



Identifying Evidence-Based Interventions for Children and Adolescents Using the Range of Possible Changes Model: A Meta-Analytic Illustration

Journal:	<i>Behavior Modification</i>
Manuscript ID:	BMOD-09-0002.R1
Manuscript Type:	Original Manuscripts
Date Submitted by the Author:	
Complete List of Authors:	De Los Reyes, Andres; University of Maryland at College Park, Department of Psychology Kazdin, Alan; Yale University, Department of Psychology
Keywords:	efficacy , effectiveness, intervention, range of possible changes



Online Supporting Material

Expanded Discussion of Methods Employed in De Los Reyes and Kazdin's
*Identifying Evidence-Based Interventions for Children and Adolescents Using the
Range of Possible Changes Model: A Meta-Analytic Illustration*

Method

Interventions Examined

The sample examined in the meta-analysis consisted of a set of exemplary randomized controlled clinical trials that tested the efficacies of two specific interventions, each developed to target a specific psychological construct: (a) youth-focused cognitive-behavioral therapy for childhood anxiety problems (hereafter referred to as CBT); and (b) parent-focused behavioral parent training for childhood conduct problems (hereafter referred to as BPT). We chose studies examining these two interventions for two reasons. First, a recent methodological meta-analysis suggests studies examining interventions for childhood anxiety and conduct problems tend to employ a relatively large number of outcome measures (Weisz, Jensen Doss, & Hawley, 2005). Thus, examining interventions for these two specific constructs would allow for an examination of within-study consistencies in research findings. Second, recent meta-analytic work has identified a number of controlled outcome studies examining each of these two interventions (Weisz, Hawley, & Jensen Doss, 2004). This suggested the ability within this population of studies to examine between-study consistencies in intervention effects.

We defined CBT using the operational definition employed by a recent meta-analysis of youth interventions (Weisz et al., 2004): An intervention focused on individual youths entailing "efforts to identify and alter cognitions that contribute to the anxiety and to identify and alter maladaptive behavior (such as avoidance of feared situations) that may serve to sustain the condition." (p. 751). Similarly, we defined BPT using a definition employed by Weisz et al. (2004): Those interventions focused on parents that aim to reduce child conduct problems by employing some or all of the following components:

- (1) parents learn basic behavioral principles relevant to child rearing;
- (2) parents learn how to define, track, and record rates of the antisocial and prosocial behaviors they

1
2
3
4 want to target; (3) parents are helped to design, role play, carry out, and refine
5
6 behavior modification programs while continuing to record rates of target behavior to
7
8 assess intervention effects. (p. 792)
9

10 *Literature Review*

11
12 The literature review and collection of studies were accomplished on two fronts. First, relevant
13
14 intervention studies published in years up to and including 2002 were derived from a previous meta-
15
16 analysis of the youth intervention literature (Weisz et al., 2004). We derived studies in this manner for a
17
18 number of reasons. Specifically, the Weisz et al. (2004) meta-analysis thoroughly reviewed the literature
19
20 for controlled experimental work examining the two specific interventions that we wished to review.
21
22 Additionally, the literature search methods employed by this meta-analysis incorporated methods taken
23
24 from prior seminal meta-analytic reviews of the youth intervention literature (Weisz, Weiss, Alicke, &
25
26 Klotz, 1987; Weisz, Weiss, Han, Granger, & Morton, 1995). Thus, to remain consistent with prior
27
28 reviews, studies conducted in years up to and including 2002 were collected from a recent meta-analytic
29
30 review employing the standard methods of literature review for the youth intervention literature.

31
32 Second, literature searches for relevant intervention studies published between the years 2003
33
34 through 2006 were conducted employing the same methods as Weisz et al. (2004). Two standard
35
36 computerized databases were employed to identify relevant studies. First, we used Psychinfo, limiting
37
38 our search from 2003 through 2006, and employing 21 psychotherapy-related keywords derived from
39
40 prior meta-analytic work (see Weisz et al., 1987; Weisz et al., 1995). Second, consistent with Weisz et al.
41
42 (2004), we conducted searches of the same years using MEDLINE, via PubMed; this is the primary
43
44 bibliographic computerized database of the National Library of Medicine. We limited our search from
45
46 2003 through 2006, and used the same search terms as Weisz et al. (2004): *Mental Disorders* with the
47
48 search limits: *clinical trial, child (3-18 years), published in English, and human subjects*.

49 *Criteria for Study Inclusion and Study List*

50
51 *Peer-Review.* Besides operational definitions of the interventions examined and literature search
52
53 methodology, we employed criteria for identifying relevant intervention studies, and including such
54
55 studies in the meta-analytic illustration. First, studies were required to have undergone some form of
56
57 peer-review. This meant that identified studies were published in peer-reviewed journals, and
58
59
60

1
2
3
4 unpublished dissertations and manuscripts, as well as book chapters were excluded from this review. We
5
6 decided to include only published studies for multiple reasons. First, we employed this criterion to
7
8 remain consistent with the prior work from which we were adopting literature search and inclusion
9
10 criteria (Weisz et al., 1987; Weisz et al., 1995; Weisz et al., 2004). In this way, studies we identified in
11
12 our own literature searches (i.e., post-2002 studies published through 2006) were retrieved identically as
13
14 those from the Weisz et al. (2004) meta-analysis. Further, studies identified by categorical classification
15
16 systems as providing supportive evidence of EBIs are largely gleaned from the peer-reviewed literature
17
18 (Lonigan, Elbert, & Johnson, 1998; Nathan & Gorman, 2007; Roth & Fonagy, 2005).

