior work, and discuss key distinctions between it and other approaches commonly used to address inconsistent findings in intervention research. Lastly, we discuss the conceptual and research implications of the framework for future intervention research, as well as the general implications of the framework for both the basic and applied psychological sciences.

Existing Classification Systems for Gauging the Evidence for Interventions

Illustrations of Current Criteria

Researchers have developed systems to identify interventions as evidence-based, based on prior well-controlled experiments. Several systems have established criteria to identify interventions that change constructs, and have been summarized and contrasted

[1] The intervention literature often distinguishes between the terms *efficacy* and *effectiveness* by the setting in which changes in constructs occur. Specifically, efficacy research examines whether an intervention can change psychological constructs under controlled experimental conditions. In contrast, effectiveness research examines whether an intervention can change psychological constructs in more naturalistic settings outside the research laboratory. Despite these differences in settings, both of these research literatures examine the ability of interventions to change multidimensional constructs. Thus, the conceptual and methodological framework we propose generally applies to the study of change in multidimensional constructs in intervention research. At the same time, the evidence we cite in support of our proposed framework is largely gleaned from the efficacy literature, particularly because of its longer history of examination and replication, relative to the effectiveness literature. Thus, it is possible to examine within- and between-study inconsistencies in evidence. Further, existing systems that identify evidence-based interventions base these identifications on the outcomes of efficacy studies. Thus, use of efficacy research to illustrate our proposed framework allows for a comparison between the proposed framework and existing classification systems.

elsewhere (e.g., Lonigan et al., 1998; Nathan & Gorman, 2002; Roth & Fonagy, 2005). These classification systems are key components of an international movement in the mental health fields to identify evidence-based interventions (e.g., Kazdin, 2000). Although the systems differ in several respects, they generally identify interventions via reviews of the empirical literature. The reviews stipulate that studies providing evidence for an intervention share several characteristics: (a) random assignment of subjects to intervention and control or comparison conditions; (b) careful specification of the population undergoing intervention; (c) use of a manual detailing the intervention; (d) multiple outcome measures (raters, if employed, are naïve to conditions); (e) statistically significant differences between intervention and comparison groups; and (f) replication of findings supporting the intervention, especially by independent investigators.

The systems often create separate categories for interventions for which well-established experimental evidence for change exists, categories for interventions for which some experimental evidence for change exists, and categories for interventions for which limited or no experimental evidence for change exists. Moreover, the systems often consider one study, or even a number of studies by the same researcher or research team as providing insufficient support for an intervention. However, the systems differ somewhat in the categories they establish, or the instances in which an intervention may be placed in a category (e.g., number of studies needed, whether one or more studies need to be conducted by independent researchers).

Although several systems for identifying evidence-based interventions are available, it is useful to highlight briefly one of the more often utilized systems (Task Force on Promotion and Dissemination of Psychological Procedures [TFPP], 1995; Chambless et al., 1998; Chambless & Hollon, 1998; Chambless & Ollendick, 2001). The criteria for this system call for identification of an intervention that has well-established efficacy if, for instance, evidence supporting the intervention exists from between-groups design experiments conducted by at least two different investigators or teams. In these experiments, the intervention must be superior to either a placebo-control condition or another intervention, or must be equivalent to an already established intervention (experiments comparing the interventions must do so with adequate sample sizes).

An intervention is considered probably efficacious if evidence for the intervention's efficacy exists from at least two between-groups design experiments that show that the intervention is superior to a wait-list control condition. Otherwise, an intervention is considered probably efficacious if one or more experiments meet criteria for well-established efficacy, but these experiments were conducted by the same investigator or research team. Lastly, an intervention is considered experimental if the intervention has not yet been tested within an experiment that meets the TFPP's criteria for methodology. Furthermore, if inconsistencies arise among multiple efficacy studies of the same intervention, then the preponderance of the evidence among the studies must be consistent with the efficacy of the intervention for that intervention to be placed within an efficacy category. Thus, the TFPP's criteria employ rank-ordered categorizations of interventions. Further, the categorizations are based on whether well-controlled studies exist to support the intervention, the number of such studies, and whether these studies have been conducted by more than one independent investigator or team.

As mentioned previously, other efforts to delineate evidence-based interventions have been developed (e.g., Lonigan et al., 1998; Nathan & Gorman, 2002; Roth & Fonagy, 2005), and have followed the pioneering efforts of TFPP. The criteria employed across systems may differ. However, the key goal of all of these systems is the same: to accumulate knowledge to reach conclusions on the ability of interventions to change specific psychological constructs.

Overview of the Need for a Conceptual Framework

Critical Issues

Systems developed to identify evidence-based interventions have significantly advanced our understanding of how to determine whether scientific knowledge exists to suggest a given intervention can bring about change. However, one fundamental issue needs to be addressed: How consistent must the evidence within a study be to suggest that the intervention it examined changes the construct it was developed to change? In light of this critical issue, we propose a theoretical and methodological framework that addresses key definitional, theoretical, methodological, and practical issues surrounding the study of change in intervention research.

Construct validity and the "test" of interventions. Recent theoretical attention to construct validity may inform the conceptualization and examination of change in intervention research. For instance, a recent conceptualization of construct validity highlights the two key factors to consider when gauging whether a test has construct validity: (a) the construct it aims to measure exists; and (b) variations in measurement outcomes are causally produced by variations in the construct (Borsboom, Mellenbergh, & van Heerden, 2004). Interestingly, the primary aims of intervention research correspond perfectly with these two factors.

Any intervention study is designed to address two questions: (a) Is this specific intervention, targeting this specific construct, capable of changing the construct? and (b) If this intervention can change the construct, does variation in the administration of the intervention causally produce changes in measures of outcome that purportedly assess intervention change? These two questions are often addressed by randomly assigning participants to either the intervention being examined or a control condition. Thus, by assigning participants to either the intervention or a control condition, a researcher varies the administration of the intervention dichotomously, so that for each participant, the intervention is either present or not present. Measures of outcome are administered to all participants, and change is examined by comparing scores on outcome measures between the two conditions. If participants' scores in the intervention condition "outperform" participants' scores in the control condition, a researcher is usually confident in concluding that the study suggests the intervention changes the construct.

Although intervention studies seem to address key issues related to testing interventions, current systems for gauging the evidence for interventions do not address key concerns when conceptualizing change. For instance, it is unclear what conceptual and empirical conditions need to be met for a researcher to conclude that they found evidence for an intervention's ability to change the targeted construct. This concern becomes difficult to address,

given the multidimensional natures and conceptualizations of constructs targeted for change.[2]

Multidimensional nature of intervention research. The difficulty in conducting intervention research can be highlighted by briefly discussing the conceptualization and examination of intervention outcomes. Perhaps the most critical issues arise because psychological constructs often targeted for intervention are quite difficult to conceptualize and measure. For instance, the psychological constructs most often targeted for intervention are specific psychological problems or disorders, presumably because such constructs often have quite negative social, emotional, and physical impacts on people experiencing them (e.g., Lambert & Ogles, 2004). Thus, examining interventions that target specific disorders are of utmost importance to researchers studying change in constructs in psychological science.

Clinical disorders (e.g., anorexia nervosa, obsessive–compulsive disorder, major depressive disorder, specific phobias, panic disorder, posttraumatic stress disorder) are often related to, influenced by, or maintained by a host of associative, risk, and causal factors (e.g., multiple behavioral, biological, cognitive, developmental, genetic, neurological, and social factors; see American Psychiatric Association, 1994; Bouton, Mineka, & Barlow, 2001; Brewin, Dalgleish, & Joseph, 1996; Guisinger, 2003; Monroe & Harkness, 2005; Öhman & Mineka, 2001; Szechtman & Woody, 2004). Understandably, the psychological constructs targeted for intervention are very complex, multifaceted, and probably difficult to change.

The difficulty in conceptualizing and measuring psychopathology in intervention research often requires relying on different perspectives or sources to gather information (e.g., self, spouse or significant other, clinician, laboratory observer, biological, institutional records, and parents, teachers, and peers, in the case of children). Furthermore, measures taken from these different sources often do not correspond (e.g., Achenbach, Krukowski, Dumenci, & Ivanova, 2005; Achenbach, McConaughy, & Howell, 1987; De Los Reyes & Kazdin, 2004, 2005). That these measures often rely on various assessment methods (e.g., rating scales, structured interviews, behavioral codings, physiological measurements) and are evaluated with diverse statistical analyses (e.g., mean differences between intervention and control conditions, diagnostic status, clinically significant change) add to the complexity of intervention research (Kazdin, 2003; Kendall, Butcher, & Holmbeck, 1999). Indeed, the complexity of psychological constructs perhaps has been a contributing factor to a longstanding concern in intervention research. Specifically, whether an investigation finds evidence to suggest an intervention changes a construct largely may be driven by both the methodology one employs to address the question (e.g., Frank et al., 1991; Kazdin & Wilson, 1978; Prien, Carpenter, & Kupfer, 1991), and the source or informant providing the information to address the question (e.g., Strupp, 1996; Strupp & Hadley, 1977).

Intervention research is conducted with an understanding that the psychological constructs often targeted for intervention must be conceptualized as multidimensional. The nature and extent of these constructs are often influenced by a number of factors, and thus, require they be assessed via multiple perspectives and measurement methodologies. Information gathered from these multiple perspectives and methodologies capture the nature and extent of these constructs in fundamentally different ways. Therefore, it is difficult to argue that the psychological constructs targeted for

intervention can be captured with a single primary measure. Further, it is difficult to argue that any individual reliable and valid measure of these psychological constructs can be relied on, over and above other reliable and valid measures of these same constructs.

In sum, psychological constructs targeted for intervention are multidimensional. This suggests that the nature of change attributable to intervention is multidimensional as well. Multiple measurement methods are used to gather information from multiple sources. After all of this information is gathered, it must be examined via multiple statistical analyses. Unfortunately, these conceptualizations and measurements often result in inconsistencies. Inconsistent evidence highlights a greater problem in conceptualizing change in intervention research.

Within-Study Inconsistencies

Given the potential for inconsistent findings, we pose a number of questions related to interpreting the findings of intervention studies. For instance, in a study of an intervention targeting childhood depression, would the evidence be considered support for change if 1 of 10 outcome measures of the targeted construct suggests change, or 5 of 10, or 8 of 10? Would a study support the intervention if measures completed by children's parents suggest change, but not measures completed by either teachers or the children themselves? Perhaps an investigation may support the intervention with analyses of mean differences between intervention and control conditions, but not diagnostic status? These questions all relate to one fundamental question: When does the evidence suggest intervention change, if multiple measures or outcomes reflect differences between conditions in some instances, but not others? To address this question, it may be necessary to assume that change within interventions is often variable.

Interventions targeting psychological constructs in children. Considerations of within-study inconsistencies are critical when determining whether evidence within a study suggests change. This is illustrated in research examining a well-developed intervention for childhood anxiety (Coping Cat; see Weisz, 2004). In the first efficacy study examining an individual cognitive–behavioral treatment for childhood anxiety disorders (Kendall, 1994), individual cognitive–behavioral treatment was compared to a wait-list control condition on outcome measures administered to parents, children, and teachers, as well as independent assessor ratings of behavioral observations. Methods of analyzing change included examinations between conditions using analyses of mean differences on outcome measures, as well as tests of diagnostic status and clinically significant change (e.g., normative comparisons; Kendall & Grove, 1988). No pretreatment outcome measure differences were evident between conditions.

---

[2] Many of the issues raised in this article have implications for conceptualizing the ability of any experimental paradigm to change a multidimensional construct. This being said, many of the issues raised in this article are raised within the context of interventions for psychopathology. This is because this area of psychology is primarily an applied science, and therefore, an area for which changing behavior (e.g., changes in thoughts, attributions, feelings, actions) is of primary importance. At the same time, where appropriate, we will discuss general implications that this article may have for experimental paradigms that change constructs in psychological science.

As can be seen in Table 1, tests of mean differences between intervention and control conditions were highly inconsistent. These inconsistencies were found on measures of the construct targeted for intervention, as well as across measures of broadband internalizing (e.g., anxiety, depression) and externalizing (e.g., hyperactivity, oppositionality, aggression) constructs. Specifically, differences were found on less than half of all of the measures examining the construct targeted for intervention. Differences on measures of the target construct were mainly found on child self-rated questionnaires. Differences on measures of the target

construct were found on one measure completed by parents. Of the eight independent assessor-completed behavioral codes examined in the study, differences were found only on the total score. Indeed, the most consistent results were on measures examining comorbid constructs (depression) and associated features of the target construct (negative affectivity, coping style, health problems), all found on child self-rated and parent-rated questionnaires.

Only parent and child were relied on to measure diagnostic status, and statistical tests comparing intervention and control conditions on diagnostic status were not reported (see Table 1).

Table 1

Intervention Outcomes for Kendall (1994): Reported Statistical Differences Between Individual Cognitive–Behavioral Treatment and Control Conditions

Informant	Measure	Method		
		Mean differences	Diagnostic status	Clinically significant change
Child	ADIS-C	—	—	—
Child	RCMAS[a]	Intervention × time	—	Intervention > control
Child	STAIC A-Trait[a]	Intervention × time	—	—
Child	STAIC A-State[a]	Intervention × time	—	—
Child	FSSC-R[a]	Intervention × time	—	—
Child	CQ-C[b]	Intervention × time	—	—
Child	CDI[b]	Intervention × time	—	Intervention > control
Child	NASSQ[b]	Intervention × time	—	—
Parent	ADIS-P	—	—	—
Parent	CBCL-Internalizing[c]	Intervention × time	—	—
Parent	CBCL-Social[b]	Intervention × time	—	—
Parent	CBCL-Health[b]	Intervention × time	—	—
Parent	CBCL-Externalizing[d]	Intervention × time	—	—
Parent	STAIC A-Trait-P[a]	Intervention × time	—	—
Teacher	TRF-Internalizing[c]	No differences	—	—
Teacher	TRF-Externalizing[d]	No differences	—	—
Independent assessor	Gratuitous Verbalizations[a]	No differences	—	—
Independent assessor	Gratuitous Body Movements[a]	No differences	—	—
Independent assessor	Trembling Voice[a]	No differences	—	—
Independent assessor	Avoiding Task[a]	No differences	—	—
Independent assessor	Absence of Eye Contact[a]	No differences	—	—
Independent assessor	Fingers in Mouth[a]	No differences	—	—
Independent assessor	Body Rigidity[a]	No differences	—	—
Independent assessor	Total Behavioral Codes[a]	Intervention × time	—	—
Statistically significant differences reported: Overall		13 of 22 Measures	N/A	2 of 2 Measures
Statistically significant differences reported: Target construct[a]		6 of 13 Measures	N/A	1 of 1 Measure
Statistically significant differences reported: Comorbid constructs/ associated features[b]		5 of 5 Measures	N/A	1 of 1 Measure
Statistically significant differences reported: Broadband internalizing constructs[c]		1 of 2 Measures	N/A	N/A
Statistically significant differences reported: Broadband externalizing constructs[d]		1 of 2 Measures	N/A	N/A

Note. ADIS-C/P = Anxiety Disorders Interview Schedule for Children: Parent and Child Versions (Silverman, 1987); RCMAS = Revised Children's Manifest Anxiety Scales (Reynolds & Richmond, 1978); STAIC = State-Trait Anxiety Inventory for Children (Spielberger, 1973); FSSC-R = Fear Survey Schedule for Children-Revised (Ollendick, 1983); CQ-C = Coping Questionnaire for Children (Kendall, 1994); CDI = Children's Depression Inventory (Kovacs, 1981); NASSQ = Children's Negative Affectivity Self-Statement Questionnaire (Ronan, Kendall, & Rowe, 1994); CBCL = Children's Behavior Checklist (Achenbach & Edelbrock, 1983); STAIC A-Trait-P = State-Trait Anxiety Inventory for Children-Parent Form (Strauss, 1987); TRF = Teacher Report Form (Achenbach & Edelbrock, 1983); — = Statistical tests not reported in the study; Intervention × time = Tests of mean differences showing a significant intervention × time interaction effect; Intervention > control = Tests of clinically significant change showing a significant difference between intervention and control in proportions of participants experiencing change. As in the original study, results summarized only reflect statistical analyses reported in the study denoting differences between individual cognitive–behavioral treatment and control conditions.
[a] Measures assessing target construct.   [b] Measures assessing comorbid constructs/associated features.   [c] Measures assessing broadband internalizing constructs.   [d] Measures assessing broadband externalizing constructs.

Moreover, parent, teacher, and child measures were relied on to measure clinically significant change. However, statistical tests comparing intervention and control conditions were only conducted on two child self-rated measures, each showing significant differences between conditions (as an aside, descriptive data of measures of clinically significant change using parent and teacher ratings were reported, but the data reported did not include all children in the conditions). Furthermore, of the two child self-rated measures of clinically significant change, both were questionnaire measures, and only one of the measures represented the construct targeted for intervention. Thus, on further review of this investigation, one conclusion can definitely be made: The investigation both did and did not find evidence for change. Moreover, these inconsistencies were found on measures purportedly assessing the construct targeted for intervention.

Except to illustrate the point, there is no need to single out Kendall (1994); the program of research from which this study comes is exemplary. More recent studies of psychological interventions for childhood anxiety (e.g., Barrett, Healy-Farrell, & March, 2004; Öst, Svensson, Hellström, & Lindwall, 2001), psychological interventions targeting constructs other than childhood anxiety (e.g., conduct problems; Sanders, Markie-Dadds, Tully, & Bor, 2000; Webster-Stratton & Hammond, 1997), and other forms of intervention (e.g., medication interventions for attention-deficit/ hyperactivity disorder; Pearson et al., 2003; Pliszka, Browne, Olvera, & Wynne, 2000) convey the same point. Conclusions vary across outcome measures, methods of analysis, and measures purportedly assessing the constructs targeted for intervention.

Interventions targeting psychological constructs in adults. Evidence of within-study inconsistencies is not limited to studies of interventions targeting psychological constructs in children. Inconsistencies can be found in studies of interventions targeting psychological constructs in adults as well. This is illustrated in research examining a well-developed intervention for generalized anxiety disorder (progressive muscle relaxation or relaxation; Barlow, Raffa, & Cohen, 2002; Bernstein & Borkovec, 1973). In an efficacy study examining relaxation for generalized anxiety disorder (Barlow, Rapee, & Brown, 1992), relaxation was compared to a wait-list control condition on outcome measures administered to participants, as well as independent assessor ratings. Methods of analysis included mean differences between conditions on outcome measures, as well as tests of responder status and high-endstate functioning (Himadi, Boice, & Barlow, 1986). A single preintervention difference was reported on a measure of neuroticism; no other differences were evident.