19
20 *Main Methodological Criteria.* Among peer-reviewed studies identified in literature searches, we
21
22 employed stringent criteria for study inclusion, based on prior work (Weisz et al., 2004). These criteria
23
24 were critical to employ in order to arrive at a set of studies for which it would be possible to examine
25
26 consistencies in outcomes findings across multiple indices of intervention effects gathered within studies
27
28 and between studies of the same intervention. Specifically, the studies examining CBT or BPT included
29
30 in the review were required to meet the following criteria: (a) the intervention being examined must have
31
32 been compared to an inert control group, such as waitlist, no treatment, placebo or other inert process; (b)
33
34 each study must have employed a prospective design and random assignment of participants to
35
36 conditions; (c) each study must have examined a sample of youths within a 3- to 18-year-old age range;
37
38 (d) each study must have examined participants selected for exhibiting the behavior or emotional
39
40 problems identified previously (child anxiety, child conduct); (e) each study must have employed a post-
41
42 intervention assessment of the construct being targeted for intervention; and (f) participants in groups
43
44 being compared to one another must not have been taking psychotropic medications.

45
46 *Criterion Excluding Comparisons of Interventions with Active Interventions.* The criterion of
47
48 excluding comparisons between alternative interventions differed slightly from that of Weisz et al.
49
50 (2004). We only included studies comparing the interventions being examined to control groups for two
51
52 reasons. First, we wished to control for between-study inconsistencies attributable to differences across
53
54 studies in what kinds of between-condition comparisons were made (e.g., waitlist controls vs. alternative
55
56 treatments). Second, we were interested in restricting the ability of the RPC Model to detect and take into
57
58 account within- and between-study inconsistencies. Indeed, allowing for only a review and classification
59
60

1
2
3
4 of control condition comparison studies would test the well-accepted notion that interventions all
5 generally outperform waitlist and other control conditions (Lambert & Ogles, 2004).
6
7

8 *Criterion Excluding Studies Only Yielding Non-Significant Outcomes.* Of the studies identified with
9 the above criteria, we required studies to report in their analyses a statistically significant benefit of the
10 intervention examined, relative to a control condition on at least one outcome measure of the target
11 construct. We employed this criterion because this requirement would likely increase consistent findings,
12 particularly between-study consistencies. Because this criterion would rule out studies that show no
13 benefit of interventions, we could examine whether applying the RPC Model to meta-analysis would
14 allow for the identification of patterns of significant effects, even among a group of studies already
15 selected for yielding statistically significant benefits of the intervention. We believed this to contribute
16 significantly to providing as conservative a test as possible of the RPC Model and its applicability to
17 meta-analysis.
18
19
20
21
22
23
24
25
26

27 *Criterion Excluding Studies That Did Not Employ Multiple Outcome Measures.* Lastly, in order to
28 ensure that it would be possible to examine patterns of consistent effects, we required that studies
29 employed at least three measures of the construct targeted for intervention (i.e., three anxiety measures for
30 studies of CBT, three conduct problem measures for studies of BPT). By “employed,” we mean that
31 studies must have prospectively administered at least three outcome measures of the target construct (i.e.,
32 pre- and post-intervention), and sufficient data must have been reported in the published study to calculate
33 effect sizes and tests of statistical significance. By “three outcome measures,” we mean that the authors
34 must have employed three measures that were each distinctly administered from each other (e.g., a study
35 was not included if the authors only reported three findings from subscale scores gleaned from the same
36 measure). We included this criterion for two reasons. Specifically, in employing at least three target
37 construct measures, it would be possible to examine a range of outcomes for consistencies in significant
38 effects and variability in treatment effects, as measured by effect sizes. Further, requiring the
39 employment of a range of outcome measures of the target construct would allow for both classifications
40 of study findings under the RPC Model categories and statistical comparisons between the upper and
41 lower limit effect sizes observed within and across studies with both each other and the mean effect size
42 observed across studies (Table 2).
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 *Study List.* The final study list included 9 studies of CBT and 7 studies of BPT, yielding 21
5
6 intervention-control comparisons (11 for CBT, 10 for BPT). Lists of the studies and descriptions of basic
7
8 demographic, methodological, and outcome measure characteristics are presented in Tables 3 and 4. The
9
10 number of studies on the list is less than the total number of controlled outcome studies of CBT and BPT
11
12 within the date range of our review. This is due in large part to our inclusion requirements of random
13
14 assignment to treatment and waitlist or other control conditions (controlled trials comparing CBT and
15
16 BPT to active treatment conditions were excluded, for reasons cited previously), employment of at least
17
18 three outcome measures of the primary target of the intervention (controlled trials employing less than
19
20 three measures were excluded, for reasons cited previously), and 3- to 18-year-old sample child age
21
22 range. At the same time, the number of studies on the list is consistent with prior reviews of carefully
23
24 selected studies of cognitive-behavioral treatments of childhood anxiety, as well as parenting and family
25
26 treatments for child behavior problems (James, Soler, & Weatherall, 2005; Woolfenden, Williams, &
27
28 Peat, 2001).

29
30 In employing the methodological criteria, we nevertheless excluded studies that otherwise met
31
32 criteria for study inclusion. For instance, we excluded one study comparing two different CBT treatments
33
34 to a waitlist control condition because post-intervention data for the control group were not reported, and
35
36 the grand majority of the statistical comparisons made in the study compared the waitlist condition to a
37
38 collapsed group of both treatment conditions (Nauta, Scholing, Emmelkamp, & Minderaa, 2003).
39
40 Because only one of the CBT treatments could be characterized as strictly youth-focused, there was
41
42 insufficient information available to both determine whether statistical comparisons made in this study
43
44 could suggest whether CBT outperformed controls, and calculate effect sizes. Similar circumstances with
45
46 inabilities to code statistical differences and effects sizes between treatment and control groups were
47
48 found with two studies comparing CBT treatments with a waitlist control: Standard deviations of outcome
49
50 measures were not provided (Bernstein, Layne, Egan, & Tennison, 2005; Williams & Jones, 1989).

51
52 Two other CBT studies were excluded because some of the outcome measures employed were not
53
54 employed prospectively (both at pre- and post-intervention; Sud, 1994; Sud & Sharma, 1990). This did
55
56 not allow for an examination of whether experimental conditions were equivalent on these measures
57
58 before intervention, and thus this study did not meet the criterion of employing at least three prospectively
59
60

1
2
3
4 administered outcome measures of the target construct. One CBT study was excluded because the
5 children in the sample met primary diagnostic criteria for Asperger syndrome, and the presence of anxiety
6 symptoms appeared to be a secondary concern for children in this sample (Sofronoff, Attwood, & Hinton,
7 2005). Finally, a BPT study was excluded because the waitlist condition did not undergo random
8 assignment along with the two other intervention conditions examined in the study (waitlist participants
9 were assigned based on therapist availability; Bernal, Klinnert, & Schultz, 1980).

16 *Study Coding Procedures*

17
18 *Coding Manual.* A coding manual was developed to describe procedures for coding information
19 from studies (manual available from the authors). Briefly, the manual was separated into multiple parts,
20 and developed to outline and describe coding procedures for basic study characteristics (e.g., sample size,
21 type, and demographics) (Part 1), outcome measure characteristics (information source, outcome measure
22 methodology) (Part 2), effect size and statistical test calculations (Part 3), and classifications of studies
23 based on the RPC Model (Part 4). Mean effect size calculations were made using statistical software.
24
25
26
27
28

29
30 *Coding Descriptions and Reliability.* Three clinical science graduate students were trained to code
31 information gleaned from studies. All three coders were blind to the study hypotheses. Two coders were
32 trained to individually code all information in each of the 16 studies. One coder with experience in
33 conducting and coding information for meta-analyses was trained as a consensus coder with the key tasks
34 of leading coding meetings. Each of the 16 studies was separately coded in their entirety by the two
35 coders, and the consensus coder led reliability meetings with both coders present. Specifically, in these
36 meetings, the consensus coder led discussions of each item coded for each study, led discussions on
37 resolutions of coding inconsistencies between the two coders, and recorded the number of instances in
38 which inconsistencies were evident between the two coders. Resolution of coding inconsistencies was
39 reached by consensus from all coders, and the consensus coder recorded a final code in these
40 circumstances. Across items coded within the 16 studies in the meta-analysis, the consensus coder
41 resolved coder inconsistencies 3.5% of the time for Parts 1-4 (156 out of 4481 items). The rate of
42 inconsistencies within each section was as follows: Part 1: 7.1% (64 out of 904 items); Part 2: 4.5% (45
43 out of 1005 items); Part 3: 1.4% (30 out of 2138 items); and Part 4: 3.9% (17 out of 434 items).
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 *Coder Training.* In order to ensure that reliability of codes gathered in one section did not influence
5 the level of reliability of codes gathered from other sections, coders first were trained and coded
6 information concerning basic study and outcome measure characteristics and statistical test and effect size
7 calculations. Once this information was coded, consensus codes were distributed to all coders for use in
8 coding information for the 16 studies, for codes relevant to classifying the evidence for individual studies.
9 Coder training for basic study and outcome measure characteristics and statistical test and effect size
10 calculations was accomplished by practicing applying the coding manual to 7 studies that were excluded
11 from the list of studies coded in the meta-analysis. These excluded studies included studies reporting
12 controlled trials of treatments for CBT and BPT. Coding practices relevant to classifying evidence under
13 the RPC Model involved having coders evaluate and code results of 14 hypothetical studies. Coding for
14 the 16 studies commenced after the study coders agreed that the coding manual and sheets were
15 sufficiently clear, and enough experience was accrued in practices for coders to report feeling adequately
16 prepared.
17
18
19
20
21
22
23
24
25
26
27
28