As can be seen in Table 2, tests of mean differences between conditions only showed differences on 6 of 16 measures. Half of the measures of the target construct (4 of 8) showed differences, 2 of 7 measures of comorbid constructs (depression) or associated features of the target construct (personality traits) showed differences, and a measure of general psychosocial functioning did not show differences. Differences were found only on two of the participant self-rated measures assessing the targeted construct. Differences were not found on well-researched participant self-rated questionnaire measures of the target construct. The remaining significant differences were found only on independent assessor ratings.

With regard to tests of responder status and high-endstate functioning, these tests were only conducted by comparing the control condition with all intervention participants grouped across inter-ventions examined in the study (relaxation, cognitive restructuring, combined relaxation and cognitive restructuring). Statistical tests of responder status and high-endstate functioning did not examine whether there were differences between participants in the relaxation condition and participants in the control condition. In brief, the evidence suggests that the investigation both did and did not find support for change. Moreover, inconsistent evidence was found on measures purportedly assessing the targeted psychological construct.

As with our previous example, the Barlow et al. (1992) study is noted merely as an illustration. The program of research from which this study comes is exemplary as well. More recent studies examining psychological interventions for adult anxiety disorders (e.g., Mersch, 1995; Schmidt et al., 2000), for psychological interventions targeting adult psychological constructs other than anxiety (e.g., eating disorders; Agras, Schneider, Arnow, Raeburn, & Telch, 1989; Agras et al., 1995), and other forms of intervention (e.g., medication interventions for bipolar disorder; Bowden et al., 2000; Nemeroff et al., 2001) convey the same point. Again, conclusions vary across measures of outcome, methods by which outcomes were examined, and even across measures assessing the psychological constructs targeted for intervention.

Between-Study Inconsistencies

Interventions targeting psychological constructs in children. Between-study inconsistencies also are evident among investigations of the same intervention, raising an additional issue: How much of the evidence gathered between studies examining the same intervention should suggest that the intervention changes the targeted construct? The importance of this issue is illustrated in a second well-conducted efficacy study of individual cognitive–behavioral treatment for childhood anxiety (Coping Koala; see Weisz, 2004). A research team independent of the researcher that conducted the first efficacy study summarized previously of individual cognitive–behavioral treatment for childhood anxiety disorders examined this same intervention in a subsequent efficacy study (Barrett, Dadds, & Rapee, 1996). In this study, individual cognitive–behavioral treatment was compared to a wait-list control condition on outcome measures administered to parents and children, as well as independent assessor ratings of children's threat and avoidance responses to ambiguity. Methods of analysis included mean differences between conditions on outcome measures, as well as tests of diagnostic status and clinically significant change. No preintervention outcome measure differences were reported.

Unlike the first study, tests of mean differences between individual cognitive–behavioral treatment and control conditions did not show a single significant difference on any of the measures assessing the target of the intervention (see Table 3). Indeed, differences between conditions were not found on measures of the target construct, even on measures for which differences were found in Kendall (1994) (cf. Tables 1 and 3). The only differences were on measures of broadband internalizing constructs completed by mother and father, and one measure of mother-rated broadband externalizing constructs. Moreover, for tests of mean differences examining independent assessor ratings of threat scores and avoidant solutions, significant effects were noted in the original study. However, results for these effects were not reported. Post-hoc univariate analyses indicated that, whereas the other interven-

Table 2
Intervention Outcomes for Barlow et al. (1992): Reported Statistical Differences Between Relaxation and Control Conditions

		Method		
Informant	Measure	Mean differences	Responder status	High-endstate functioning
Self	STAI-Trait[a]	No differences	—	—
Self	CSAQ-Cognitive[a]	No differences	—	—
Self	CSAQ-Somatic[a]	No differences	—	—
Self	FQ-Total[b]	No differences	—	—
Self	FQ-Anxiety/Depression[b]	No differences	—	—
Self	BDI[b]	Intervention > control	—	—
Self	EPI-Extraversion[b]	No differences	—	—
Self	EPI-Neuroticism[b]	No differences	—	—
Self	SSS[c]	No differences	—	—
Self	Average Daily Anxiety[a]	Intervention > control	—	—
Self	Average Daily Depression[b]	No differences	—	—
Self	Weekly Intense Anxiety Episodes[a]	No differences	—	—
Self	Percent of Day Worrying[a]	Intervention > control	—	—
Independent assessor	ADIS-R Clinical Severity[a]	Intervention > control	—	—
Independent assessor	Hamilton Anxiety[a]	Intervention > control	—	—
Independent assessor	Hamilton Depression[b]	Intervention > control	—	—
Self and Independent assessor	Composite Scale-TR	—	—	—
Self and Independent assessor	Composite Scale-HEF	—	—	—
Statistically significant differences reported: Overall		6 of 16 Measures	N/A	N/A
Statistically significant differences reported: Target construct[a]		4 of 8 Measures	N/A	N/A
Statistically significant differences reported: Comorbid constructs/ associated features[b]		2 of 7 Measures	N/A	N/A
Statistically significant differences reported: General psychosocial functioning[c]		0 of 1 Measure	N/A	N/A

Note. STAI-Trait = State-Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970); CSAQ = Cognitive-Somatic Anxiety Questionnaire (Schwartz, Davidson, & Goleman, 1978); FQ = Fear Questionnaire (Marks & Mathews, 1979); BDI = Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961); EPI = Eysenck Personality Inventory (Eysenck & Eysenck, 1968); SSS = Subjective Symptoms Scale (Barlow et al., 1992); ADIS-R = Anxiety Disorders Interview Schedule-Revised (Di Nardo & Barlow, 1988); Hamilton Anxiety/Depression = Hamilton Anxiety and Depression Scales (Hamilton, 1959, 1960); Composite Scale-TR = Scales combining scores from ADIS-R, Hamilton Anxiety, Average Daily Anxiety, and SSS to determine responder status; Composite Scale-HEF = Scales combining scores from ADIS-R, Hamilton Anxiety, Average Daily Anxiety, FQ, and SSS to determine those exhibiting high endstate functioning; — = Statistical tests not reported in the study; Intervention > control = Tests of mean differences showing a significant difference between intervention and control conditions. As in the original study, results summarized only reflect statistical analyses reported in the study denoting differences between relaxation and control conditions.
[a] Measures assessing target construct. [b] Measures assessing comorbid constructs/associated features. [c] Measures assessing general psychosocial functioning.

tion examined in the study besides individual cognitive–behavioral treatment had lower outcome scores, relative to individual cognitive–behavioral treatment and controls, individual cognitive–behavioral treatment was not reported to have lower scores, relative to controls. Thus, unlike the first study, this second study found that mean differences did not suggest any differences on measures of the targeted construct. Differences were found mainly on mother- and father-completed internalizing construct subscales drawn from one questionnaire measure.

Between-study inconsistencies were also found on tests of the diagnostic status of the targeted construct and clinically significant change. For example, unlike the first study, this second study reported statistical tests comparing intervention and control conditions on diagnostic status. However, these tests were conducted by combining the diagnostic status of children receiving the individual cognitive–behavioral treatment examined by Kendall (1994) with children receiving another intervention. This combined group was then compared to the control condition. Thus,

statistical tests of diagnostic status did not examine whether there were differences between children in the individual cognitive–behavioral treatment condition and children in the control condition. Moreover, unlike the first study, this second study did not report statistical tests of clinically significant change, and clinically significant change was measured using only parent-completed measures (as an aside, between-study inconsistencies across diagnostic status and clinically significant change methods of analysis were due, in part, to between-study inconsistencies in use of the methods, or inconsistent data-analytic approaches. However, findings derived from the only method of analysis for which between-study consistencies in methods and analytic approaches were apparent [mean differences] were highly inconsistent between studies). Therefore, a review of between-study inconsistencies suggests that change was both founded and unfounded across these two investigations. Most importantly, the between-study change attributable to the intervention was unfounded across measures of the psychological construct targeted for intervention, and

Table 3
Intervention Outcomes for Barrett et al. (1996): Reported Statistical Differences Between Individual Cognitive–Behavioral Treatment and Control Conditions

Informant	Measure	Method		
		Mean differences	Diagnostic status	Clinically significant change
Child	RCMAS[a]	No differences	—	—
Child	FSSC-R[a]	No differences	—	—
Child	CDI[b]	No differences	—	—
Mother	CBCL-Internalizing[c]	Intervention × time	—	—
Mother	CBCL-Externalizing[d]	Intervention × time	—	—
Father	CBCL-Internalizing[c]	Intervention × time	—	—
Father	CBCL-Externalizing[d]	No differences	—	—
Parent and child	ADIS-C/P	—	—	—
Independent assessor	Threat Scores[a]	No differences	—	—
Independent assessor	Avoidant Solutions[a]	No differences	—	—
Independent assessor	Percentage Choosing Avoidant Solutions	—	—	—
Statistically significant differences reported: Overall		3 of 9 Measures	N/A	N/A
Statistically significant differences reported: Target construct[a]		0 of 4 Measures	N/A	N/A
Statistically significant differences reported: Comorbid constructs/ associated features[b]		0 of 1 Measure	N/A	N/A
Statistically significant differences reported: Broadband internalizing constructs[c]		2 of 2 Measures	N/A	N/A
Statistically significant differences reported: Broadband externalizing constructs[d]		1 of 2 Measures	N/A	N/A

Note. RCMAS = Revised Children's Manifest Anxiety Scales; FSSC-R = Fear Survey Schedule for Children-Revised; CDI = Children's Depression Inventory; CBCL = Children's Behavior Checklist (Achenbach & Edelbrock, 1991); ADIS-C/P = Anxiety Disorders Interview Schedule for Children: Parent and Child Versions (Silverman & Nelles, 1988); — = Statistical tests not reported in the study; Intervention × time = Tests of mean differences showing a significant intervention × time interaction effect. As in the original study, results summarized only reflect statistical analyses reported in the study denoting differences between individual cognitive-behavioral treatment and control conditions.
[a] Measures assessing target construct. [b] Measures assessing comorbid constructs/associated features. [c] Measures assessing broadband internalizing constructs. [d] Measures assessing broadband externalizing constructs.

inconsistencies between studies were found on identical outcome measures.

As before, we have provided an example of the inconsistencies evident between studies examining the efficacy of the same intervention, without any implication that these studies are unique in any negative sense. Studies examining the same psychological intervention for child constructs other than anxiety (e.g., Coping with Depression Course for Adolescents; Clarke, Rohde, Lewinsohn, Hops, & Seeley, 1999; Lewinsohn, Clarke, Hops, & Andrews, 1990) and other forms of intervention (e.g., fluoxetine for obsessive–compulsive disorder; Geller et al., 2001; Liebowitz et al., 2002) convey the same point. Inconsistent findings made across two (or more) studies support the conclusion that identifying an intervention as changing a construct when inconsistent evidence exists between studies is at best arguable.

Interventions targeting psychological constructs in adults. Similar to within-study inconsistencies, between-study inconsistencies are not limited to interventions with children. This is illustrated in a second well-conducted study of relaxation that was conducted independent from Barlow et al. (1992). In an efficacy study examining relaxation for generalized anxiety disorder, relaxation was compared to a wait-list control condition on outcome measures administered to participants, as well as independent assessor ratings (Butler, Fennell, Robson, & Gelder, 1991). The

single method used to compare conditions was analyses of mean differences. Unlike Barlow et al. (1992), there were no other comparisons between conditions using any other method of examination. No preintervention outcome measure differences were reported.

As can be seen in Table 4, tests of mean differences between conditions only showed differences on 4 of 16 measures. Unlike Barlow et al. (1992), differences on measures of the target construct were found only on one measure. Perhaps more importantly, the two studies administered two identical measures of the target construct, and neither found differences on one measure, and only Barlow et al. (1992) found differences on the other measure (cf. Tables 2 and 4). Unlike Barlow et al. (1992), no differences were found on independent assessor ratings.

Similar to our example of inconsistent findings in studies examining interventions for children, the efficacy of this adult intervention was both founded and unfounded across these two investigations.[3] Other studies examining the same psychological

---

[3] The examples we cite from the adult intervention literature may appear dated. We emphasize that these articles are important to cite, as they form part of the evidence base for this intervention (progressive muscle relaxation for generalized anxiety disorder; see Barlow et al., 2002). Further,

Table 4
Intervention Outcomes for Butler et al. (1991): Reported Statistical Differences Between Relaxation and Control Conditions

		Method		
Informant	Measure	Mean differences	Responder status	High-endstate functioning
Self	BAI[a]	No differences	—	—
Self	STAI-Trait[a]	No differences	—	—
Self	Leeds-Anxiety[a]	No differences	—	—
Self	Leeds-Depression[b]	No differences	—	—
Self	0–8 Self-Anxiety[a]	Intervention > control	—	—
Self	0–8 Self-Depression[b]	No differences	—	—
Self	BDI[b]	Intervention > control	—	—
Self	DAS[b]	Intervention > control	—	—
Self	CCL[b]	Intervention > control	—	—
Self	SPQ[b]	No differences	—	—
Self	Interpretation[b]	No differences	—	—
Self	FNE[b]	No differences	—	—
Self	Individual Thoughts[b]	No differences	—	—
Independent assessor	0–8 Assessor-Anxiety[a]	No differences	—	—
Independent assessor	0–8 Assessor-Depression[b]	No differences	—	—
Independent assessor	Hamilton-Anxiety[a]	No differences	—	—
Statistically significant differences reported: Overall		4 of 16 Measures	N/A	N/A
Statistically significant differences reported: Target construct[a]		1 of 6 Measures	N/A	N/A
Statistically significant differences reported: Comorbid constructs/ associated features[b]		3 of 10 Measures	N/A	N/A

Note. BAI = Beck Anxiety Inventory (Beck, Brown, Epstein, & Steer, 1988); STAI-Trait = State-Trait Anxiety Inventory; Leeds-Anxiety/Depression = Leeds Scales for the Self-Assessment of Anxiety and Depression (Snaith, Bridge, & Hamilton, 1976); 0–8 Self-Anxiety/Depression = 9-point rating scale of anxiety and depression rated by patient (Snaith et al., 1976); BDI = Beck Depression Inventory; DAS = Dysfunctional Attitude Scale (Rush, Weissenburger, & Eaves, 1986); CCL = Cognition Checklist (Beck, Brown, Eidelson, Steer, & Riskind, 1987); SPQ = Subjective Probabilities Questionnaire (Butler, Gelder, Hibbert, Cullington, & Klimes, 1987); Interpretation = measure of interpretations of ambiguous material in a threatening way (Butler et al., 1987); FNE = Fear of Negative Evaluation Scale (Watson & Friend, 1969); 0–8 Assessor-Anxiety/Depression = 9-point rating scale of anxiety and depression rated by assessors (Watson & Marks, 1971); Hamilton = Hamilton Anxiety Scale; — = Statistical tests not reported in the study; Intervention > control = Tests of mean differences showing a significant difference between intervention and control conditions. As in the original study, results summarized only reflect statistical analyses reported in the study denoting differences between relaxation and control conditions.
[a] Measures assessing target construct. [b] Measures assessing comorbid constructs/associated features.

intervention for adult psychological constructs other than anxiety disorders (e.g., Coping with Depression Course; Brown & Lewinsohn, 1984; Dowrick et al., 2000), and other forms of intervention (e.g., acamprosate for alcohol dependence; Chick, Howlett, Morgan, & Ritson, 2000; Tempesta, Janiri, Bignamini, Chabac, & Potgieter, 2000) convey the same point. Again, inconsistent findings made across two (or more) studies of the same intervention present difficulties in concluding that the intervention changes the targeted construct.

---

recent work in classifying interventions cites these studies as evidence supporting the intervention (see Chambless & Ollendick, 2001; Nathan & Gorman, 2002; Roth & Fonagy, 2005). Moreover, the purpose of citing these studies was to remain consistent with the examples drawn from the child intervention literature. Indeed, the child intervention research we cite examined these interventions relative to control conditions. However, much of the recent research examining adult interventions has advanced beyond comparing interventions to control conditions, and now compares interventions to alternative interventions (see Lambert & Ogles, 2004). Thus, we both cite Barlow et al. (1992) and Butler et al. (1991) and illustrate the RPC Model using these studies in order to maintain parallel methodologies across the examples of intervention research drawn from the adult and child intervention literatures.

The Range of Possible Changes Framework

Rationale for the Framework

The purpose of comparing and contrasting the findings made both within and between intervention studies was not to suggest that inconsistencies are found across indicators of change. Indeed, as mentioned previously, intervention research acknowledges the multidimensional nature of constructs often targeted for intervention. Further, indicators often encompass a diverse spectrum of functioning (e.g., symptom/diagnostic presentation of the target construct, symptom/diagnostic presentation of comorbid constructs, associated, risk, and causal factors of the target construct). Therefore, in many instances, different indicators should produce discrepant findings, and change should often be highly variable. However, we cited and summarized previously evidence that inconsistencies are found on indicators that purportedly represent change in the constructs targeted for intervention. Moreover, we cited instances in which these inconsistencies were found between investigations of the same intervention. In some cases, inconsistencies between investigations were found on the same indicators.

The purpose of noting within- and between-study inconsistencies highlights a broader issue. Specifically, when acknowledging that constructs targeted for intervention are multidimensional, in-

tervention research ought to also acknowledge that interventions developed to change these constructs will likely not change them uniformly. Change in intervention research is highly variable, and current systems that gauge evidence-based interventions do not take this variability into account. For instance, one can envision a variety of instances in which an intervention may change measures of a construct, and at the same time, envision a variety of other instances in which change may not occur (e.g., change in participants' self-perceptions of the target construct, but no change in spousal perceptions; change in symptoms of the target construct, but no change in complete diagnostic recovery; change in independent assessors' perceptions of the target construct, but no change in physiological reactivity to stimuli that promote increased levels of the target construct). Even this dichotomous distinction does not enter into the larger theoretical and empirical discussion of intervention change.

The intervention research literature seemingly treats change as uniform and invariable. This observation is evident by the fact that current systems are identifying interventions as evidence-based, and are not qualifying this evidence to any significant degree. Indeed, beyond such distinctions that the evidence across studies suggests an intervention changes a construct, or probably changes a construct, or does not change a construct, current systems do not acknowledge the variability of change in any significant sense.

Intervention research and theory are in dire need of a theoretical and methodological framework to guide conceptualizations and examinations of change within and between studies. Indeed, if change is variable, investigations must acknowledge, conceptualize, and specify this variability a priori. A model that acknowledges the within- and between-study variability of change can guide the development and testing of theory-driven hypotheses to examine instances in which interventions may change constructs, and instances in which change in constructs may not occur. In addition, such a model may be used to conceptualize gradations of change.