29 *Post-Intervention Effect Size Calculations.* Calculations of effect sizes were performed for each of
30 the methods employed by studies to examine intervention outcomes (mean differences, diagnostic status,
31 clinically significant change; e.g., Cohen, 1988; Cohen, Cohen, West, & Aiken, 2003; Rosenthal &
32 DiMatteo, 2001). Mean differences calculations were made by subtracting the control group mean from
33 the intervention group mean, and dividing this difference by the control group's standard deviation at
34 outcome (Glass's Δ ; see Rosenthal & DiMatteo, 2001). Glass's Δ is an effect size metric that meta-
35 analysts consider being within the d family metric of effect sizes (Rosenthal & DiMatteo, 2001). Thus,
36 we maintained a consistent presentation of effects across methods of analysis by presenting Glass's Δ
37 results using the d symbol. The studies under examination were derived from the efficacy literature
38 examining interventions for children. Thus, we calculated mean differences effect sizes consistent with
39 prior meta-analytic work examining child intervention research (e.g., Weisz et al., 1987; Weisz et al.,
40 1995; Weisz, McCarty, & Valeri, 2006).
41
42
43
44
45
46
47
48
49
50
51
52

53 Diagnostic status and clinically significant change calculations were calculated using the Phi (Φ)
54 coefficient to examine differences in proportions between conditions (see Cohen et al., 2003). There were
55 instances in which only results of statistical tests were available (e.g., t statistic). Thus, effect sizes
56
57
58
59
60

1
2
3
4 (using the r metric) were estimated in these instances using test statistics, as suggested elsewhere
5
6 (Rosenthal & DiMatteo, 2001). Given the use of r effect size measures for some calculations, and that Φ
7
8 is an r effect size measure as well, effect sizes calculated using Φ and r were converted to d , in order to
9
10 construct effect size ranges along a common metric (Rosenthal & DiMatteo, 2001). Lastly, all effect
11
12 sizes were adjusted to take into account small sample bias, employing Hedges small sample correction
13
14 (Hedges & Olkin, 1985).

15
16 *Calculating and Coding Statistical Tests of Intervention Outcomes.* In addition to evaluating the
17
18 strength of intervention effects via calculating effect sizes, we evaluated the consistency of statistical tests
19
20 of differences between intervention and control conditions. Studies included in the meta-analysis were
21
22 quite variable in the methods of statistical tests employed to examine statistical differences between
23
24 conditions. However, the statistical power of significance tests is influenced by sample size and the type
25
26 of statistical test employed (Cohen, 1988). Thus, calculations of statistical differences between conditions
27
28 were kept constant across examinations of statistical test outcomes within and between studies.
29
30 Specifically, for tests of mean differences, coders recorded the post-treatment means and standard
31
32 deviations for each intervention and control condition and for each outcome measure, and employed an
33
34 online independent samples t test calculator to code the results (Graphpad Software, Inc., 2005).

35
36 For tests of diagnostic status and clinically significant change, coders recorded the post-intervention
37
38 frequencies of participants in intervention and control conditions for each dichotomous outcome measure,
39
40 and employed an online chi square test statistical calculator to code the results of significant tests of
41
42 diagnostic status and clinically significant change (Ball, 2003). All statistical tests were conducted as
43
44 two-tailed tests to take into account the possibility of both positive and negative effects, and the threshold
45
46 for statistical significance was set at $p < .05$. Sometimes studies reported an intervention outcome using
47
48 only the result of a statistical test such as t test or chi square. In these instances, coders recorded the
49
50 statistical information, and employed that information to both calculate effect sizes, and code statistical
51
52 significance.

53
54 *Pre-Intervention Group Comparability.* Finally, as an added check on results of post-intervention
55
56 analyses as well as the fidelity of random assignment in each study, coders recorded pre-intervention
57
58 information for each of the measures, using procedures identical to codes of post-intervention results.
59
60

1
2
3
4 *RPC Model: Incorporating Study Codes, Effect Sizes, and Intervention Effects.* Coders employed the
5
6 RPC Model to classify the findings gleaned from each of the 21 intervention-control comparisons
7
8 conducted across the 16 studies. To make RPC Model categorical classifications coders reviewed
9
10 information they coded previously on outcome measure and statistical outcome characteristics (outcome
11
12 measure methodology, outcome measure source, method of statistical analysis), along with the results of
13
14 statistical outcomes (statistical tests, effect size calculations). Coders were provided with a copy of the
15
16 original table from De Los Reyes and Kazdin (2006) denoting criteria for the RPC Model categories. The
17
18 RPC Model's origin was not disclosed to coders during data collection.

19
20 Coders made RPC Model classifications for each intervention-control comparison (see Table 1).
21
22 Specifically, for each intervention-control comparison, coders identified the percentages of findings that
23
24 were statistically significant, based on information coded previously. These percentages of findings were
25
26 employed to determine the RPC Model category within which a given study would be classified. Further,
27
28 if specificity in significant effects could be identified (e.g., three or more findings based on parent report
29
30 yielding consistently significant effects and could be classified in the Evidence for Informant Specific
31
32 change category) and findings could not be classified in a non-specific effect category (e.g., Evidence for
33
34 Probable Change), the study was classified in an RPC Model category denoting specificity in intervention
35
36 effects (e.g., Evidence for Informant-Specific Change). For the purposes of these RPC Model
37
38 classifications as well as for all other calculations and classifications, we defined "finding" as any single
39
40 instance in which an intervention-control comparison was made on an outcome measure of the construct
41
42 targeted for intervention. Under such a definition, a single outcome measure could contribute more than
43
44 one finding if: (a) the measure was examined using more than one statistical method; and/or (b) more than
45
46 one subscale within that measure was examined using one or more statistical methods. Study
47
48 classifications using the RPC Model as well as calculations of mean effect sizes were based on the nature
49
50 and extent of these findings. For a discussion on the development and rationale for the structure and
51
52 criteria of the individual RPC Model categories, see De Los Reyes and Kazdin (2006).

53
54 Further, each intervention-control comparison was coded for the range of effect sizes (i.e., upper and
55
56 lower limit effect sizes) observed within its RPC Model classification. Coders coded these effect size
57
58 ranges for findings within the RPC Model category classifications for each of the 21 intervention-control
59
60

1
2
3
4 comparisons. For each intervention-control comparison and its RPC Model category classification(s),
5
6 coders identified the highest and lowest observed effect sizes. Under all circumstances, effect size ranges
7
8 consisted of all findings employed to reach the RPC Model category classification, irrespective of
9
10 whether the finding was statistically significant or if the effect size was negative (i.e., intervention
11
12 condition had worse scores, relative to controls). Additionally, coders employed Cohen's (1988) effect
13
14 size conventions of small (.20), medium (.50) and large (.80) effects to categorize effect size ranges.
15
16 Further, coders were instructed to consider any effect sizes below .20 (including negative effect sizes,
17
18 where the intervention had worse scores, relative to controls) under a new category: Below small. The
19
20 criteria employed to construct these ranges are presented in Table 2. Broadly, the effect size range
21
22 categories captured every possible upper and lower effect size limit that could be reached, based on
23
24 Cohen's (1988) effect size conventions. Coders employed this system to categorize effect size ranges.

25
26 Intervention-control comparisons were coded under a single effect size range, even if they were
27
28 ascribed more than one RPC Model category. For example, if an intervention-control comparison yielded
29
30 evidence specific to parent-rated outcomes, measured via questionnaire, then the effect size range for this
31
32 comparison would encompass only the findings within these category classifications. Similarly, if an
33
34 intervention-control comparison was classified under an RPC Model category denoting specificity of
35
36 change (Table 1), then the effect size range only encompassed findings within this category classification.

37
38 *Quantifying Mean Effect Sizes.* We were interested in comparing effect size findings gleaned from
39
40 the RPC Model to the mean effect size gleaned across studies. Given that study inclusion criteria
41
42 pertaining to number of outcome measures employed resulted in each study providing more than one
43
44 effect size, effect sizes were aggregated within the study so that a mean effect size could be attained for
45
46 each intervention-control comparison. Additionally, in cases in which an RPC Model classification of a
47
48 study was for specificity in change (Table 1), this might result in the inclusion of some effect sizes
49
50 gleaned from the study included in effect size ranges and others left out. Therefore, in order to provide as
51
52 conservative a test as possible, all effect sizes gleaned from the study were employed to calculate mean
53
54 effect sizes for the study, regardless of the RPC Model classification for that study. A further
55
56 consideration is that the RPC Model was developed to classify individual intervention-control
57
58 comparisons. Thus, studies examining more than one intervention yielded more than one data point.
59
60

1
2
3
4 Therefore, we calculated and reported a mean effect size for each intervention-control comparison, and
5 categorized this mean effect size as a below small, small, medium, or large effect, consistent with
6 Cohen's (1988) effect size conventions (Table 2).
7
8

9
10 *Data-Analytic Plan.* The main analyses compared the mean upper and lower limit effect size
11 findings identified within studies both to each other and to estimates of mean effect sizes across studies.
12 These comparisons were addressed in two ways. First, we conducted a paired-samples t test to compare
13 the mean lower limit effect size and the mean upper limit effect size gleaned from individual intervention-
14 control comparisons. Second, we conducted two one-sample t tests, one comparing mean lower limit
15 effect sizes across intervention-control comparisons with the mean effect size across intervention-control
16 comparisons, and another comparing mean upper limit effect sizes across intervention-control
17 comparisons with the mean effect size across intervention-control comparisons. Results reported below
18 were consistent, regardless of whether one-sample or paired-samples t tests were employed. Further, we
19 ran the same analyses comparing RPC Model upper and lower limit effect sizes with the mean
20 intervention-control comparison effect size, excluding the upper and lower limit effect sizes for each
21 intervention-control comparison from calculations of the mean. Results from these analyses were
22 consistent with results of comparisons of upper and lower limit effect sizes with mean comparison effect
23 sizes, which included the upper and lower limit effects in the mean, suggesting that an outlier effect
24 specific to either upper or lower limit effects could not explain the findings reported below.
25
26
27
28
29
30
31
32
33
34
35
36
37
38

39 All analyses were conducted within the entire sample for a specific intervention (i.e., separate for
40 CBT and BPT studies). Additionally, hypotheses were directional and tested with low statistical power,
41 given the small number of studies examined in the meta-analysis. Therefore, all tests comparing the RPC
42 Model to other approaches were conducted as one-tailed significance tests, and effect sizes for analyses
43 were calculated using Cohen's d based on the test statistics yielded from statistical analyses.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