Overview of Theoretical Conceptualization

We refer to the proposed framework as the Range of Possible Changes (RPC) Model. The premise of the RPC Model is that targeting multidimensional constructs for intervention necessitates the multidimensional conceptualization of intervention change. In turn, a multidimensional conceptualization of change requires conceptualizing change as highly variable and existing along a range of possible changes. Most importantly, because change is conceptualized as multidimensional, the RPC Model encourages researchers to make a priori conceptualizations of the magnitudes by which interventions change constructs. The RPC Model assumes that changes within interventions will not be uniform. Therefore, it is critically important to hypothesize the extent and/or instances in which interventions change constructs.

The theoretical conceptualization of the RPC Model is presented in Figure 1. The figure highlights that an intervention can target any of a number of dimensions of a construct. Further, the change construct of that intervention is conceptualized along dimensions in conjunction with the dimension or dimensions of the targeted construct. This conceptualization of change is ultimately that of a construct that is highly variable and multidimensional. Specifically, a multidimensional change construct requires conceptualization of: (a) the extent to which the intervention will change the

dimension or dimensions of the targeted construct; (b) the instances in which change is conceptualized to occur and not occur; and (c) the instances in which change is conceptualized to occur, but vary in magnitude along a range of possible changes.

Overview of Measurement Conceptualization

The measurement conceptualization of the RPC Model is presented in Figure 2. The key premise behind the measurement conceptualization of the framework is that the multidimensional conceptualization of change necessitates the incorporation of both multiple indicators of change and multiple methods of examining change. Under these circumstances, the RPC Model conceptualizes a single finding as an item on a measure of change. Within this measure, there are a number of "subscales" (e.g., findings made by examining diagnostic status; findings made by examining measures completed by intervention participants). Within any given dimension, the extent to which items across subscales of that dimension consistently support change may suggest the strength of the evidence for change. For instance, consistent findings within the dimension of change in the symptom and diagnostic presentation of the target construct may suggest the intervention's ability to change the symptom and diagnostic presentation of the construct. Conversely, the extent to which subscales differ in whether they suggest change may provide insight into how the intervention may change the construct in some instances, but not others (e.g., findings gathered via mean differences method suggest change, but not findings gathered via other methods of analysis; multiple informant data suggest change, but not performance-based or physiological data).

The measurement conceptualization of the RPC Model assumes that findings of change will exist along a range of possible changes. This idea is highlighted by the incorporation of multiple indicators and methods of examination to reach RPCs (Ranges of Possible Changes). Each RPC is anchored at each end by a lower limit indicator of change and an upper limit indicator of change. In Figure 2, the sequences of, and arrows leading from different informants to RPCs for different methods acknowledge the use of multiple informant data in constructing RPCs. These sequences do not suggest which informant data result in which types of observed intervention effects (e.g., caregiver/significant other findings suggest lower or higher observed intervention effects, relative to effects observed using data from other informants). Relatedly, the sequences of, and arrows leading from different RPCs for different methods acknowledge the use of multiple methods in constructing RPCs. These sequences do not suggest which method data result in which observed effects (e.g., diagnostic status findings suggesting lower or higher observed effects, relative to findings derived from data examined using other methods).

Most importantly, the measurement conceptualization of the RPC Model ought to be aligned with the conceptualized dimension of change under investigation. An example may be if the focus of conceptualizing change is the dimension of the targeted construct's symptom and diagnostic presentation. The measurement conceptualization of the RPC Model ought to inform the measurement and examination of this specific dimension of change. Of course, this does not mean that theoretical and measurement conceptualizations of a particular or single dimension of change should be the only conceptualization tested within investigations (e.g., this does not mean that investigations should be concerned only with the
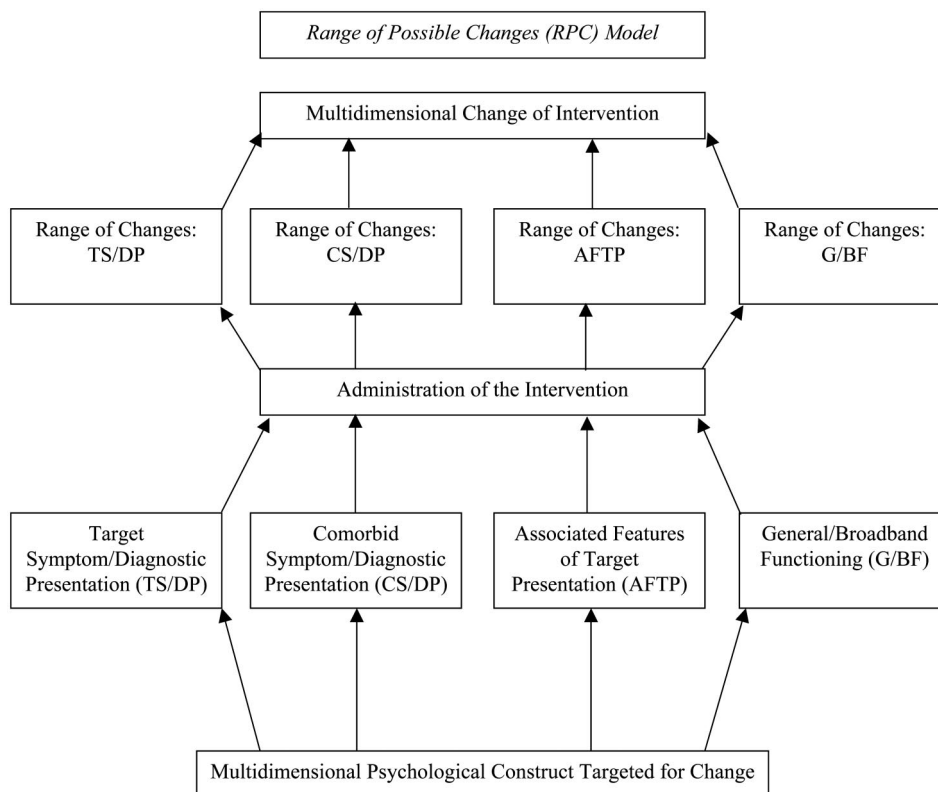
Figure 1. Theoretical conceptualization of the Range of Possible Changes (RPC) Model. Change within interventions targeting multidimensional psychological constructs is conceptualized as highly variable and multidimensional. Dimensions of the target psychological construct (e.g., target symptom/diagnostic presentation, associated features of target presentation) are not meant to be exhaustive. Dimensions of the targeted construct may change as conceptualizations of change within the intervention are revised or are complemented by additional or revised dimensions of the construct, and thus, by additional or revised dimensions of change.

dimension representing change in the targeted construct's symptom and diagnostic presentation). The purpose of clarifying this aspect of the measurement conceptualization of the RPC Model is to emphasize that, for each dimension of the targeted construct, and thus, for each dimension of change, a corresponding measurement conceptualization of these underlying dimensions ought to be constructed.

How the RPC Framework Addresses Within-Study Inconsistencies

The RPC Model posits that change ought to be interpreted and classified across a range of possible changes. At the same time, the consistency by which an intervention brings about change across indicators, or whether the majority of indicators suggest change can be used to arrive at sound interpretations and conceptualizations of change. Table 5 presents classification categories extrapolated from the theoretical and measurement conceptualizations of the RPC Model. These categories can be used to interpret an investigation's findings as to the specific dimension of change examined. Broadly, under the RPC Model categories, the ideal or best-case scenario for evidence for change ("Best Evidence for Change") requires the RPC for a particular dimension be statistically significant on 80% of the range of informants, measures, and

methods employed. Less ideal categorizations of change primarily depend on the percentage of significant findings, and/or whether findings are consistent across use of specific informants, measures, and/or methods of examination.

Second, the conceptual basis of the RPC Model acknowledges that, because the constructs targeted for intervention are multidimensional, conceptualizations of change must be multidimensional. Thus, the RPC Model must be able to incorporate multiple indicators that may each be representative of one of multiple dimensions. A strength of the RPC Model is that its classification categories can be used to make separate interpretations of the evidence for change, depending on the dimension. Indeed, as illustrated later, separate RPCs can be constructed for classifications of change, depending on the dimension or dimensions examined.

Third, the RPC framework acknowledges that both within and between categories of classifying change, variability will exist in classifications within studies. For instance, a study may find that a majority of indicators suggests change, but the magnitudes of change may vary widely across indicators. The RPC Model addresses additional concerns with regard to interpreting and classifying change by incorporating effect size conventions (e.g., Cohen's *d* measures of small [.2], medium [.5] and large [.8] effects;
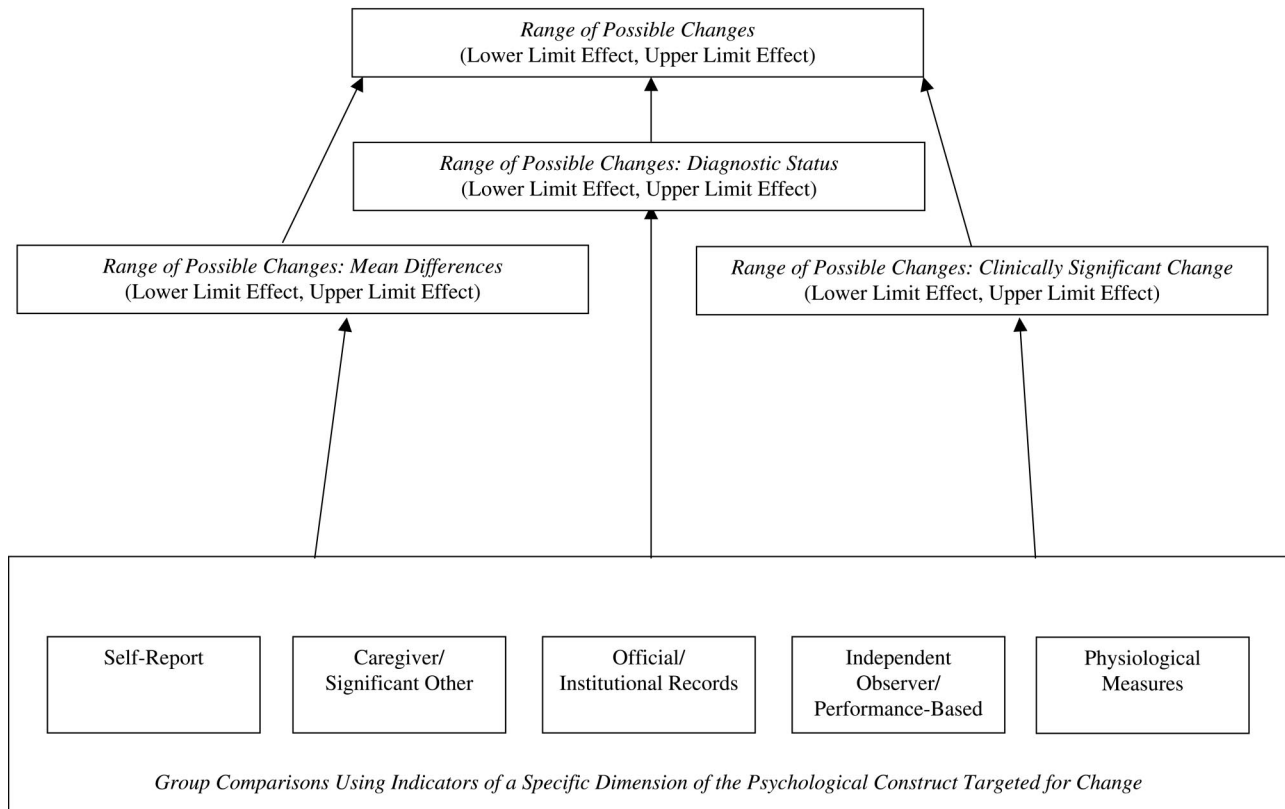
Figure 2. Measurement conceptualization of the Range of Possible Changes (RPC) Model. Multiple method and source data are utilized to construct RPCs for multiple methods of examining change. Ranges of Possible Changes for multiple methods can be used to construct RPCs for data gathered across methods and sources. Under the RPC Model, source data are gathered using multiple methods of measurement as well (e.g., self-report data gathered using questionnaire rating scales and structured interviews; institutional records gathered from multiple methods/contexts of functioning such as occupational or salary records, hospital admittance rates, criminal records, and grade point average or absentee rates from academic institutions).

Cohen, 1988) to complement the RPC Model's categories. Thus, along with the RPC Model category classification, the range of observed magnitudes of change should be characterized using accepted effect size conventions. An example may be an investigation interested in classifying changes in the symptom presentation of the targeted construct. In this scenario, if a study is classified under the "Best Evidence for Change" category, with observed effect sizes ranging from *d*s of .3 to 1.2, such an investigation may provide "best evidence" to suggest small-to-large positive magnitudes of the intervention's ability to change the symptom presentation of the construct.

How the RPC Framework Addresses Between-Study Inconsistencies

The RPC Model assumes that investigations of the same dimension of the same intervention may differ in the consistency by which indicators will suggest change. Through the incorporation of the RPC Model's categories, along with effect size conventions, the evidence provided by indicators across studies can be compared to decipher whether they classify change at similar magnitudes. As a result, the RPC Model can be used along with effect size conventions to classify the evidence provided by two or more

investigations of the same dimension of an intervention. An example may be if the evidence provided by two studies for the same dimension could each be classified within the "Best Evidence for Change" category, with findings ranging from *d*s of .5 to 1.0. In this example, the two studies may be classified as suggesting similar consistent change, and similar estimates of medium-to-large positive change.

Defining the RPC Model's Categories

A strength of the RPC Model is that its basic principles can be used to construct categories to interpret the evidence for change. Descriptions of the RPC Model classification categories are presented in Table 5, and elaborated on below. Classifications are placed into two categories: (a) evidence for nonspecific change (i.e., no informant-, context-, measure-, or method-specific change); and (b) evidence for change specific to an informant, context, measure, and/or method. Classifications for nonspecific change are hierarchical, and can be distinguished by percentages of significant effects found. For instance, the "Best Evidence for Change" category denotes studies finding significant change on 80% of a range of informants, measures, and methods of analysis, whereas the "Evidence for Probable Change" category denotes

Table 5
Recommended Categories for Classifying the Within-Study Change for Investigations Using the Range of Possible Changes (RPC) Model

Category	Criteria	Interpretation
Best Evidence for Change	At least 80% of the findings from multiple informants, measures, and methods of analyzing outcomes show significant results; predominantly significant results found on three or more informant's ratings, measures, and methods; no clear informant-specific, measure-specific, or method-specific pattern of significant results.	Grand majority of evidence for change across range of informants, measures, and methods indicates that conditions are significantly different; investigation provides sufficient evidence to suggest the intervention changes the dimension of the construct.
Evidence for Probable Change	More than 50% of the findings from multiple informants, measures, and methods of analyzing outcomes show significant results; significant results found on simple majority of three or more informant's ratings, measures, and methods; no clear informant-specific, measure-specific, or method-specific pattern of significant results.	Simple majority of evidence for change across range of informants, measures, and methods indicates evidence for probable differences between conditions; investigation may suggest probable change in the dimension of the construct, and suggest reasons why inconsistent changes across measures were found.
Limited Evidence for Change	Either 50% or less of the findings from three or more informant's ratings, measures, and methods show significant results, or less than grand majority (less than 80%) of findings from specific informant's ratings, measures, and/or methods show significant results; significant results are either sporadically found across a range of informant's ratings, measures, or methods of analysis, or are not found on specific informant's ratings, measures, and/or methods, to a degree that warrants classification in a category denoting specificity of change; no clear pattern of significant results.	Sparse evidence for change; intervention may not change the dimension of the construct.
No Evidence for Change	No significant results are observed.	No evidence for change; intervention likely may not change the dimension of the construct.
Evidence for Contextual- or Informant-Specific Change	Significant results are found on grand majority (80%) of ratings provided by specific informant(s), and limited or no evidence (50% or less) is found on ratings of other informant(s); clear contextual or informant-specific pattern of significant results.	No definitive evidence for change; investigation may suggest evidence for change in the dimension of the construct that is perhaps specific to when the construct is exhibited in specific context(s) or in interactions with specific informant(s); future experimental work would be needed to examine whether the intervention changes the dimension of the construct, but only when the construct is exhibited in specific contexts or situations, or when the construct is exhibited in interactions between the participant and specific informant(s).
Evidence for Measure- or Method-Specific Change	Significant results are found on grand majority (80%) of specific measure(s) or method(s) of analyzing intervention outcomes, and limited or no evidence (50% or less) is found using other measures or methods; clear measure- and/or method-specific pattern of significant results.	No definitive evidence for change; investigation may suggest evidence for change in the dimension of the construct that is specific to when the construct is measured via either measure(s) for which findings were made, or method(s) of analysis for which findings were made, or both; future experimental work would be needed to examine whether change is measure- and/or method-specific.

studies finding significant change at a lower percentage (greater than 50%). In contrast, classifications for specific change are not hierarchical, and deal mainly with the type of specificity observed.

Categories should not be used to classify change across all indicators, unless all indicators in the investigation relate to a single dimension. When examining an intervention across two or more studies to decipher between-study consistencies, change should be classified within the same dimension. Further, when selecting a range of sources to study change, this range should not necessarily be restricted to self- or observer-assessments. Change can be examined using a number of sources, such as self-report, reports from significant others, use of performance-based or physiological measures, institutional records, and reports from independent laboratory observations (see Tables 6–9; Figure 2; Kazdin, 2003; Kendall et al., 1999). Indeed, multiple sources (e.g., parent, child, teacher, clinician, independent observer, physiological measurements) are employed in research examining interventions targeting constructs perceived as difficult to measure using methods

other than self-assessments (e.g., childhood anxiety; Kendall, 1994; Barrett et al., 1996, 2004; Geller et al., 2001; Liebowitz et al., 2002; Öst et al., 2001).