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

- Ball, C. N. (2003). *Georgetown linguistics: Web chi square calculator*. Retrieved January 22, 2007, from http://www.georgetown.edu/faculty/ballc/webtools/web_chi.html.
- *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.
- Bernal, M. E., Klinnert, M. D., & Schultz, L. A. (1980). Outcome evaluation of behavioral parent training and client-centered parent counseling for children with conduct problems. *Journal of Applied Behavior Analysis, 13*, 677-691.
- Bernstein, G. A., Layne, A. E., Egan, E. A., & Tennison, D. M. (2005). School-based interventions for anxious children. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*, 1118-1127.
- 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. (2006). Conceptualizing changes in behavior in intervention research: The range of possible changes model. *Psychological Review, 113*, 554-583.
- De Los Reyes, A., & Kazdin, A. E. (2008). When the evidence says, "Yes, no, and maybe so": Attending to and interpreting inconsistent findings among evidence-based interventions. *Current Directions in Psychological Science, 17*, 47-51.
- *Flannery-Schroeder, E. C., & Kendall, P. C. (2000). Group and individual cognitive-behavioral treatments for youth with anxiety disorders: A randomized clinical trial. *Cognitive Therapy and Research, 24*, 251-278.
- *Gallagher, H. M., Rabian, B. A., & McCloskey, M. S. (2004). A brief group cognitive-behavioral intervention for social phobia in childhood. *Journal of Anxiety Disorders, 18*, 459-479.
- Graphpad Software, Inc. (2005). *Quickcalcs online calculators for scientists: t test calculator*. Retrieved January 22, 2007, from <http://www.graphpad.com/quickcalcs/ttest1.cfm?Format=SD>.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

- 1
2
3
4 James, A., Soler, A., & Weatherall, R. (2005). Cognitive behavioural therapy for anxiety disorders in
5 children and adolescents. *Cochrane Database of Systematic Reviews*. New York: John Wiley &
6 Sons.
7
8
9
10 *Kendall, P. C. (1994). Treating anxiety disorders in children: Results of a randomized clinical trial.
11 *Journal of Consulting and Clinical Psychology, 62*, 100-110.
12
13 *Kendall, P. C., Flannery-Schroeder, E., Panichelli-Mindel, S. M., Southam-Gerow, M., Henin, A., &
14 Warman, M. (1997). Therapy for youths with anxiety disorders: A second randomized clinical
15 trial. *Journal of Consulting and Clinical Psychology, 65*, 366-380.
16
17
18
19 *King, N. J., Tonge, B. J., Mullen, P., Myerson, N., Heyne, D., Rollings, S., et al. (2000). Treating
20 sexually abused children with posttraumatic stress symptoms: A randomized clinical trial.
21 *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 1347-1355.
22
23
24
25 Lambert, M. J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M.J. Lambert
26 (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 139-
27 193). New York: John Wiley & Sons.
28
29
30
31 *Leal, L. L., Baxter, E. G., Martin, J., & Marx, R. W. (1981). Cognitive modification and systematic
32 desensitization with test anxious high school students. *Journal of Counseling Psychology, 28*,
33 525-528.
34
35
36
37 *Leung, C., Sanders, M. R., Leung, S., Mak, R., & Lau, J. (2003). An outcome evaluation of the
38 implementation of the Triple P-Positive Parenting Program in Hong Kong. *Family Process, 42*,
39 531-544.
40
41
42
43 Lonigan, C. J., Elbert, J. C., & Johnson, S. B. (1998). Empirically supported psychological interventions
44 for children: An overview. *Journal of Clinical Child Psychology, 27*, 138-145.
45
46
47 *McMurray, N. E., Bell, R. J., Fusillo, A. D., Morgan, M., & Wright, F. A. C. (1986). Relationship
48 between locus of control and effects of coping strategies on dental stress in children. *Child &*
49 *Family Behavior Therapy, 8*, 1-17.
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4 Nauta, M. H., Scholing, A., Emmelkamp, P. M. G., & Minderaa, R. B. (2003). Cognitive-behavioral
5
6 therapy for children with anxiety disorders in a clinical setting: No additional effect of a cognitive
7
8 parent training. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42, 1270-
9
10 1278.
- 11
12 Nathan, P. E., & Gorman, J. M. (Eds.) (2007). *A guide to treatments that work* (3rd ed.). New York:
13
14 Oxford University Press.
- 15
16 Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods
17
18 for literature reviews. *Annual Review of Psychology*, 52, 59-82.
- 19
20 Roth, A., & Fonagy, P. (2005). *What works for whom?: A critical review of psychotherapy research* (2nd
21
22 ed.). New York: Guilford Press.
- 23
24 Sofronoff, K., Attwood, T., & Hinton, S. (2005). A randomized controlled trial of a CBT intervention for
25
26 anxiety in children with Asperger syndrome. *Journal of Child Psychology and Psychiatry*, 46,
27
28 1152-1160.
- 29
30 *Spence, S. H., Donovan, C., & Brechman-Toussaint, M. (2000). The treatment of childhood social
31
32 phobia: The effectiveness of a social skills training-based, cognitive-behavioral intervention, with
33
34 and without parental involvement. *Journal of Child Psychology and Psychiatry*, 41, 713-726.
- 35
36 Sud, A. (1994). Attentional skills training/cognitive modeling: Short term therapeutic cognitive
37
38 interventions for test anxiety. *Psychological Studies*, 39, 1-7.
- 39
40 Sud, A., & Sharma, S. (1990). Two short-term, cognitive interventions for the reduction of test anxiety.
41
42 *Anxiety Research*, 3, 131-147.
- 43
44 *Webster-Stratton, C. (1984). Randomized trial of two parent-training programs for families with
45
46 conduct-disordered children. *Journal of Consulting and Clinical Psychology*, 52, 666-678.
- 47
48 *Webster-Stratton, C. (1990). Enhancing the effectiveness of self-administered videotape parent training
49
50 for families with conduct-problem children. *Journal of Abnormal Child Psychology*, 18, 479-492.
- 51
52 *Webster-Stratton, C. (1992). Individually administered videotape parent training: "Who benefits?"
53
54 *Cognitive Therapy and Research*, 1992, 16, 31-52.
- 55
56
57
58
59
60