Categories Classifying Nonspecific Change

Best Evidence for Change. The category "Best Evidence for Change" classifies findings made under optimal conditions. Under this category, 80% of findings from multiple informants, measures, and methods of analysis (i.e., three or more of each) should reflect evidence for change; no clear informant-specific, measure-specific, or method-specific change should be found. This optimal scenario is reserved for instances in which the range of findings is predominantly significant, and there is evidence for change across a range of informants, measures, and methods of examination. One can draw a parallel between this category and internal consistency (e.g., Kazdin, 2003; Nunnally & Bernstein, 1994). Indeed, researchers and theorists commonly conceptualize the ideal scenar-

Table 6

Intervention Outcomes for Kendall (1994): Illustration of the Range of Possible Changes (RPC) Model to Examine Differences Between Individual Cognitive–Behavioral Treatment and Control Conditions

Informant	Measure	Method		
		Mean differences (between-group effect size [*d*])	Diagnostic status (between-group effect size [*d*])	Clinically significant change (between-group effect size [*d*])
Child	ADIS-C	—	—	—
Child	RCMAS[a]	.78*	—	.80*
Child	STAIC A-Trait[a]	1.05*	—	—
Child	STAIC A-State[a]	1.09*	—	—
Child	FSSC-R[a]	.40*	—	—
Child	CQ-C[b]	1.07*	—	—
Child	CDI[b]	.48*	—	.85*
Child	NASSQ[b]	.46*	—	—
Parent	ADIS-P	—	—	—
Parent	CBCL-Internalizing[c]	1.52*	—	—
Parent	CBCL-Social[b]	−.45*	—	—
Parent	CBCL-Health[b]	.99*	—	—
Parent	CBCL-Externalizing[d]	.97*	—	—
Parent	STAIC A-Trait-P[a]	.36*	—	—
Teacher	TRF-Internalizing[c,e]	.36	—	—
Teacher	TRF-Externalizing[d]	−.17	—	—
Independent assessor	Gratuitous Verbalizations[a]	0	—	—
Independent assessor	Gratuitous Body Movements[a]	0	—	—
Independent assessor	Trembling Voice[a]	0	—	—
Independent assessor	Avoiding Task[a]	0	—	—
Independent assessor	Absence of Eye Contact[a]	0	—	—
Independent assessor	Fingers in Mouth[a]	0	—	—
Independent assessor	Body Rigidity[a]	0	—	—
Independent assessor	Total Behavioral Codes[a]	2.14*	—	—
		RPC: (−.45*, 2.14*), % *p* < .05 = 59	RPC: N/A	RPC: N/A

Overall RPC: (−.45*, 2.14*), % *p* < .05 = 63
Symptom presentation of target construct[a] RPC: (0, 2.14*), % *p* < .05 = 50
(Evidence for Change Specific to Child Self-Rated Questionnaires, Examined Via Mean Differences [.40*, 1.09*], % *p* < .05 = 100)
Symptom presentation of comorbid constructs/associated features of target construct[b] RPC: (−.45*, 1.07*), % *p* < .05 = 100
(Evidence for Change Specific to Child Self-Rated Questionnaires, Examined Via Mean Differences [.46*, 1.07*], % *p* < .05 = 100)
Broadband internalizing constructs[c] RPC: N/A
Broadband externalizing constructs[d] RPC: N/A

Note. ADIS-C/P = Anxiety Disorders Interview Schedule for Children: Parent and Child Versions; RCMAS = Revised Children's Manifest Anxiety Scales; STAIC = State-Trait Anxiety Inventory for Children; FSSC-R = Fear Survey Schedule for Children-Revised; CQ-C = Coping Questionnaire for Children; CDI = Children's Depression Inventory; NASSQ = Children's Negative Affectivity Self-Statement Questionnaire; CBCL = Children's Behavior Checklist; STAIC A-Trait-P = State-Trait Anxiety Inventory for Children-Parent Form; TRF = Teacher Report Form; * *p* < .05; — = Comparisons between intervention and control conditions not conducted in the study, or could not be estimated because of insufficient information.
[a] Measures assessing symptom presentation of target construct. [b] Measures assessing symptom presentation of comorbid constructs/associated features of target construct. [c] Measures assessing broadband internalizing constructs. [d] Measures assessing broadband externalizing constructs. [e] Analyses on this measure were conducted for both the total sample and a subgroup of children. This finding was made for the entire sample.

ios for the internal consistency of a measure to be those instances in which the items comprising the measure correlate highly with each other. For instance, intervention researchers and theorists often define this scenario to be instances in which 80% of the proportion of variance in scores from a measure is accounted for by "true score," or the construct the measure purports to measure (e.g., Nunnally & Bernstein, 1994; Ryan, Lopez, & Sumerall, 2001). The RPC Model categories classify change within single dimensions; instances when it would be optimal for findings to be highly correlated. Therefore, the concept of internal consistency readily applies to classifying change using the RPC Model.

Evidence for Probable Change. Change categorized under "Evidence for Probable Change" denotes investigations for which a simple majority (greater than 50%) of findings suggests change across multiple informants, measures, and methods of analysis

(i.e., three or more of each). Further, no clear informant-specific, measure-specific, or method-specific effect should be found. This category classifies instances in which the majority of the range of findings is significant, and these findings are present across a range of informants, measures, and methods of examination. One can conceptualize studies under this category as those for which the available evidence suggests that there is a greater likelihood than not that the intervention changes the dimension of interest. Given this classification, additional work should examine whether the intervention is likely to produce nonspecific changes in the dimension, or whether there is indeed a limit (e.g., contextually or methodologically specific) to change.

Limited or No Evidence for Change. Two final categories classify the poorest nonspecific evidence. First, "Limited Evidence for Change" denotes those studies for which either: (a) 50% or

Table 7
Intervention Outcomes for Barrett et al. (1996): Illustration of the Range of Possible Changes (RPC) Model to Examine Differences Between Individual Cognitive–Behavioral Treatment and Control Conditions

Informant	Measure	Method		
		Mean differences (between-group effect size [d])	Diagnostic status (between-group effect size [d])	Clinically significant change (between-group effect size [d])
Child	RCMAS[a]	.43	—	—
Child	FSSC-R[a]	.44	—	—
Child	CDI[b]	.43	—	—
Mother	CBCL-Internalizing[c]	.91*	—	—
Mother	CBCL-Externalizing[d]	.43*	—	—
Father	CBCL-Internalizing[c]	.61*	—	—
Father	CBCL-Externalizing[d]	.10	—	—
Parent and child	ADIS-C/P[a]	—	.65*	—
Independent assessor	Threat Scores[a]	.56	—	—
Independent assessor	Avoidant Solutions[a]	.28	—	—
Independent assessor	Percentage Choosing Avoidant Solutions, Post-Discussion	—	—	—
		RPC: (.10, .91*), % p < .05 = 33	RPC: N/A	RPC: N/A

Overall RPC: (.10, .91*), % p < .05 = 40
Symptom/diagnostic presentation of target construct[a] RPC: (.28, .65*), % p < .05 = 20
(Limited Evidence for Change)
Symptom presentation of comorbid constructs[b] RPC: N/A
Broadband internalizing constructs[c] RPC: N/A
Broadband externalizing constructs[d] RPC: N/A

Note. RCMAS = Revised Children's Manifest Anxiety Scales; FSSC-R = Fear Survey Schedule for Children-Revised; CDI = Children's Depression Inventory; CBCL = Children's Behavior Checklist; ADIS-C/P = Anxiety Disorders Interview Schedule for Children: Parent and Child Versions; * $p <$ .05; — = Comparisons between intervention and control conditions not conducted in the study, or could not be estimated because of insufficient information.
[a] Measures assessing symptom/diagnostic presentation of target construct. [b] Measures assessing symptom presentation of comorbid constructs. [c] Measures assessing broadband internalizing constructs. [d] Measures assessing broadband externalizing constructs.

fewer of the findings from three or more informants, measures, and methods suggest change; or (b) fewer than 80% of findings from specific informants, measures, and/or methods suggest change. Significant results are either sporadically found across a range of informants, measures, or methods of analysis, or not found for specific informants, measures, and/or methods, to a degree that warrants classification in a category denoting specificity of change. Second, studies falling under the "No Evidence for Change" category do not report any evidence for change.

Categories Classifying Specific Change

Evidence for Contextual- or Informant-Specific Change. The category "Evidence for Contextual- or Informant-Specific Change" classifies studies for which a clear pattern of significant findings (80%, similar to "Best Evidence for Change") exists for a given informant or informants. Further, limited or no evidence (50% or less) is found employing other informants' or sources' ratings. Classifications in such a category may suggest that the intervention may only change the construct's expression in, for instance, specific contexts (e.g., home, school), or in interactions between the participant in the intervention and the source providing the information (e.g., mother, significant other). Needless to say, no informant or source can provide definitive or "gold standard" information of a construct expressed within a given context (i.e., teacher ratings of a child in an intervention cannot be interpreted as a definitive description of how the child is behaving in

school). Thus, additional experimental work would be necessary to suggest that the intervention may change the targeted dimension of the construct only when expressed within a given context or within specific interactions with people in the participant's life (e.g., additional intervention studies experimentally examining a version of the intervention that specifically targets the construct when displayed in the identified context).

Evidence for Measure- or Method-Specific Change. Studies classified in the "Evidence for Measure- or Method-Specific Change" category may be characterized by a clear pattern of significant findings (80%, similar to "Best Evidence for Change") on either a specific type of indicator, or specific method of analyzing change. Moreover, limited or no evidence (50% or less) is found on other measures or methods. Studies in this category suggest that findings are reflected only when employing specific methods of assessing change, or specific methods of examination.

Findings classified in this category may be attributable to a number of factors. For example, findings may result from systematic use of insensitive or unreliable types of indicators (e.g., use of unreliable questionnaire measures, but reliable structured interview measures). Alternatively, findings may result from the possibility that the intervention may change the examined dimension in one fashion but not another (e.g., findings suggesting change on continuous measures of the symptom and diagnostic presentation of the target construct, but not categorical measures). Moreover, inconsistencies may be the result of the interactive effects of

Table 8
Intervention Outcomes for Barlow et al. (1992): Illustration of the Range of Possible Changes (RPC) Model to Examine Differences Between Relaxation and Control Conditions

		Method		
Informant	Measure	Mean differences (between-group effect size [d])	Responder status (between-group effect size [d])	High-endstate functioning (between-group effect size [d])
Self	STAI-Trait[a]	.56	—	—
Self	CSAQ-Cognitive[a]	1.26	—	—
Self	CSAQ-Somatic[a]	1.55	—	—
Self	FQ-Total[b]	−.09	—	—
Self	FQ-Anxiety/Depression[b]	1.15	—	—
Self	BDI[b]	.84*	—	—
Self	EPI-Extraversion[b]	.75	—	—
Self	EPI-Neuroticism[c]	−1.64	—	—
Self	SSS[d]	.83	—	—
Self	Average Daily Anxiety[a]	.44*	—	—
Self	Average Daily Depression[b]	0	—	—
Self	Weekly Intense Anxiety Episodes[a]	.53	—	—
Self	Percent of Day Worrying[a]	.85*	—	—
Independent assessor	ADIS-R Clinical Severity[a]	2.00*	—	—
Independent assessor	Hamilton Anxiety[a]	1.85*	—	—
Independent assessor	Hamilton Depression[b]	.92*	—	—
Self and independent assessor	Composite Scale-TR	—	—	—
Self and independent assessor	Composite Scale-HEF	—	—	—
		RPC: (−.09, 2.00*), % $p < .05$ = 40	RPC: N/A	RPC: N/A

Overall RPC: (−.09, 2.00*), % $p < .05$ = 40
Symptom presentation of target construct[a] RPC: (.44*, 2.00*), % $p < .05$ = 50
(Limited Evidence for Change)
Symptom presentation of comorbid constructs/associated features of target construct[b] RPC: (−.09, 1.15), % $p < .05$ = 33
(Limited Evidence for Change)
General functioning[d] RPC: N/A

Note. STAI-Trait = State-Trait Anxiety Inventory; CSAQ = Cognitive-Somatic Anxiety Questionnaire; FQ = Fear Questionnaire; BDI = Beck Depression Inventory; EPI = Eysenck Personality Inventory; SSS = Subjective Symptoms Scale; ADIS-R = Anxiety Disorders Interview Schedule-Revised; Hamilton Anxiety/Depression = Hamilton Anxiety and Depression Scales; Composite Scale-TR = Scales combining scores from ADIS-R, Hamilton Anxiety, Average Daily Anxiety, and SSS to determine responder status; Composite Scale-HEF = Scales combining scores from ADIS-R, Hamilton Anxiety, Average Daily Anxiety, FQ, and SSS to determine those exhibiting high endstate functioning; * $p < .05$; — = Comparisons between intervention and control conditions not conducted in the study, or could not be estimated due to insufficient information.
[a] Measures assessing symptom presentation of target construct. [b] Measures assessing symptom presentation of comorbid constructs/associated features of target construct. [c] The effect size denoting differences between intervention and control conditions on the EPI-Neuroticism scale was not included in any of the RPCs, because significant pre-intervention differences between intervention and control conditions were reported in the original article. [d] Measures assessing general psychosocial functioning.

method of examination, type of indicator, and/or differential measurement unreliability. Although it appears that separate categories should be constructed for both measure- and method-specificity, it is important to maintain a category that encompasses findings that are measure- and/or method-specific, because of the interrelations between measures of change, and methods of examining change. Similar to studies under the "Evidence for Contextual- or Informant-Specific Change" category, additional experimental work should examine whether the intervention changes the dimension in some ways (e.g., reduction of symptoms) but not others (e.g., diagnostic recovery).

Critical Considerations

Dimensions of Change as Foci of Intervention Research

Use of the RPC Model to study intervention change raises a number of critical issues. Perhaps the most critical consideration

involves which dimensions to examine. Indeed, how does one decide which dimension to focus on, given a multidimensional conceptualization of change? Conceptualizing change as multidimensional is not synonymous with the idea that all dimensions are equal, and therefore, should be focused upon equally. Dimension selection depends heavily on the type of intervention and at what period the intervention is being evaluated. Table 10 lists recommendations for dimension selection in intervention research.

Treatment research/immediate postadministration. In treatment research, the psychological construct being targeted for change is most often present in the participants being administered the intervention. As mentioned previously, psychological constructs targeted for intervention are often specific psychological problems or disorders; constructs that have negative social, emotional, and physical impacts on people experiencing them. Thus, when research evaluates a treatment immediately after it is administered, perhaps the most critical goal is to provide participants immediate psychosocial relief from the construct targeted for

Table 9

Intervention Outcomes for Butler et al. (1991): Illustration of the Range of Possible Changes (RPC) Model to Examine Differences Between Relaxation and Control Conditions

Informant	Measure	Method		
		Mean differences (between-group effect size [d])	Responder status (between-group effect size [d])	High-endstate functioning (between-group effect size [d])
Self	BAI[a]	.47	—	—
Self	STAI-Trait[a]	.30	—	—
Self	Leeds-Anxiety[a]	.16	—	—
Self	Leeds-Depression[b]	.24	—	—
Self	0–8 Self-Anxiety[a]	1.06*	—	—
Self	0–8 Self-Depression[b]	.13	—	—
Self	BDI[b]	.44*	—	—
Self	DAS[b]	.61*	—	—
Self	CCL[b]	.49*	—	—
Self	SPQ[b]	.39	—	—
Self	Interpretation[b]	.53	—	—
Self	FNE[b]	.19	—	—
Self	Individual Thoughts[b]	.27	—	—
Independent assessor	0–8 Assessor-Anxiety[a]	.56	—	—
Independent assessor	0–8 Assessor-Depression[b]	.06	—	—
Independent assessor	Hamilton-Anxiety[a]	.08	—	—
		RPC: (.06, 1.06*), $\% p < .05 = 25$	RPC: N/A	RPC: N/A

Overall RPC: (.06, 1.06*), $\% p < .05 = 25$
Symptom presentation of target construct[a] RPC: (.08, 1.06*), $\% p < .05 = 17$
(Limited Evidence for Change)
Symptom presentation of comorbid constructs/associated features of target construct[b] RPC: (.06, .61*), $\% p < .05 = 30$
(Limited Evidence for Change)

Note. BAI = Beck Anxiety Inventory; STAI-Trait = State-Trait Anxiety Inventory; Leeds-Anxiety/Depression = Leeds Scales for the Self-Assessment of Anxiety and Depression; 0–8 Self-Anxiety/Depression = 9-point rating scale of anxiety and depression rated by patient; BDI = Beck Depression Inventory; DAS = Dysfunctional Attitude Scale; CCL = Cognition Checklist; SPQ = Subjective Probabilities Questionnaire; Interpretation = measure of interpretations of ambiguous material in a threatening way; FNE = Fear of Negative Evaluation Scale; 0–8 Assessor-Anxiety/Depression = 9-point rating scale of anxiety and depression rated by assessors; Hamilton = Hamilton Anxiety Scale; * $p < .05$; — = Comparisons between intervention and control conditions not conducted in the study, or could not be estimated due to insufficient information.
[a] Measures assessing symptom presentation of target construct.   [b] Measures assessing symptom presentation of comorbid constructs/associated features of target construct.

change. Often times, "immediate relief" can be conceptualized as targeting the dimension of the construct's symptom and diagnostic presentation. This is because this presentation often defines the targeted construct (e.g., *Diagnostic and Statistical Manual of Mental Disorders* [*DSM–IV*]; American Psychiatric Association, 1994), and treatments are often developed to specifically target this presentation for change. Further, this presentation often defines, in part, the social, emotional, and physical impacts that these constructs have on people experiencing them.

The previous conceptualization of immediate relief may be appropriate for many constructs targeted for change (e.g., anxiety and depressive disorders), because symptom and diagnostic presentations of many of these constructs are amenable to change via the short-term treatments examined in research (for a review, see Lambert & Ogles, 2004). However, for more chronic or recurrent constructs (e.g., schizophrenia), symptom and diagnostic presentations may not be amenable to change. Thus, it may be appropriate in these instances to conceptualize "immediate relief" as change in a dimension that may

Table 10

Suggested Dimensions of Change as Foci of Intervention Research

Program type	Period of examination	Suggested dimensions of change
Treatment	Immediate postadministration	Target construct's symptom/diagnostic presentation (or adaptive functioning to cope with target construct, if more feasible).
	Longitudinal follow-up	Target construct's symptom/diagnostic presentation (or adaptive functioning to cope with target construct, if more feasible); factors conceptualized to play risk or causal roles in the relapse or recurrence of the target construct's symptom/diagnostic presentation (or risk or relapse in maladaptive functioning, if more feasible).
Prevention	Immediate postadministration	Factors conceptualized to play risk or causal roles in the development and/or expression of the target construct's symptom/diagnostic presentation.
	Longitudinal follow-up	Factors conceptualized to play risk or causal roles in the development and/or expression of the target construct's symptom/diagnostic presentation; target construct's symptom/diagnostic presentation.

be more amenable to change, such as adaptive functioning to cope with the psychological construct (e.g., enhanced social skills).

Treatment research/longitudinal follow-up. For research evaluating a treatment immediately after its administration, evaluating the treatment's ability to provide immediate psychosocial relief may be the most prudent goal. However, when a treatment is being evaluated at a longitudinal follow-up stage (i.e., months or years have elapsed since administration of the intervention), the dimension or dimensions examined may change. To be clear, aspects that are present in immediately postadministration research may still be present in longitudinal follow-up research. An example may be when the treatment examined is conceptualized as producing additional change long after its administration (i.e., "sleeper effects"). In such an instance, reduction in the symptom and diagnostic presentation of the target construct (or changes in adaptive functioning, if more feasible) may remain a focus of investigations of the treatment.

At the same time, the focus of research at longitudinal follow-up may become more complex, relative to research conducted immediately postadministration. Specifically, recurrence or relapse in the symptom and diagnostic presentation of the target construct (or maladaptive functioning, if more feasible) becomes a dimension of concern. This concern, of course, is held under the assumption that evidence supporting the treatment exists for evaluations immediately postadministration. An exception may be if change is conceptualized as specific to preventing relapse of disorder. An example may be treatment for relapse of depressive episodes. In this case, participants may have depression at the time of treatment, or are currently in remission, and the goal in such a case is to reduce the likelihood of future depressive episodes.

Relatedly, a dimension of change in factors that may play a risk or causal role in the relapse of the construct's symptom and diagnostic presentation ought to be evaluated or monitored as well. Indeed, if relapse in the construct's symptom and diagnostic presentation (or maladaptive functioning, if more feasible) is of interest in longitudinal follow-up treatment research, then a dimension of factors that pose risk to relapse ought to be examined as well. Relationships between relapse of the dimension initially targeted for treatment and factors posing risk to relapse of this dimension may inform subsequent examinations of interventions for relapse prevention.

Prevention research/immediate postadministration. The dimension a researcher focuses on when examining a prevention program immediately postadministration may differ markedly from an investigation examining a treatment program immediately postadministration. We emphasize that the prevention programs we refer to deal with *primary prevention*, where participants being administered the program are not currently experiencing the psychological construct targeted for prevention (for a review, see Kaplan, 2000). The goal of these programs is to reduce the likelihood that the target construct will ever develop. Primary prevention programs can be distinguished from *secondary prevention* programs, which deal with identifying an existing construct at an early stage (or before it becomes a more legitimate or pressing psychosocial concern) and attempting to reduce it. We focus on primary prevention, rather than secondary prevention, because primary prevention programs often focus squarely on behavior change. Further, prior work suggests primary prevention programs offer substantial benefits, relative to secondary prevention programs (Kaplan, 2000). Thus, our focus on primary prevention

programs is consistent with the RPC Model's focus on changes in behavior.

The goal of primary prevention programs is to reduce the chances that the targeted construct will ever develop. Whether the construct develops is often defined as whether one experiences the symptom and diagnostic presentation of the construct. Therefore, prevention research done immediately postadministration ought to examine the ability of the program to change the dimension representing factors conceptualized to play risk or causal roles in the development and/or expression of the target construct's symptom and diagnostic presentation. Of course, selection of which risk or causal factors to target for change weighs heavily on the factor or factors that the program is conceptualized to target and change.

Prevention research/longitudinal follow-up. Finally, primary prevention research examining the program at longitudinal follow-up differs markedly from the types of research discussed previously. Change in risk or causal factors continues to be a primary focus. Further, longitudinal reductions in these factors, or relapse or recurrence of these factors (or both) may be of interest, depending on the conceptualization of change. In addition, given the main goals of prevention research (reducing risk of development or expression of target construct), a dimension that represents changes in the development or expression of the symptom and diagnostic presentation of the target construct should also be examined.

Selection of Indicators to Represent Change

Acknowledging inconsistencies across indicators of change raises the issue of selection of indicators to employ within investigations. Specifically, one could argue that classifying change based, in part, on the consistency by which indicators suggest change may result in researchers selecting highly similar indicators to reach requisite levels of consistency (e.g., employing only self-rated questionnaire scales, employing only indicators that were developed by the same investigator, employing multiple indicators derived from the same structured interview measure).

Perhaps the positive implications of the RPC Model outweigh any negative implications. For instance, under the RPC Model, researchers that employ highly similar indicators of any dimension are simply limiting the potential inferences they may draw from their findings. Findings based on a restricted range of indicators would result in inferring, at best, restricted conceptualizations of change. Less restricted conceptualizations would only come from studies finding consistent evidence across a wide range of indicators.

We propose a key strength of the RPC Model is that it may encourage researchers to carefully consider the selection of indicators when conceptualizing the instances in which the interventions they examine may change the constructs they target. Indeed, promoting the development of a priori hypotheses on the instances in which change may or may not occur may encourage researchers to select indicators that reflect these hypotheses. Therefore, use of the RPC Model would likely lead to the careful selection of indicators, commensurate with researchers' conceptualizations of change. In our discussion that follows of the research implications of the RPC Model, we elaborate on how the framework may inform the development and testing of hypotheses in future work, and provide recommendations on the selection of indicators.

Differential Psychometric Properties and Power

Another critical issue to consider when using the RPC Model relates to the differential psychometric properties of measures, and the statistical power to detect significant effects. Inconsistent findings may be influenced by use of indicators that differ in their psychometric properties (e.g., differential measurement reliability, construct validity, floor and ceiling effects), as well as by whether requisite levels of statistical power are reached to detect significant effects. This is a critical issue to consider when interpreting findings within intervention studies, or indeed, when addressing any research question in science. It would be foolhardy to address a research question if one were uncertain of the availability of either reliable, valid measures of the constructs germane to the question, or adequate statistical power to detect hypothesized effects.

The RPC Model assumes that researchers have at their disposal adequate empirical conditions by which to conduct research. However, if such conditions are not readily available, we believe a strength of the RPC Model lies in its emphasis on conceptualizing change in light of the reality that the evidence for change may be variable. The RPC framework places great emphasis on conceptualizing the instances in which evidence for change may be observed. Therefore, if an investigator anticipates, for instance, that change will be suggested only by indicators for which prior work suggests attain a given threshold of reliability and validity, then this anticipation should be hypothesized and incorporated into their conceptualization of change.

If indicators that differ in their reliability and validity are employed within a study, then the extent to which differential reliability and validity relate to inconsistencies should be examined and incorporated into a conceptualization of change. Such examinations would inform the selection of indicators in subsequent research. Relatedly, if an investigator hypothesizes that an intervention's effects may be quite small, and thus, difficult to detect without a large sample, then the idea that evidence for change may be found only under certain conditions of sampling must factor into the conceptualization and measurement of change.

Perhaps the broader point to be made is that when one is conceptualizing and measuring change, one is doing so under the assumption that the available evidence must be relied on to gauge or infer change. Whether an intervention changes a construct cannot be assumed on faith. Inconsistent findings due, in part, to the poor psychometric properties of measures employed in a study or inadequate power may not reflect anything about an intervention's ability to change a construct. However, they may play a key role in how one interprets the available evidence. The RPC Model was developed to conceptualize, classify, and examine change, based on available evidence. Therefore, the framework assumes that the conditions in which evidence is gathered are empirically sound. Like any experimental endeavor, if a researcher is interested in examining an intervention, it is important that extraneous factors or influences be ruled out in their study of the intervention (e.g., reliable and valid measures of relevant constructs are carefully chosen, plans for achieving adequate statistical power are drawn). In this way, the evidence will likely reflect the ability (or absence thereof) of the intervention to change the targeted construct, and not reflective of poor study design or selection of poor measures.

Multiple Classifications

In classifying studies under the RPC Model, one possible scenario may be that, for any one dimension, classifications of change may seemingly fall under multiple categories. We emphasize that such classifications should be limited to those studies that find multiple specificities of effects, or combinations of informant-, measure-, and/or method-specific evidence. We emphasize that a study cannot be classified under categories denoting "Best Evidence for Change," "Evidence for Probable Change," or "Limited Evidence for Change" if clear patterns of significant effects (80%) are found for specific informants or sources, measures, and/or methods, and limited evidence (50% or less) is found using other informants, measures, and/or methods. Conversely, a study cannot be classified under categories denoting specificity of effects simply because effects were found only under specific circumstances. To classify a study as suggesting specificity of change, a clear pattern of significant findings (80%) must be observed for specific informants, measures, and/or methods (e.g., criteria met for classifications in "Evidence for Contextual- or Informant-Specific Change" and/or "Evidence for Measure- or Method-Specific Change" categories). Thus, although studies may at times seem to meet criteria for a number of categories, we argue that multiple classifications ought to be reserved for those instances in which studies show multiple specificities of effects (see Table 5).

Comparisons With Already Established Interventions

One final consideration concerns instances in which research examines an intervention by comparing it to another intervention already identified as being able to change the targeted construct. The RPC Model categories, as currently conceptualized, inform the classification of change when comparing an intervention to a control condition, or an alternative or unestablished intervention (e.g., intervention not identified by a given system as changing the construct). Systems that identify evidence-based interventions often allow for the identification of an intervention if a well-conducted between-groups experiment finds that the intervention is at least equivalent to an already established intervention. Under these circumstances, the RPC Model may be used to classify within- and between-study consistencies in findings suggesting the equivalence of an intervention, relative to an already established intervention. For instance, the RPC Model categories can be modified, such that the threshold for findings is set to equivalency between conditions. Change would be suggested by indicators that find the two interventions are at least equivalent (e.g., findings suggest nonsignificant differences between conditions). Furthermore, if there are differences, the intervention of interest in the study should be suggested to make changes that are superior to those made by the already established intervention.

Illustration of the RPC Model

An illustration of the RPC Model using the findings from the studies reviewed previously is presented in Tables 6–9. For this illustration, we calculated effect sizes for all indicators and methods of examination available. This was done to provide as conservative an illustration of the RPC Model as possible. The RPCs presented in Tables 6–9 were constructed by placing the lowest observed effect at the lower end of the interval, and placing the

highest observed effect at the upper end. An RPC was constructed in each case where there were at least three findings to include in the range. For example, informant-specific RPCs and classifications of change could be made in instances in which at least three measures from that informant could be included in the range. The percentage of statistically significant effects was reported for the effects within each RPC. For each study, an overall RPC was presented that denotes observed effects found across indicators and methods of examination. In addition, RPCs were constructed for each dimension of indicators employed in the study (e.g., symptom/diagnostic presentation of the target construct, symptom/diagnostic presentation of comorbid constructs). The classification of change using the RPC Model's categories was given for each dimension, and an interpretation was provided of the strength of the effects observed, consistent with *d* effect size conventions (e.g., Cohen, 1988; Rosenthal & DiMatteo, 2001).

Method

Calculations of effect sizes across mean differences, diagnostic status, and clinically significant change methods were conducted consistent with prior work (e.g., Cohen, Cohen, West, & Aiken, 2003; Rosenthal & DiMatteo, 2001). Calculations of effect sizes for tests of responder status and high-endstate functioning were not calculated for studies of relaxation for adult generalized anxiety disorder. In the original studies, these tests did not directly compare relaxation and control conditions, and insufficient information was provided to calculate such effects. Specifically, in cases in which comparisons could be made between relaxation and controls on outcomes gauged via tests of responder status and high-endstate functioning, percentages of participants were the only statistics reported. In these cases, actual frequencies of participants were necessary to calculate effect sizes, given that not all participants completed the intervention, and some that did complete the intervention did not provide data on outcomes. This made it unclear as to how many participants were examined in these tests. Similarly, calculations of effect sizes for diagnostic status findings in Kendall (1994), as well as for clinically significant change findings in Barrett et al. (1996), could not be conducted because only percentages of participants were reported. Nevertheless, the broader point of the illustration is that change is highly variable and multidimensional (see Tables 6 through 9). These effects, although not calculated, would not detract from the general findings or overall points of the illustration.

Effect Size Metric and Characteristics

Effect sizes in the illustration are both presented and interpreted via effect size conventions of small (.2), medium (.5), and large (.8) *d* metric effects (Cohen, 1988). Positive effect sizes denote the intervention condition had better outcomes, relative to the control condition, and negative effect sizes denote the intervention condition had worse outcomes. Each RPC was constructed using all relevant observed effects, regardless of whether such effects were positive or negative. Although negative effects of interventions are obviously outside the scope of the hypothesized effects of interventions examined in research (i.e., intervention researchers are interested in examining interventions that produce favorable outcomes), we emphasize that it is important to examine the range of possible effects that an intervention may have. This issue is especially critical, given prior work suggesting the possibility that interventions intended to produce favorable outcomes may actually, at times produce unfavorable outcomes (see Lambert & Ogles, 2004). Furthermore, including negative effects in RPCs is consistent with how intervention research is conducted, given that although hypotheses in intervention research are often directional, statistical tests conducted in such research are by default, nondirectional.

Results of statistical analyses were derived from the results of statistical tests reported in the original articles. Specifically, for a given indicator of change, any statistically significant difference between intervention and control conditions that was reported in the original study was counted as a significant difference between conditions on that indicator (e.g., main effect of treatment, treatment × time interaction). However, sometimes the results of statistical analyses were not reported in the article, or the result was stated, but the relevant statistics were not available. In these instances, effect sizes and statistical significance were estimated using reported means and standard deviations (i.e., mean differences) or frequencies (i.e., diagnostic status, clinically significant change) using suggested procedures (Cohen et al., 2003; Rosenthal & DiMatteo, 2001). When a study reported nonsignificant differences, and did not provide relevant statistics, effect sizes were conservatively set at 0. In cases where a dimension or category of method of analysis had fewer than three findings, an RPC was not constructed; these cases were denoted as N/A.

Effect Size Calculations

Calculations of effect sizes for mean differences between conditions were made by subtracting the control group mean from the intervention group mean, and dividing this difference by the control group's standard deviation at outcome (Glass's Δ; Rosenthal & DiMatteo, 2001). Glass's Δ is an effect size metric for which meta-analytic theorists and researchers consider being within the *d* family metric of effect sizes (Rosenthal & DiMatteo, 2001). Thus, to maintain a consistent presentation of effects across methods of analysis, results using Glass's Δ are presented using the *d* symbol. Our examples of intervention research were derived from the efficacy literature examining interventions for both children and adult generalized anxiety disorder. Thus, we conducted calculations of mean differences effect sizes consistent with prior meta-analytic work of both child intervention research (e.g., Weisz, Weiss, Alicke, & Klotz, 1987; Weisz, Weiss, Han, Granger, & Morton, 1995) and research examining interventions for adult generalized anxiety disorder (Gould, Otto, Pollack, & Yap, 1997).

Effect sizes for analyses of diagnostic status in Barrett et al. (1996) were calculated using the phi (Φ) coefficient to examine differences in proportions between conditions (see Cohen et al., 2003). For clinically significant change findings in Kendall (1994), only results of statistical tests were available (*t* statistic). Thus, effect sizes (using the *r* metric) were estimated using test statistics, as suggested elsewhere (Rosenthal & DiMatteo, 2001). Lastly, given the use of *r* effect size measures for some calculations, and that Φ is an *r* effect size measure as well, effect sizes calculated using Φ and *r* were converted to *d*, in order to construct RPCs along a common effect size metric (Rosenthal & DiMatteo, 2001).

Results

Individual Cognitive–Behavioral Treatment for Childhood Anxiety Disorders

Findings interpreted using the RPC Model present a different picture of the evidence for change (see Tables 6 and 7). For each study examining individual cognitive–behavioral treatment for childhood anxiety disorders, we present findings across and within multiple dimensions, and illustrate how the model highlights the substantial variability across indicators. The test conducted by Kendall (1994) of the efficacy of individual cognitive–behavioral treatment for childhood anxiety disorders suggests that change within this intervention is multidimensional (see Table 6). Across all dimensions, effect sizes ranged from significant and moderately detrimental ($d = -.45$) to significant and positively large ($d = 2.14$). Moreover, even when separate RPCs were constructed for two separate dimensions (i.e., symptom/diagnostic presentation of the target construct, symptom presentation of comorbid constructs/associated features of the target construct), the only consistency

across dimensions was that changes were largely specific to questionnaire-rated information provided by the child, examined via mean differences between conditions. Most importantly, the RPCs for both of these dimensions suggested wide variability (see Table 6).

Consistent with the RPC Model, the test conducted by Barrett et al. (1996) of the efficacy of individual cognitive–behavioral treatment for childhood anxiety disorders suggests, again, that the efficacy of individual cognitive–behavioral treatment for childhood anxiety disorders is highly variable and multidimensional (see Table 7). Across all dimensions, effect sizes ranged from nonsignificant and positively small ($d = .10$) to significant and positively large ($d = .91$). The only RPC that could be constructed was for the symptom/diagnostic presentation of the target construct. Interestingly, the RPC for this dimension suggested limited evidence for change. Indeed, a key observation to make in the RPC Model classification of changes in the symptom/diagnostic presentation of the targeted construct is that change in this dimension was classified into different categories for Kendall (1994) and Barrett et al. (1996). The differential classification of this dimension between these two studies is surprising, given that many of the indicators employed across studies were identical (cf. Tables 6 and 7). Thus, between-study inconsistencies were found across classifications for Kendall (1994) and Barrett et al. (1996).

Relaxation for Adult Generalized Anxiety Disorder

An illustration of the RPC Model using the studies of relaxation for adult generalized anxiety disorder summarized previously highlights the within- and between-study inconsistencies in change (see Tables 8 and 9).[4] The test conducted by Barlow et al. (1992) of relaxation for adult generalized anxiety disorder highlights the variable and multidimensional nature of change (see Table 8). Overall, indicators ranged from nonsignificant and negligibly negative ($d = -.09$; relaxation had higher but nonsignificantly different pretreatment scores on this variable than controls) to significant and positively large ($d = 2.00$). Interestingly, similarly wide ranges were found for indicators of the dimension representing changes in the symptom presentation of the target construct. The RPC for this dimension was classified as providing limited evidence for change. The RPC representing changes in the symptom presentation of comorbid constructs/associated features of the target construct also showed wide variability.

Lastly, the test conducted by Butler et al. (1991) of the efficacy for relaxation for adult generalized anxiety disorder once again suggests that the change construct for this intervention is multidimensional and highly variable (see Table 9). Overall, indicators ranged from nonsignificant and small ($d = .06$) to significant and positively large ($d = 1.06$). Wide ranges in change were found for indicators encompassing the dimension representing change in the symptom presentation of the target construct. Moreover, similar to Barlow et al. (1992), the RPC for this dimension was classified as providing limited evidence, although strikingly fewer significant effects were found (cf. Tables 8 and 9). Indeed, fewer than 20% of the indicators suggested the intervention changed the symptom presentation of the target construct. Further, of the two indicators of change in the target construct that Barlow et al. (1992) and Butler et al. (1991) shared in common, the one participant self-rated measure did not suggest change in either investigation. The other indicator, an independent assessor-rated measure of the tar-

get construct, suggested strikingly different changes in each investigation (cf. Tables 8 and 9). This indicator suggested the lowest changes ($d = .08$) in Butler et al. (1991), and suggested one of the highest changes ($d = 1.85$) in Barlow et al. (1992). Thus, between-study inconsistencies predicted by the RPC Model were found across classifications of change in these studies.