- 1
2
3
4 *Webster-Stratton, C., & Hammond, M. (1997). Treating children with early-onset conduct
5
6 problems: A comparison of child and parent training interventions. *Journal of Consulting and*
7
8 *Clinical Psychology, 65*, 93-109.
9
- 10 *Webster-Stratton, C., Kolpacoff, M., & Hollinsworth, T. (1988). Self-administered videotape therapy for
11
12 families with conduct-problem children: Comparison with two cost-effective treatments and a
13
14 control group. *Journal of Consulting and Clinical Psychology, 56*, 558-566.
15
- 16 *Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-onset conduct
17
18 problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child*
19
20 *and Adolescent Psychology, 33*, 105-124.
21
- 22 Weisz, J. R., Hawley, K. M., & Jensen Doss, A. (2004). Empirically tested psychotherapies for youth
23
24 internalizing and externalizing problems and disorders. *Child and Adolescent Psychiatric Clinics*
25
26 *of North America, 13*, 729-815.
27
- 28 Weisz, J. R., Jensen Doss, A., & Hawley, K. M. (2005). Youth psychotherapy outcome research: A
29
30 review and critique of the evidence base. *Annual Review of Psychology, 56*, 337-363.
31
- 32 Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with
33
34 children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical*
35
36 *Psychology, 55*, 542-549.
37
- 38 Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with
39
40 children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological*
41
42 *Bulletin, 117*, 450-468.
43
- 44 Weisz, J. R., McCarty, C. A., & Valeri, S. M. (2006). Effects of psychotherapy for depression in children
45
46 and adolescents: A meta-analysis. *Psychological Bulletin, 132*, 132-149.
47
- 48 Williams, C. E., & Jones, R. T. (1989). Impact of self-instructions on response maintenance and
49
50 children's fear of fire. *Journal of Clinical Child Psychology, 18*, 84-89.
51
- 52 Woolfenden, S. R., & Williams, K., & Peat, J. (2001). Family and parenting interventions in children and
53
54 adolescents with conduct disorder and delinquency aged 10-17. *Cochrane Database of Systematic*
55
56 *Reviews*. New York: John Wiley & Sons.
57
58
59
60

Table 1	
<i>Description and Criteria of RPC Model Categories^a</i>	
<i>Category</i>	<i>Criteria</i>
Best Evidence for Change	At least 80% of the findings from three or more informants, measures, and analytic methods show differences, and at least three findings were gleaned from each of the informants, measures, and methods. There is no clear informant-specific, measure-specific, or method-specific pattern of findings. The evidence suggests the intervention successfully targets the construct.
Evidence for Probable Change	More than 50% of the findings from three or more informants, measures, and analytic methods show differences, and at least three findings were gleaned from each of the informants, measures, and methods. There is no clear informant-specific, measure-specific, or method-specific pattern of findings. The evidence suggests the intervention probably changes the targeted outcome domain, yet future work ought to examine why inconsistencies occurred.
Limited Evidence for Change	Either 50% or less of the findings from three or more informants, measures, and analytic methods show differences, or less than the grand majority (less than 80%) of findings from specific informant's ratings, measures, and/or methods show differences. Any differences found are either scattered across outcomes from multiple informants, measures, or methods, or are not found predominantly on outcomes from specific informants, measures, and/or methods. The evidence is inconclusive.
No Evidence for Change	No differences are observed. The evidence is completely inconclusive.
Evidence for Informant-Specific Change	Differences are found on the grand majority (80%) of ratings provided by specific informant(s), and at least three findings were gleaned from the informant(s) for which specificity of findings were observed. The evidence suggests the treatment might change the domain when it is exhibited in specific situations or in interactions with specific informant(s).
Evidence for Measure- or Method-Specific Change	Differences are found on the grand majority (80%) of specific measure(s) or analytic method(s), and at least three findings were gleaned from the measure(s) or method(s) for which specificity of findings were observed. The evidence suggests the intervention might change the domain when it is measured with specific kinds of measure(s), method(s), or both.