General Discussion of the Illustration of the RPC Model

Our illustration of the RPC Model suggests that intervention change is highly variable and multidimensional. That change is variable is suggested by within- and between-study inconsistencies suggested by different indicators, and different methods of analysis. That change is multidimensional is suggested by the variability in changes observed both within and between different dimensions of change. Most importantly, variability in change was evident, particularly among indicators encompassing the same dimension. Wide variability was often observed among indicators of the specific construct targeted for change. Moreover, this variability was observed both within and between investigations, even when identical measures were employed across investigations. Therefore, the illustration provides evidence for the utility of the RPC Model for conceptualizing change, and classifying evidence within and between studies.

The RPC Model's Relationship to Other Approaches

Selecting a Primary Outcome Measure in Advance of the Study

The RPC Model can be contrasted with other approaches commonly employed to address inconsistent findings. For instance, researchers commonly select a single indicator in advance of the investigation, and designate it as a "primary outcome measure" (e.g., Bowden et al., 2000; Hayward et al., 2000; Hazell & Stuart, 2003; Michelson et al., 2004; Wigal et al., 2004). In investigations that utilize this approach, often times other measures are employed as well; these measures are often deemed "secondary outcome measures." One could argue that this is a defensible approach to examining change in an intervention, because the primary indicator is identified a priori. Further, often the primary outcome measure is the only indicator employed that assesses the targeted construct.

As mentioned previously, constructs targeted for change are multidimensional and highly variable. As such, it is difficult to argue that adequate examinations of change in interventions targeting such complex constructs can be captured with a single indicator. Indeed, evidence presented previously suggests wide within-study variability in findings, particularly among multiple

---

[4] Although nonsignificant effect sizes reported in Tables 8 and 9 are larger than some of the smaller effect sizes showing significant effects, it must be noted that analyses of covariance (ANCOVAs) were conducted in these studies, using pretreatment variables as covariates. Thus, it is likely that, for some of the nonsignificant effect sizes noted in the study, although they were sometimes larger than other smaller significant effects, the variance explained by preintervention scores may have contributed to postintervention differences, leading to nonsignificant results in these larger effect sizes (see Table 2 in Barlow et al., 1992; Tables 2 and 3 in Butler et al., 1991).

indicators of the targeted construct. Furthermore, this evidence suggests wide variability between investigations of the same intervention, even on identical indicators.

The RPC Model highlights how difficult it is to accumulate knowledge of change based on single or primary indicators. This difficulty arises particularly in instances in which a "primary" indicator suggests change, whereas indicators representing identical dimensions as the "primary" indicator either do not suggest change, or suggest change of varying magnitudes. A key strength of the RPC Model is that the multidimensional and variable nature of change is acknowledged and accounted for when conceptualizing, examining, and classifying change.

Traditional "and/or" Rules

Another approach commonly used to address inconsistent findings is the employment of "and/or" rules. For instance, a participant can be identified as benefiting from an intervention if outcome measures from all informants suggest that the participant does not meet diagnostic criteria for the targeted construct (i.e., "and" rule; Bird, Gould, & Staghezza, 1992; Piacentini, Cohen, & Cohen, 1992). Conversely, using the "or" rule, a participant can be identified as benefiting from an intervention if an outcome measure from any single informant suggests that the participant does not meet diagnostic criteria for the targeted construct.

Prior work suggests that "and/or" rules may be problematic. First, using "and/or" rules may not be more incrementally reliable than interpreting the information from multiple informants' ratings independent from one another (e.g., Offord et al., 1996). In addition, prior work has found that using these decision rules can mask associated features of the construct that are identified when informants' ratings of the construct are considered independently (e.g., Offord et al., 1996; Rubio-Stipec, Fitzmaurice, Murphy, & Walker, 2003). This suggests that identifying change via "and/or" rules may hinder efforts to subsequently identify moderators or predictors of change. Moreover, prior work suggests that "and/or" rules may lead to either overestimated ("or" rule) or underestimated ("and" rule) rates of change (Piancentini et al., 1992; Youngstrom, Findling, & Calabrese, 2003). This is because these methods may fail to correctly identify participants who either do or do not meet diagnostic criteria for the targeted construct.

Perhaps most importantly, "and/or" decision rules do not acknowledge the multidimensional nature of change. Indeed, these rules do not take into consideration that inconsistent findings may represent instances in which change occurs, and instances in which change does not occur. For instance, inconsistent findings among different sources of indicators (self-perceived vs. spouse-perceived symptoms of social phobia) may very well suggest that the intervention changes one source's (self) perceptions, but not another source's (spouse) perceptions. Thus, use of these rules may lead to a loss of information of the instances in which interventions bring about change. The issues that make use of "and/or" decision rules problematic are accounted for by the RPC Model, in that inconsistencies are not so much resolved as they are utilized to understand the instances in which an investigation both suggests and does not suggest change.

High-Endstate Functioning

Another approach to resolving inconsistencies is referred to as high-endstate functioning (for a review, see Hill & Lambert,

2004). This approach evaluates change without reference to pre-intervention functioning. Interestingly, high-endstate functioning can take a number of different forms. For instance, the approach can gauge change, relative to normative functioning. High-endstate functioning can also be employed by selecting multiple indicators, and setting a priori thresholds of change for each indicator. If a participant passes the set thresholds on all measures, that participant has achieved high-endstate functioning (Hill & Lambert, 2004).

The RPC Model does not dispute the various principles of high-endstate functioning (e.g., gauging change, relative to normative functioning; use of multiple indicators). At the same time, high-endstate functioning is one of a number of methods of examining change, and as noted previously, findings derived from multiple methods of measurement and analysis are often highly variable. Moreover, the RPC Model highlights the reality that multiple indicators are highly variable. As a result, conceptualizing and classifying change solely through a method by which multiple measures have to all suggest change may be potentially problematic. A key strength of the RPC Model is that it acknowledges that there are varieties of approaches or methods by which to examine change, and different methods may lead to different conclusions. The RPC Model highlights that use of a sole method is not a defensible approach to examining change, and thus both encourages and accounts for use of multiple methods.

Averaging Effects Across All Indicators in a Study

Another method to address inconsistent findings is the averaging of effects observed across indicators, an approach often used in meta-analytic research (see Matt, 1989 for a review). Measures of central tendency are not used within the RPC Model because the purpose of conceptualizing and classifying change is, in part, to take into account consistency in findings. Relying solely on measures of the average effects of interventions may mask the wide variability of change found both within and between investigations of interventions, thus hindering the ability to note consistencies and inconsistencies in the evidence for change. Thus, the RPC Model focuses on the consistency by which indicators suggest change, and not on attaining an estimate of central tendency of change. The RPC Model is consistent with a primary aim of intervention research: replication of findings made across controlled experiments (e.g., Lambert & Ogles, 2004). Indeed, the RPC Model simply extends this primary aim of intervention research a step further, by taking into account whether findings of change replicate or are consistent across multiple findings made both within and between controlled experiments.

Mixing and Matching Informants' Contexts and Perspectives

The RPC Model can be compared and contrasted with another approach to handling inconsistent findings (Kraemer et al., 2003). The Kraemer et al. (2003) approach assumes inconsistencies among informants' ratings can be integrated to form a single indicator, by mixing and matching ratings based on the discrepant perspectives and contexts by which informants provide information. For instance, the method posits that a single reliable indicator of child psychopathology can be attained by first gathering child ratings (provided from a self-perspective in both nonhome and

home contexts), parent ratings (provided from an other- or observer-perspective in a home context), and teacher ratings (provided from an other- or observer-perspective in a nonhome context). Once the ratings are gathered, a principal-component analysis can be performed to extract shared variance among the ratings. This shared variance may then be attributed to three components: (a) trait or characteristic of interest (e.g., child psychopathology); (b) perspective by which each informant observes the characteristic (i.e., teacher and parent are observers of the child's behavior, child is providing information of him- or herself); and (c) context in which each informant observes the characteristic (i.e., teacher observes the child's behavior in a nonhome context, parent in the home context, and child in both home and nonhome contexts). The "trait" variance explained among informants' ratings becomes an indicator that may approximate a "gold standard" indicator of the construct.

The RPC Model does not dispute the underlying assumption of the Kraemer et al. (2003) approach: It is important to both gather and use multiple indicators to assess constructs. Indeed, this practice is one of many critical practices one must undertake when conducting intervention research. At the same time, there are two reasons why the RPC Model conceptualizes change along a range of possible changes. First, the RPC Model acknowledges that, in intervention research, indicators derived from informants' perceptions of the participant's behavior in different contexts are but one method to gauge change. As mentioned previously, measurement methods other than informants' ratings, such as performance-based measures, indicators of physiological functioning, and institutional records are often used in intervention research. These methods often measure constructs in ways that are independent of the contexts in which constructs are expressed or exhibited, or the perspectives by which people in participants' lives observe these constructs. Thus, it would be difficult to combine indicators that may be context- and perspective-independent with other indicators that are context- and perspective-dependent, particularly when using approaches that attempt to integrate them into a single measure. Further, indicators such as lab-based and performance-based measures are often both integral in examining change, and result in findings that are discrepant from those made with informants' ratings. A strength of the RPC Model is its incorporation of multiple indicators, regardless of whether they are derived from methods that are context- and perspective-dependent.

Second, because of the difficulty in gathering information from multiple contexts, informants are often assessed for their perceptions of the participant's behavior in contexts or situations for which they have very little or no access. For example, current measures of child psychopathology (e.g., Child Behavior Checklist [Achenbach, 1991]; Diagnostic Interview Schedule for Children [Shaffer et al., 1996]) often assess for parents' ratings of children's behavior in contexts other than in the home (e.g., school, peer interactions). It is quite likely that parents use information they have received from other informants in order to provide information for items that assess the nature and extent of these behaviors (e.g., teachers calling home to inform parents that their child is misbehaving at school). Therefore, informants that presumably provide information of behavior from a specific perspective and observed context are often asked to provide information of behavior when it is exhibited or expressed outside of the confines of that specific perspective and context. As a result, when informants are assessed by measures that ask them to rate constructs for which

they do not have access, they may rely on information provided informally to them by other informants. Thus, the RPC Model does not attempt to integrate information from multiple informants into a single indicator, because in intervention research the perspectives and contexts by which sources provide information are often blurred, and seldom mutually exclusive.

Implications of the RPC Model

Conceptual Implications

Research and theory on interventions in psychological science. The RPC Model has implications for conceptualizing, examining, and classifying interventions that purport to change multidimensional constructs. Much of this article discussed and illustrated the RPC Model in the context of research examining interventions targeting psychological disorders because the change construct is critically important to this area of psychology. As mentioned previously, it is readily acknowledged within clinical science that the constructs targeted for change are often multidimensional. There is clear evidence to suggest that the intervention change construct within clinical science also ought to be conceptualized as multidimensional. Furthermore, intervention research within clinical science is a fairly broad area of psychology, interested in ameliorating a broad range of psychological problems. Therefore, clinical science is an area that is generally focused on applying psychological principles and theory to the study of change, and is a prime candidate through which to propose and illustrate the RPC Model.

At the same time, we argue that the RPC Model would generally apply to interventions purporting to change constructs in psychological science. Indeed, recent conceptualizations of a variety of constructs purporting to explain specific normative behaviors acknowledge that such constructs are not one-dimensional, but encompass a variety of components, processes, and mechanisms (biases in group judgments, false-memory editing, self-enhancement bias, skill learning; Brainerd, Reyna, Wright, & Mojardin, 2003; Kwan, John, Kenny, Bond, & Robins, 2004; Sun, Slusarz, & Terry, 2005; Van Rooy, Van Overwalle, Vanhoomissen, Labiouse, & French, 2003). It is likely that change within interventions targeting these constructs would have to be conceptualized as multidimensional as well. Therefore, we encourage applied research to use the RPC Model when conceptualizing, examining, and classifying change.

Basic research. The RPC Model has implications for basic research and theory on both normative and abnormal psychological processes. Many of the issues raised and addressed in this paper relate directly to research and theory in the applied psychological sciences. However, when discussing broad issues relating directly to either the applied or basic domains of psychological science, it is helpful to note that these domains of research often have reciprocal impacts on one another. An example may be formative. Consider research and theory on childhood aggression and antisocial behavior. Highly influential basic research in this area examined parenting and family dysfunction factors and their relation to aggression and antisocial behavior in children. This seminal research identified specific factors that may pose risk for aggression in children (e.g., authoritarian parenting, corporal punishment, positive reinforcement of negative behaviors; Patterson, 1982; Reid, Patterson, & Snyder, 2002). Identification of these factors gave rise to the idea that aggression and antisocial behavior may be

reduced by changing these factors. As a result, interventions were designed to target maladaptive parenting and family dysfunction (e.g., parent management training; Kazdin, 2005). These interventions are some of the most rigorously studied interventions in psychological science, and multiple studies suggest they ameliorate aggression in children (e.g., Kazdin, 2005).

This brief example highlights how carefully conducted basic research and theory informs applied research and theory. However, as mentioned previously, the influence that these domains may have on one another is reciprocal. For instance, identifying a number of risk or causal factors to target in parent management training may perhaps assume that, in the absence of intervention, these factors may be playing some role in the development and/or maintenance of aggressive and antisocial behavior in children. In addition, assume that, under the RPC Model, future research examining parent management training finds that, immediately after its administration, the intervention reduces both the symptom and diagnostic presentation of childhood aggression, and a dimension representing authoritarian parenting.

Under the RPC Model, follow-up or longitudinal examinations of these two dimensions may examine whether recurrence of problems in these dimensions are related to one another. This research could provide an invaluable resource for both generating hypotheses and building theory that ultimately inform basic research on the longitudinal processes underlying continuity and discontinuity of childhood aggression. An example may be a long-term investigation of parent management training that finds that relapse or recurrence in the symptom and diagnostic presentation of childhood aggression is related to relapse in the authoritarian parenting dimension (i.e., authoritarian parenting returns, and that is related to increases in aggressive symptoms and behavior disorder diagnoses). Such an investigation may inform additional basic research on the relapse or recurrence of aggression in children over the course of development, leading to an increased understanding of how and why aggression in children may increase and decrease over time. Thus, although the RPC Model we propose most directly informs applied psychological science, we must emphasize that such research may ultimately inform basic research on the same psychological processes intervened upon in applied psychological science.

Accommodating for modifications to the change construct. The RPC Model allows for future modifications in conceptualizations of change, particularly in accordance with developments in research and theory on constructs. The RPC Model was developed to take into account the multidimensional nature of constructs. Needless to say, the multiple dimensions noted and illustrated both within this paper and within the RPC Model (see Figure 1) are not meant to be exhaustive. On the contrary, dimensions of constructs ought to be examined in light of the dimensions of change of interest. The RPC Model was developed to guide the study of change within interventions that target any dimension or dimensions of constructs. In this manner, the RPC Model was developed to take into account accumulations of knowledge in the understanding of constructs, particularly if research identifies additional dimensions that are critical to target. We encourage use of the RPC Model in the future study of change, particularly future research that identifies dimensions that are most critical to target in intervention research.

Informing existing classification systems. The RPC Model may inform existing systems. The framework provides useful categories by which to classify change within and between studies (see Table 5). Thus, existing systems ought to use the RPC Model to inform future efforts in classification. For instance, existing systems can use the RPC Model by, first, identifying the construct of interest, along with an intervention developed to change this construct. A dimension of the construct can be identified, along with the corresponding dimension of change. Once the dimensions of the construct and change are identified, an acceptable or threshold RPC Model classification category can be identified. Under this RPC Model category, a range of change can be identified that would be acceptable for a given study to be counted as providing evidence for change. A review of the literature can then be conducted to examine whether investigations examining that intervention meet or surpass these thresholds.

An example using the TFPP guidelines may be helpful. Under TFPP, an intervention targeting the *DSM–IV* symptom presentation of childhood social phobia can be classified under the *probably efficacious* category if: (a) at least one study found evidence for the intervention's moderate-to-large change, relative to a control condition; and (b) the evidence is classified either in the RPC Model's "Best Evidence for Change" or "Evidence for Probable Change" categories (see Table 5). This example is not a recommendation; it is simply an illustration. Although the criteria used may change markedly from system to system, the point is that the RPC Model can inform future efforts in employing existing systems to classify interventions.

Research Implications

Informing study design.   The RPC Model has implications for future work examining change. For example, the RPC Model can inform the design of future studies that conceptualize, examine, and classify change. Figure 3 outlines how the RPC Model can inform study design. Broadly, the RPC Model highlights the need to take various issues into account when designing a study. First, once the construct targeted for intervention is identified, it is important to also identify which dimension or dimensions of the construct are going to be targeted for change. Indeed, as mentioned previously, some dimensions of constructs may be more important to target for intervention than other dimensions. Deciding which dimensions to target may depend on such factors as the type of intervention, the point in time that the intervention is evaluated, and the construct being targeted (see Table 10). Thus, if the targeted construct is, for example, adulthood major depressive disorder, and knowledge of the intervention being examined to target this construct is still at a preliminary stage, the dimension to target may simply be the symptom presentation of the construct, and change may be evaluated immediately postintervention.

Second, once the dimension or dimensions of the construct are identified, it is important to identify which dimension or dimensions of change will be conceptualized, examined, and classified. Key in the selection of the dimension or dimensions of change is that they align with the dimension or dimensions of the targeted construct (e.g., selection of the symptom presentation of major depressive disorder as the dimension targeted for change = studying the symptom presentation dimension of change). Third, once the dimension or dimensions of change are identified, a critical issue involves the hypothesizing and conceptualizing of the instances in which change will occur. In essence, the RPC Model highlights the importance of placing change in a proper conceptual
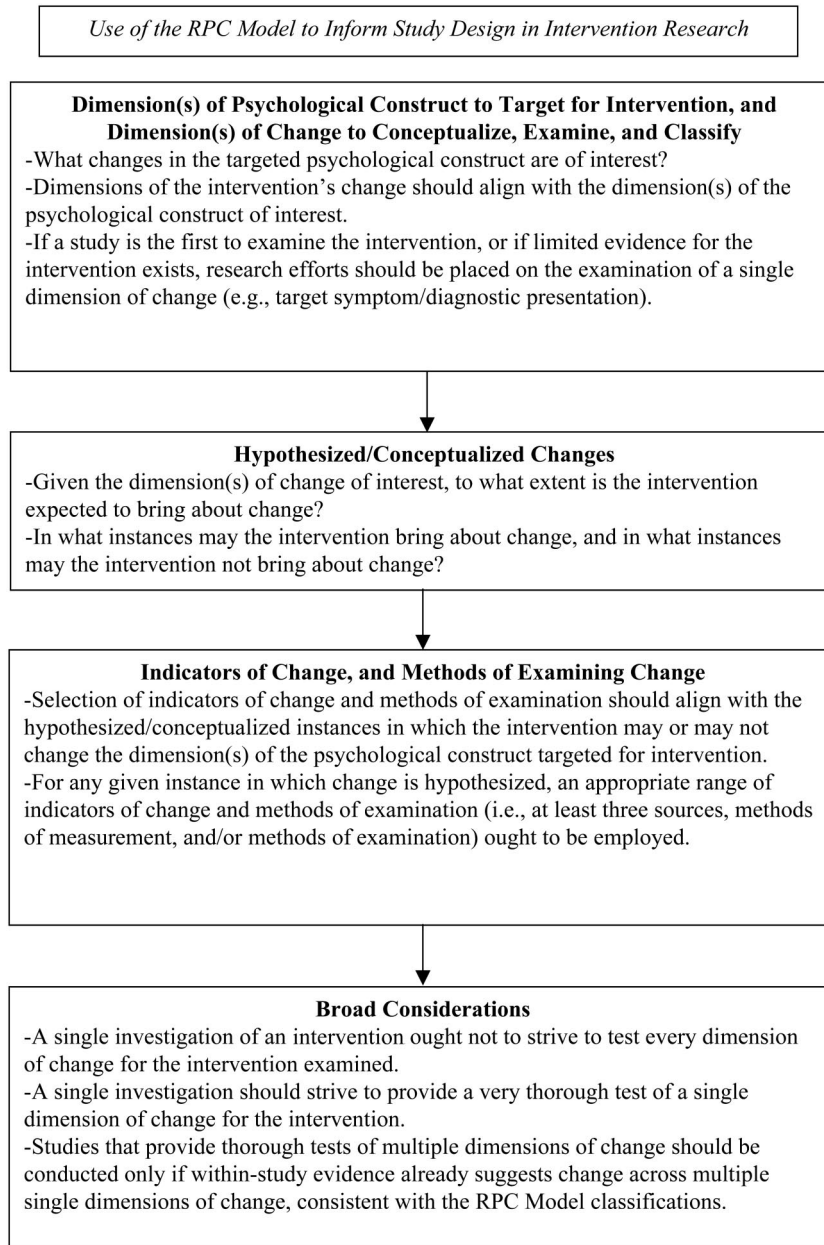
| Use of the RPC Model to Inform Study Design in Intervention Research |

Dimension(s) of Psychological Construct to Target for Intervention, and Dimension(s) of Change to Conceptualize, Examine, and Classify
-What changes in the targeted psychological construct are of interest?
-Dimensions of the intervention's change should align with the dimension(s) of the psychological construct of interest.
-If a study is the first to examine the intervention, or if limited evidence for the intervention exists, research efforts should be placed on the examination of a single dimension of change (e.g., target symptom/diagnostic presentation).

Hypothesized/Conceptualized Changes
-Given the dimension(s) of change of interest, to what extent is the intervention expected to bring about change?
-In what instances may the intervention bring about change, and in what instances may the intervention not bring about change?

Indicators of Change, and Methods of Examining Change
-Selection of indicators of change and methods of examination should align with the hypothesized/conceptualized instances in which the intervention may or may not change the dimension(s) of the psychological construct targeted for intervention.
-For any given instance in which change is hypothesized, an appropriate range of indicators of change and methods of examination (i.e., at least three sources, methods of measurement, and/or methods of examination) ought to be employed.

Broad Considerations
-A single investigation of an intervention ought not to strive to test every dimension of change for the intervention examined.
-A single investigation should strive to provide a very thorough test of a single dimension of change for the intervention.
-Studies that provide thorough tests of multiple dimensions of change should be conducted only if within-study evidence already suggests change across multiple single dimensions of change, consistent with the RPC Model classifications.

Figure 3. Conceptualization of how the Range of Possible Changes (RPC) Model informs study design. Key components are: (a) the dimension(s) of both the constructs targeted for intervention, and, relatedly, the dimension(s) of intervention change that will be conceptualized, examined, and classified; (b) the hypothesized or conceptualized changes of the intervention; and (c) the selection of both indicators and methods of analysis. Broad considerations of study design are briefly highlighted.

context, which entails acknowledging that the intervention will likely bring about change in some instances but not others. For example, a researcher may conceptualize change in a short-term intervention targeting the symptom presentation dimension of adulthood major depressive disorder as such that the researcher expects changes in self-perceived target symptom presentation, but not in self-perceived target diagnostic presentation. The researcher may anticipate that, because knowledge of the intervention is still in its infancy, the duration of the intervention may still need to be properly studied and calibrated. Thus, as tested in the study, the

researcher may conceptualize the intervention as limited to reducing participants' self-perceived symptom presentations of the targeted construct. However, such reductions may not be large enough to change participants' self-perceived diagnostic recoveries, let alone other sources' perceptions of symptom presentations.

Fourth, the selection of indicators and methods of analysis should be commensurate with the a priori conceptualization of change. The methodological and statistical design of the study should correspond with the conceptualized instances in which the intervention may change the dimension or dimensions of the

targeted construct. Key in this component of study design is the idea that, for any given instance of hypothesized change, an appropriate range of indicators and methods of examination ought to be employed. To return to our previous example, if the researcher conceptualizes change as changes in participants' self-perceived symptom presentations, the researcher may employ three methodologically distinct measures of self-perceived symptoms of major depressive disorder. Further, the researcher may examine changes via mean differences, and two different methods of clinically significant change (e.g., high-endstate functioning, reliable change index; Hill & Lambert, 2004; Jacobson & Truax, 1991).

Perhaps most importantly, the RPC Model highlights the importance of not making a single investigation a test of every dimension of change within the intervention. Such a test will likely not lead to a thorough understanding of any one of the multiple dimensions by which an intervention may or may not bring about change. Instead, such a test may lead to, at best, an inconclusive or limited understanding of change. We propose that a single study ought to strive to provide a very thorough test of a single dimension. Within that dimension, a single study ought to conceptualize under what instances or circumstances the study will identify change. Thorough tests of multiple dimensions should be conducted only if prior within-study examinations of that intervention already suggest change across multiple single dimensions.

Emphasizing that an investigation ought to strive to provide a thorough test of a single dimension may appear an odd recommendation. Indeed, one could argue that this is impractical, given that intervention studies, particularly for psychological problems and disorders, often require numerous years to complete and massive amounts of funding. We argue that, in trying to examine a plurality of dimensions within an investigation, researchers may risk finding, at best, inconsistent evidence across these numerous dimensions, and weak evidence for any one dimension. Alternatively, we argue that, if resources within an investigation are specifically placed on a single dimension, researchers are far more likely to find consistent, strong evidence for a specific instance in which the intervention may change the construct.

The overarching point is that the RPC Model highlights a key consideration. Namely, intervention research ought not to try to examine too much within any given investigation. For instance, if a researcher is interested in studying an intervention that has not been examined, then the initial study of that intervention ought to focus squarely on one dimension of change. After this study is conducted, and if sound evidence is found for the dimension, then research can proceed to examining other dimensions of change for that intervention. We encourage future work to use the RPC Model to inform study design, and aid in the accumulation of knowledge of change within specific interventions, over the course of multiple investigations.

Classifications of limited change within interventions. Use of the RPC Model may lead to findings gathered via methodologically rigorous and well-controlled studies being classified as providing limited or no evidence for change within interventions. If a study's evidence for an intervention is classified as limited or nonexistent, then this classification may potentially hinder the likelihood that the study is considered for publication. It is imperative that future research incorporating the RPC Model be evaluated as much for its methodological rigor as for its findings, so that studies reporting evidence that is classified as providing limited or

no evidence be included in the published literature. Such empirical work is as critical to our understanding of change as empirical work that finds moderate or strong evidence.

The critical importance of including studies reporting limited or no evidence for change is highlighted by intervention research in the general medical sciences. In this area, well-conducted replication studies of prior intervention research are often published, regardless of whether null or significant results are found. This research is critical to the understanding of intervention change in this area of science. Indeed, often times initial studies finding positive effects of interventions are subsequently contradicted by further well-conducted research on these same interventions; subsequent research often finds no evidence for positive intervention effects, or even evidence of significantly negative intervention effects (e.g., Ioannidis, 2005; Ioannidis & Lau, 2001; Ioannidis & Trikalinos, 2005; Trikalinos et al., 2004).

It is likely that if replication studies in the published psychological intervention literatures were as common as they are in published literatures of the general medical sciences, these same inconsistencies may be observed. Thus, all well-conducted intervention studies, regardless of their RPC Model classification, speak to the wide variability of change. This knowledge of the variability of change is critical in gaining an understanding of change and how it can be conceptualized, examined, and classified. It is important that well-conducted studies be included in the published literature, even if these studies find limited or nonexistent evidence for change.

Meta-analytic reviews of intervention research. Similar to the RPC Model, prior meta-analytic work suggests wide variability in the findings of intervention research, depending on the indicator. For instance, one of the first meta-analyses of child intervention research found average effect sizes between intervention and control conditions ranged from .16 to 1.14 (Casey & Berman, 1985); a later meta-analysis found average effect sizes between intervention and control conditions ranged from .33 to 1.08, depending on the informant or source (Weisz et al., 1987).[5] Indeed, these discrepant findings have led researchers to conclude that the outcomes of children's interventions, and thus their available evidence base, may differ depending on the informant (Weisz et al., 1995). Moreover, meta-analytic work examining adult interventions suggests the same inconsistencies. For example, average pre-to-postintervention effect sizes of interventions for adult agoraphobia range from .44 to 2.66, depending, in part, on the source (e.g., clinician, self, performance-based, physiological; Ogles, Lambert, Weight, & Payne, 1990). Further, outcomes for interventions of adult depression differ markedly, depending on whether the source was the clinician or participant (Lambert, Hatch, Kingston, & Edwards, 1986).

Given prior work suggesting these inconsistencies, it is important that future meta-analytic research be conducted, using the RPC Model. Indeed, prior meta-analytic work suggests the average

---

[5] Effect sizes were calculated slightly differently between Casey and Berman (1985) and Weisz et al. (1987), with Casey and Berman (1985) dividing differences between intervention and control conditions by the pooled intervention and control group standard deviation, and Weisz et al. (1987) using the control group standard deviation. Despite these differences in effect size calculation, the general finding of wide variability across findings, depending on the informant was consistent in both studies.

effects of interventions vary widely across different indicators. However, these average effects speak in a limited sense to the variability or multidimensional nature of change. As mentioned previously, meta-analytic reviews often aggregate findings using measures of central tendency; use of measures of central tendency may mask large variability in change.

The variability in indicators of change is a critical issue when interpreting the findings of meta-analytic research. For instance, theorists have noted that a primary aim of intervention research is replication of findings supporting evidence-based interventions (Lambert & Ogles, 2004; Nathan & Gorman, 2002). Indeed, the aim of intervention research is readily present in the efforts of classification systems that identify evidence-based interventions. Inherent in these systems is the notion that in order for an intervention to be considered evidence-based, consistent supportive evidence must be present between at least two studies. Moreover, the goal in both identifying interventions and conducting meta-analytic reviews is identical: Examining whether evidence suggests that an intervention works. Thus, if the aim of meta-analytic reviews is to evaluate whether evidence supports interventions, then the variability and multidimensional nature of change within interventions must be acknowledged.

In line with the need for future meta-analytic work, we recommend future research use the RPC Model to conduct quantitative reviews. Indeed, the RPC Model readily lends itself to conducting such reviews. For instance, our illustration of the RPC Model incorporated the methods by which meta-analytic reviews are conducted. In addition, the RPC Model can be used to classify change both within and between investigations (see Table 5). Change examined within a study of a specific intervention can be treated as the unit of analysis, and effect sizes can be calculated, as illustrated previously (see Tables 6–9). Further, change examined between multiple studies of the same intervention can be treated as the unit of analysis. Classifications of evidence for change gathered within each of multiple investigations can be plotted along the classification categories of the RPC Model. Most importantly, the RPC Model provides a heuristic to conceptualize, examine, and classify change, using methods of quantitative review. We encourage future work to use the RPC Model when conducting such reviews.

Limitations of the RPC Model

The RPC Model raises a number of methodological and conceptual issues that warrant further attention. First, the RPC Model informs the study of change within interventions that target multidimensional constructs. Thus, the utility of the RPC Model may be restricted to examining interventions within which change is likely highly variable and multidimensional. At the same time, this limitation can also be viewed as a key strength of the RPC Model. Indeed, constructs often targeted for intervention are multidimensional. Therefore, researchers ought to commonly expect that interventions may change the multiple dimensions of these constructs in a highly variable fashion. Nevertheless, the limitation of the RPC Model in not informing intervention research targeting one-dimensional constructs may speak to the limits of the RPC Model in informing research in other areas of the behavioral and medical sciences that examine change in one-dimensional constructs (e.g., vision correction for detached retina interventions).

Second, the framework does not inform the study of moderators and mechanisms of change (e.g., Kraemer, Wilson, Fairburn, & Agras, 2002). As such, it remains unclear how the evidence for moderators and mechanisms of change ought to be conceptualized. Conceptualizations of the instances, characteristics, or contexts that qualify or explain change may be particularly challenging, because moderators and mechanisms may be multidimensional as well. Thus, conceptualizations of moderators and mechanisms of change may also need to take into account their multidimensionality and variability, in addition to the multidimensionality and variability of change itself. At the same time, the purpose of this paper was to propose a conceptualization of intervention change. Such a conceptualization should be developed before work conceptualizes moderators and mediators of intervention change. Nevertheless, future work should develop conceptualizations of moderators and mechanisms of change.

Third, we suggest a framework to conceptualize change, within the context of the dimensions often sought for change in intervention research. The issue of whether specific dimensions are more important to change, relative to others, is an empirical issue that warrants further focused attention, and may perhaps be outside the scope of the RPC Model, as currently conceptualized. For instance, we suggest dimensions of change to target, based on such factors as the type of intervention implemented, the time point at which change is examined, and the construct being targeted for change (see Table 10). These dimensions of change logically follow from current intervention research and the constructs targeted for change in such research.

Future work should continue to empirically scrutinize whether targeting the dimensions of change currently targeted in intervention research brings about proper change. By "proper change," we mean change that sufficiently reduces the negative effects and/or enhances the positive effects that targeted constructs have on people affected by them. Most importantly, we emphasize that, in advancing the RPC Model as a conceptualization of change, we encourage researchers to not treat specific dimensions as "gold standard" dimensions of change. An intervention's successfully targeting and changing a single dimension of change within a given construct may not be sufficient to completely, or even adequately address the effects (whether positive or negative) a construct may have on people affected by them. It is likely that, given the complexity of constructs targeted for change, multiple dimensions of change may have to be examined over multiple investigations of an intervention. In this way, psychological science can gain an understanding of the circumstances in which specific interventions bring about change, or whether any one intervention brings about "enough" change in "enough" dimensions for one to conclude that it stands alone as an intervention for a specific multidimensional construct.

Lastly, although we conceptualize intervention change as multidimensional, and both guide and encourage the multidimensional study of change, it remains unclear how many dimensions of change are important to target within an intervention. The dimensions of change we discuss are not meant to be exhaustive. At the same time, it is important to mention that other dimensions not directly discussed in this article (e.g., cost-effectiveness of the intervention, neurological correlates of intervention change within specific dimensions) may be critical to consider as well. It is important that future work delineate both which and how many

dimensions are important to change within specific interventions targeting specific constructs.

Concluding Comments

Interventions are developed to change specific constructs, so that they alter the expression or presence of these constructs. However, psychological science has progressed sufficiently to suggest that the presence of psychological constructs is often highly variable and multifaceted. Thus, the conceptualization and examination of whether interventions actually change constructs must align itself with the complexity underlying the constructs targeted for intervention.

The goal of this article was to provide a viable conceptual and methodological framework to guide the study of intervention change. The RPC Model has heuristic value and can be utilized to conceptualize, examine, and classify both findings gathered within and between prior investigations, and inform the conceptualization, examination, and classification of future work examining interventions. In line with the proposed RPC Model, we encourage further conceptual and empirical work on intervention change. Most importantly, we recommend that future work incorporate the RPC Model in conceptualizing, examining, and classifying change within interventions, to account for the multidimensional conceptualization of the psychological constructs targeted for intervention. Specifically, we recommend that future research use the RPC Model to conduct: (a) meta-analytic reviews to conceptualize, examine, and classify prior findings; and (b) future investigations studying intervention change.

In psychological science, the constructs targeted for intervention often are very difficult to change, and their presence or expression often have very large implications for the people for which they affect. Thus, the main goal of interventions developed to deal with these constructs is often to either get rid of their negative effects, or enhance their positive effects. However, conducting an investigation to examine the effects of an intervention comes at a price: It is very difficult to get a straight answer as to whether the intervention changes what it was developed to change. The RPC Model conceptualizes change so that investigators may arrive at straight answers to the most critical questions surrounding the ability of interventions to change psychological constructs. Future theoretical and empirical work ought to use the RPC Model to properly conceptualize and examine the complexity of change within interventions, given the intention of interventions to change complex constructs.

References

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist 14–18 and 1991 profile.* Burlington: University of Vermont.

Achenbach, T. M., & Edelbrock, C. S. (1983). *Manual for the Child Behavior Checklist and profile.* Burlington: University of Vermont.

Achenbach, T. M., & Edelbrock, C. S. (1991). *Manual for the Child Behavior Checklist and profile.* Burlington: University of Vermont.

Achenbach, T. M., Krukowski, R. A., Dumenci, L., & Ivanova, M. Y. (2005). Assessment of adult psychopathology: Meta-analyses and implications of cross-informant correlations. *Psychological Bulletin, 131,* 361–382.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213–232.

Agras, W. S., Schneider, J. A., Arnow, B., Raeburn, S. D., & Telch, C. F. (1989). Cognitive-behavioral and response-prevention treatments for bulimia nervosa. *Journal of Consulting and Clinical Psychology, 57,* 215–221.

Agras, W. S., Telch, C. F., Arnow, B., Eldredge, K., Detzer, M. J., Henderson, J., et al. (1995). Does interpersonal therapy help patients with binge eating disorder who fail to respond to cognitive-behavioral therapy? *Journal of Consulting and Clinical Psychology, 63,* 356–360.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Barlow, D. H., Raffa, S. D., & Cohen, E. M. (2002). Psychosocial treatments for panic disorders, phobias, and generalized anxiety disorder. In P. E. Nathan & J. M. Gorman (Eds.), *A guide to treatments that work* (2nd ed., pp. 301–335). New York: Oxford University Press.

Barlow, D. H., Rapee, R. M., & Brown, T. A. (1992). Behavioral treatment of generalized anxiety disorder. *Behavior Therapy, 23,* 551–570.

Barrett, P., Healy-Farrell, L., & March, J. S. (2004). Cognitive-behavioral family treatment of childhood obsessive-compulsive disorder: A controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 43,* 46–62.

Barrett, P. M., Dadds, M. R., & Rapee, R. M. (1996). Family treatment of childhood anxiety: A controlled trial. *Journal of Consulting and Clinical Psychology, 64,* 333–342.

Beck, A. T., Brown, G., Eidelson, J. I., Steer, R. A., & Riskind, J. H. (1987). Differentiating anxiety and depression: A test of the cognitive content-specificity hypothesis. *Journal of Abnormal Psychology, 96,* 179–183.

Beck, A. T., Brown, G., Epstein, N., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56,* 893–897.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 41,* 561–571.

Bernstein, D. A., & Borkovec, T. D. (1973). *Progressive relaxation training.* Champaign, IL: Research Press.

Bird, H. R., Gould, M. S., & Staghezza, B. (1992). Aggregating data from multiple informants in child psychiatry epidemiological research. *Journal of the American Academy of Child and Adolescent Psychiatry, 31,* 78–85.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111,* 1061–1071.

Bouton, M. E., Mineka, S., & Barlow, D. H. (2001). A modern learning theory perspective on the etiology of panic disorder. *Psychological Review, 108,* 4–32.

Bowden, C. L., Calabrese, J. R., McElroy, S. L., Gyulai, L., Wassef, A., Petty, F., et al. (2000). A randomized, placebo-controlled 12-month trial of divalproex and lithium in treatment of outpatients with bipolar I disorder. *Archives of General Psychiatry, 57,* 481–489.

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review, 110,* 762–784.

Brewin, C. R., Dalgleish, T., & Joseph, S. (1996). A dual representation theory of posttraumatic stress disorder. *Psychological Review, 103,* 670–686.

Brown, R. A., & Lewinsohn, P. M. (1984). A psychoeducational approach to the treatment of depression: Comparison of group, individual, and minimal contact procedures. *Journal of Consulting and Clinical Psychology, 52,* 774–783.

Butler, G., Fennell, M., Robson, P., & Gelder, M. (1991). Comparison of behavior therapy and cognitive behavior therapy in the treatment of generalized anxiety disorder. *Journal of Consulting and Clinical Psychology, 59,* 167–175.

Butler, G., Gelder, M., Hibbert, G., Cullington, A., & Klimes, I. (1987).

Anxiety management: Developing effective strategies. *Behaviour Research and Therapy, 25,* 517–522.

Casey, R. J., & Berman, J. S. (1985). The outcomes of psychotherapy with children. *Psychological Bulletin, 98,* 388–400.

Chambless, D. L., Baker, M., Baucom, D. H., Beutler, L. E., 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.

Chick, J., Howlett, H., Morgan, M. Y., & Ritson, B. (2000). United Kingdom multicentre acamprosate study (UKMAS): A 6-month prospective study of acamprosate versus placebo in preventing relapse after withdrawal from alcohol. *Alcohol and Alcoholism, 35,* 176–187.

Clarke, G. N., Rohde, P., Lewinsohn, P. M., Hops, H., & Seeley, J. R. (1999). Cognitive-behavioral treatment of adolescent depression: Efficacy of acute group treatment and booster sessions. *Journal of the American Academy of Child and Adolescent Psychiatry, 38,* 272–279.

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

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

De Los Reyes, A., & Kazdin, A. E. (2004). Measuring informant discrepancies in clinical child research. *Psychological Assessment, 16,* 330–334.

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131,* 483–509.

Di Nardo, P. A., & Barlow, D. H. (1988). *Anxiety Disorders Interview Schedule-Revised (ADIS-R).* Albany, NY: Phobia and Anxiety Disorders Clinic, State University of New York.

Dowrick, C., Dunn, G., Ayuso-Mateos, J. L., Dalgard, O. S., Page, H., Lehtinen, V., et al. (2000). Problem solving treatment and group psychoeducation for depression: Multicentre randomised controlled trial. *British Medical Journal, 321,* 1450–1454.

Eysenck, H. J., & Eysenck, S. B. G. (1968). *Manual for the Eysenck Personality Inventory.* San Diego: EdITS.

Frank, E., Prien, R. F., Jarrett, R. B., Keller, M. B., Kupfer, D. J., Lavori, P. W., et al. (1991). Conceptualization and rationale for consensus definitions of terms in major depressive disorder. *Archives of General Psychiatry, 48,* 851–855.

Geller, D. A., Hoog, S. L., Heiligenstein, J. H., Ricardi, R. K., Tamura, R., Kluszynski, S., et al. (2001). Fluoxetine treatment for obsessive-compulsive disorder in children and adolescents: A placebo-controlled clinical trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 773–779.

Gould, R. A., Otto, M. W., Pollack, M. H., & Yap, L. (1997). Cognitive behavioral and pharmacological treatment of generalized anxiety disorder: A preliminary meta-analysis. *Behavior Therapy, 28,* 285–305.

Guisinger, S. (2003). Adapted to flee famine: Adding an evolutionary perspective on anorexia nervosa. *Psychological Review, 110,* 745–761.

Hamilton, M. (1959). The measurement of anxiety states by rating. *British Journal of Medical Psychology, 32,* 50–55.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23,* 56–62.

Hayward, C., Varady, S., Albano, A. M., Thienemann, M., Henderson, L., & Schatzberg, A. F. (2000). Cognitive-behavioral group therapy for social phobia in female adolescents: Results of a pilot study. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 721–726.

Hazell, P. L., & Stuart, J. E. (2003). A randomized controlled trial of clonidine added to psychostimulant medication for hyperactive and aggressive children. *Journal of the American Academy of Child and Adolescent Psychiatry, 42,* 886–894.

Hill, C. E., & Lambert, M. J. (2004). Methodological issues in studying psychotherapy processes and outcomes. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 84–135). New York: Wiley.

Himadi, W. G., Boice, R., & Barlow, D. H. (1986). Assessment of agoraphobia II: Measurement of clinical change. *Behaviour Research and Therapy, 24,* 321–332.

Hoagwood, K., Hibbs, E., Brent, D., & Jensen, P. (1995). Introduction to the special section: Efficacy and effectiveness in studies of child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 683–687.

Ioannidis, J. P. A. (2005). Contradicted and initially stronger effects in highly cited clinical research. *Journal of the American Medical Association, 294,* 218–228.

Ioannidis, J. P. A., & Lau, J. (2001). Evolution of treatment effects over time: Empirical insight from recursive cumulative metaanalyses. *Proceedings of the National Academy of Sciences USA, 98,* 831–836.

Ioannidis, J. P. A., & Trikalinos, T. A. (2005). Early extreme contradictory estimates may appear in published research: The proteus phenomenon in molecular genetics research and randomized trials. *Journal of Clinical Epidemiology, 58,* 543–549.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12–19.

Kaplan, R. M. (2000). Two pathways to prevention. *American Psychologist, 55,* 382–396.

Kazdin, A. E. (2000). *Psychotherapy for children and adolescents: Directions for research and practice.* New York: Oxford University Press.

Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.

Kazdin, A. E. (2005). *Parent management training: Treatment for oppositional, aggressive, and antisocial behavior in children and adolescents.* New York: Oxford University Press.

Kazdin, A. E., & Wilson, G. T. (1978). Criteria for evaluating psychotherapy. *Archives of General Psychiatry, 35,* 407–416.

Kendall, P. C. (1994). Treating anxiety disorders in children: Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology, 62,* 100–110.

Kendall, P. C., Butcher, J. N., & Holmbeck, G. N. (Eds.). (1999). *Handbook of research methods in clinical psychology* (2nd ed.). New York: Wiley.

Kendall, P. C., & Grove, W. M. (1988). Normative comparisons in therapy outcome. *Behavioral Assessment, 10,* 147–158.

Kovacs, M. (1981). Rating scales to assess depression in school-aged children. *Acta Paedopsychiatrica, 46,* 305–315.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry, 160,* 1566–1577.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, 59,* 877–883.

Kwan, V. S. Y., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: An interpersonal approach. *Psychological Review, 111,* 94–110.

Lambert, M. J., Hatch, D. R., Kingston, M. D., & Edwards, B. C. (1986). Zung, Beck, and Hamilton Rating Scales as measures of treatment outcome: A meta-analytic comparison. *Journal of Consulting and Clinical Psychology, 54,* 54–59.

Lambert, M. J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook*

of psychotherapy and behavior change (5th ed., pp. 139–193). New York: Wiley.

Lewinsohn, P. M., Clarke, G. N., Hops, H., & Andrews, J. (1990). Cognitive-behavioral treatment for depressed adolescents. *Behavior Therapy, 21,* 385–401.

Liebowitz, M. R., Turner, S. M., Piacentini, J. C., Beidel, D. C., Clarvit, S. R., Davies, S. O., et al. (2002). Fluoxetine in children and adolescents with OCD: A placebo-controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 41,* 1431–1438.

Lonigan, C. J., Elbert, J. C., & Johnson, S. B. (1998). Empirically supported psychological interventions for children: An overview. *Journal of Clinical Child Psychology, 27,* 138–145.

Marks, I. M., & Mathews, A. M. (1979). Brief standard self-rating for phobic patients. *Behaviour Research and Therapy, 17,* 263–267.

Matt, G. E. (1989). Decision rules for selecting effect sizes in meta-analysis: A review and reanalysis of psychotherapy outcome studies. *Psychological Bulletin, 105,* 106–115.

Mersch, P. P. A. (1995). The treatment of social phobia: The differential effectiveness of exposure *in vivo* and an integration of exposure *in vivo*, rational emotive therapy and social skills training. *Behaviour Research and Therapy, 33,* 259–269.

Michelson, D., Buitelaar, J. K., Danckaerts, M., Gillberg, C., Spencer, T. J., Zuddas, A., et al. (2004). Relapse prevention in pediatric patients with ADHD treated with atomoxetine: A randomized, double-blind, placebo-controlled study. *Journal of the American Academy of Child and Adolescent Psychiatry, 43,* 896–904.

Monroe, S. M., & Harkness, K. L. (2005). Life stress, the "kindling" hypothesis, and the recurrence of depression: Considerations from a life stress perspective. *Psychological Review, 112,* 417–445.

Nathan, P. E., & Gorman, J. M. (2002). Efficacy, effectiveness, and the clinical utility of psychotherapy research. In P. E. Nathan & J. M. Gorman (Eds.), *A guide to treatments that work* (2nd ed., pp. 643–654). New York: Oxford University Press.

Nemeroff, C. B., Evans, D. L., Gyulai, L., Sachs, G. S., Bowden, C. L., Gergel, I. P., et al. (2001). Double-blind, placebo-controlled comparison of imipramine and paroxetine in the treatment of bipolar depression. *American Journal of Psychiatry, 158,* 906–912.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Offord, D. R., Boyle, M. H., Racine, Y., Szatmari, P., Fleming, J. E., Sanford, M., et al. (1996). Integrating assessment data from multiple informants. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 1078–1085.

Ogles, B. M., Lambert, M. J., Weight, D. G., & Payne, I. R. (1990). Agoraphobia outcome measurement: A review and meta-analysis. *Psychological Assessment, 2,* 317–325.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108,* 483–522.

Ollendick, T. H. (1983). Reliability and validity of the Revised Fear Survey Schedule for Children (FSSC-R). *Behaviour Research and Therapy, 21,* 685–692.

Öst, L-G., Svensson, L., Hellström, K., & Lindwall, R. (2001). One-session treatment of specific phobias in youths: A randomized clinical trial. *Journal of Consulting and Clinical Psychology, 69,* 814–824.

Patterson, G. R. (1982). *Coercive family process.* Eugene, OR: Castalia.

Pearson, D. A., Santos, C. W., Roache, J. D., Casat, C. D., Loveland, K. A., Lachar, D., et al. (2003). Treatment effects of methylphenidate on behavioral adjustment in children with mental retardation and ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry, 42,* 209–216.

Piacentini, J. C., Cohen, P., & Cohen, J. (1992). Combining discrepant diagnostic information from multiple sources: Are complex algorithms better than simple ones? *Journal of Abnormal Child Psychology, 20,* 51–63.

Pliszka, S. R., Browne, R. G., Olvera, R. L., & Wynne, S. K. (2000). A double-blind, placebo-controlled study of Adderall and methylphenidate in the treatment of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 619–626.

Prien, R. F., Carpenter, L. L., & Kupfer, D. J. (1991). The definition and operational criteria for treatment outcome of major depressive disorder. *Archives of General Psychiatry, 48,* 796–800.

Reid, J. B., Patterson, G. R., & Snyder, J. (Eds.). (2002). *Antisocial behavior in children and adolescents: A developmental analysis and model for intervention.* Washington, DC: American Psychological Association.

Reynolds, C. R., & Richmond, B. O. (1978). What I think and feel: A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology, 6,* 271–280.

Ronan, K. R., Kendall, P. C., & Rowe, M. (1994). Negative affectivity in children: Development and validation of a self-statement questionnaire. *Cognitive Therapy and Research, 18,* 509–528.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52,* 59–82.

Roth, A., & Fonagy, P. (2005). *What works for whom?: A critical review of psychotherapy research* (2nd ed.). New York: Guilford Press.

Rubio-Stipec, M., Fitzmaurice, G., Murphy, J., & Walker, A. (2003). The use of multiple informants in identifying the risk factors of depressive and disruptive disorders: Are they interchangeable? *Social Psychiatry and Psychiatric Epidemiology, 38,* 51–58.

Rush, A. J., Weissenburger, J., & Eaves, G. (1986). Do thinking patterns predict depressive symptoms? *Cognitive Therapy and Research, 10,* 225–236.

Ryan, J. J., Lopez, S. J., & Sumerall, S. W. (2001). Understanding test construction. In W. I. Dorfman & M. Hersen (Eds.), *Understanding psychological assessment* (pp. 1–15). New York: Kluwer Academic/Plenum Publishers.

Sanders, M. R., Markie-Dadds, C., Tully, L. A., & Bor, W. (2000). The Triple P-Positive Parenting Program: A comparison of enhanced, standard, and self-directed behavioral family intervention for parents of children with early onset conduct problems. *Journal of Consulting and Clinical Psychology, 68,* 624–640.

Schmidt, N. B., Woolaway-Bickel, K., Trakowski, J., Santiago, H., Storey, J., Koselka, M., et al. (2000). Dismantling cognitive-behavioral treatment for panic disorder: Questioning the utility of breathing retraining. *Journal of Consulting and Clinical Psychology, 68,* 417–424.

Schwartz, G. E., Davidson, R. J., & Goleman, D. J. (1978). Patterning of cognitive and somatic processes in the self-regulation of anxiety: Effects of meditation versus exercise. *Psychosomatic Medicine, 40,* 321–328.

Shaffer, D., Fisher, P., Dulcan, M. K., Davies, M., Piacentini, J. C., Schwab-Stone, M. E., et al. (1996). The NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3): Description, acceptability, prevalence rates, and performance in the MECA study. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 865–877.

Silverman, W. K. (1987). *Anxiety Disorders Interview Schedule for Children (ADIS)*. State University of New York at Albany: Graywind Publications.

Silverman, W. K., & Nelles, W. B. (1988). The Anxiety Disorders Interview Schedule for Children. *Journal of the American Academy of Child and Adolescent Psychiatry, 27,* 772–778.

Snaith, R. P., Bridge, C. W. K., & Hamilton, M. (1976). The Leeds scales for the self-assessment of anxiety and depression. *British Journal of Psychiatry, 128,* 156–165.

Spielberger, C. D. (1973). *Manual for State-Trait Anxiety Inventory for Children.* Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Strauss, C. (1987). *Modification of trait portion of State-Trait Anxiety Inventory for Children-parent form.* (Available from the author, Department of Clinical and Health Psychology, College of Public Health and Health Professions, University of Florida, P. O. Box 100165, Gainesville, FL 32610–0165.)

Strupp, H. H. (1996). The tripartite model and the *Consumer Reports* study. *American Psychologist, 51,* 1017–1024.

Strupp, H. H., & Hadley, S. W. (1977). A tripartite model of mental health and therapeutic outcomes. *American Psychologist, 32,* 187–196.

Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review, 112,* 159–192.

Szechtman, H., & Woody, E. (2004). Obsessive-compulsive disorder as a disturbance of security motivation. *Psychological Review, 111,* 111–127.

Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically-validated psychological treatments: Report and recommendations. *Clinical Psychologist, 48,* 3–23.

Tempesta, E., Janiri, L., Bignamini, A., Chabac, S., & Potgieter, A. (2000). Acamprosate and relapse prevention in the treatment of alcohol dependence: A placebo-controlled study. *Alcohol and Alcoholism, 35,* 202–209.

Trikalinos, T. A., Churchill, R., Ferri, M., Leucht, S., Tuunainen, A., Wahlbeck, K., et al. (2004). Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *Journal of Clinical Epidemiology, 57,* 1124–1130.

Van Rooy, D., Van Overwalle, F., Vanhoomissen, T., Labiouse, C., & French, R. (2003). A recurrent connectionist model of group biases. *Psychological Review, 110,* 536–563.

Watson, J. P., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology, 33,* 448–457.

Watson, J. P., & Marks, I. M. (1971). Relevant and irrelevant fear in flooding: A crossover study of phobic patients. *Behavior Therapy, 2,* 275–293.

Webster-Stratton, C., & Hammond, M. (1997). Treating children with early-onset conduct problems: A comparison of child and parent training interventions. *Journal of Consulting and Clinical Psychology, 65,* 93–109.

Weisz, J. R. (2004). *Psychotherapy for children and adolescents: Evidence-based treatments and case examples.* New York: Cambridge University Press.

Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology, 55,* 542–549.

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117,* 450–468.

Wigal, S., Swanson, J. M., Feifel, D., Sangal, R. B., Elia, J., Casat, C. D., et al. (2004). A double-blind, placebo-controlled trial of dexmethylphenidate hydrochloride and *d,l-threo*-methylphenidate hydrochloride in children with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 43,* 1406–1414.

Youngstrom, E. A., Findling, R. L., & Calabrese, J. R. (2003). Who are the comorbid adolescents? Agreement between psychiatric diagnosis, youth, parent, and teacher report. *Journal of Abnormal Child Psychology, 31,* 231–245.