Note. ^a Adapted from De Los Reyes and Kazdin (2006) and De Los Reyes and Kazdin (2008). In the categories above, by “informants” we mean reporters of outcomes (e.g., self, spouse or significant other, clinician, laboratory observer, biological, institutional records); by “measures” we mean ways to assess outcomes (e.g., questionnaire or symptom-count measures, laboratory observations, diagnostic interviews); by “analytic methods” we mean statistical strategies (e.g., tests of mean differences, tests of diagnostic status). For a discussion on the development and rationale for the structure and criteria of the individual RPC Model categories, see De Los Reyes and Kazdin (2006).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 2	
<i>Description and Criteria of Effect Size Ranges to be Employed in Conjunction With Range of Possible Changes Model Categories</i>	
<i>Category</i>	<i>Criteria</i>
Below Small to Below Small	*Lower end includes any effect size below .20. *Upper end includes any effect size below .20.
Below Small to Small	*Lower end includes any effect size below .20. *Upper end includes any effect size greater than or equal to .20, but less than .50.
Below Small to Medium	*Lower end includes any effect size below .20. *Upper end includes any effect size greater than or equal to .50, but less than .80.
Below Small to Large	*Lower end includes any effect size below .20. *Upper end includes any effect size greater than or equal to .80.
Small to Small	*Lower end includes any effect size greater than or equal to .20, but less than .50. *Upper end includes any effect size greater than or equal to .20, but less than .50.
Small to Medium	*Lower end includes any effect size greater than or equal to .20, but less than .50. *Upper end includes any effect size greater than or equal to .50, but less than .80.
Small to Large	*Lower end includes any effect size greater than or equal to .20, but less than .50. *Upper end includes any effect size greater than or equal to .80.
Medium to Medium	*Lower end includes any effect size greater than or equal to .50, but less than .80. *Upper end includes any effect size greater than or equal to .50, but less than .80.
Medium to Large	*Lower end includes any effect size greater than or equal to .50, but less than .80. *Upper end includes any effect size greater than or equal to .80.
Large to Large	*Lower end includes any effect size greater than or equal to .80. *Upper end includes any effect size greater than or equal to .80.

Table 3							
<i>Demographic Characteristics of Cognitive-Behavioral Therapy (CBT) and Behavioral Parent Training (BPT) Studies Included in the Meta-Analysis</i>							
Study	Type of Sample	Pre-Treated Sample Size	Age Range of Total Sample	% Boys	Pre-Treated Sample Size of Experimental Conditions, Meta-Analysis	Number of Intervention Groups, Meta-Analysis	Number of Intervention Groups, Total Sample
<i>CBT Studies</i>							
Barrett et al. (1996)	Diagnosed outpatients, clinic-referred	79	7-14	56.96	54	1	2
Flannery-Schroeder & Kendall (2000)	Diagnosed outpatients, clinic-referred	45	8-14	51.51	45	2	2
Gallagher et al. (2004)	Diagnosed outpatients, recruited sample	23	8-11	47.83	23	1	1
Kendall (1994)	Diagnosed outpatients, clinic-referred	47	9-13	60.00	47	1	1
Kendall et al. (1997)	Diagnosed outpatients, clinic-referred	118	9-13	62.00	118	1	1
King et al. (2000)	Symptomatic outpatients, clinic-referred	36	5-17	30.56	24	1	2
Leal et al. (1981)	Symptomatic school sample	30	10 th grade students	N/A ^a	30	2	2
McMurray et al. (1986)	Symptomatic school sample	80	9-12	50.00	80	1	1
Spence et al. (2000)	Diagnosed outpatients, clinic-referred	50	7-14	62.00	33	1	2
<i>BPT Studies</i>							
Leung et al. (2003)	Symptomatic outpatients, clinic-referred	91	3-7	63.77	91	1	1
Webster-Stratton (1984)	Symptomatic outpatients, clinic-referred	35	3-8	71.43	24	1	2
Webster-Stratton et al. (1988)	Symptomatic outpatients, clinic-referred	114	3-8	69.30	114	3	3
Webster-Stratton (1990)	Symptomatic outpatients, clinic-referred	47	3-8	79.07	47	2	2
Webster-Stratton (1992)	Symptomatic outpatients, clinic-referred	100	3-8	72.00	100	1	1
Webster-Stratton & Hammond (1997)	Diagnosed outpatients, clinic-referred	97	4-8	74.23	48	1	3
Webster-Stratton et al. (2004)	Diagnosed outpatients, clinic-referred	159	4-8	90.00	57	1	5

Note. ^a Leal et al. (1981) did not provide this information.

Table 4								
<i>Methodological and Outcome Characteristics of Cognitive-Behavioral Therapy (CBT) and Behavioral Parent Training (BPT) Studies Included in the Meta-Analysis</i>								
Study	Intervention-Control Comparison	Number/ Informants	Number/ Measure Methods	Number/ Analytic Methods	Number/ Outcome Measures	Number Outcome Findings	Statistically Significant Findings (%)	Pre-Treatment Difference
<i>CBT Studies</i>								
Barrett et al. (1996)	ICBT vs. WL	3	2	2	4	5	2 (40.00)	No
Flannery-Schroeder & Kendall (2000)	ICBT vs. WL	3	1	1	6	6	4 (67.00)	No
	GCBT vs. WL	3	1	1	6	6	6 (100.00)	Yes ^a
Gallagher et al. (2004)	GCBT vs. WL	3	2	2	5	6	3 (50.00)	No
Kendall (1994)	ICBT vs. WL	3	2	2	6	7	5 (71.43)	Yes ^b
Kendall et al. (1997)	ICBT vs. WL	3	1	1	5	6	5 (83.33)	Yes ^c
King et al. (2000)	ICBT vs. WL	3	2	1	4	7	3 (42.86)	No
Leal et al. (1981)	GCBT (CM) vs. WL	2	2	1	3	3	0	No
	GCBT (SD) vs. WL	2	2	1	3	3	0	No
McMurray et al. (1986)	GCBT vs. Placebo	2	2	1	3	3	1 (33.33)	N/A ^d
Spence et al. (2000)	GCBT vs. WL	4	3	1	5	5	1 (20.00)	No
<i>BPT Studies</i>								
Leung et al. (2003)	GBPT vs. WL	1	1	1	3	4	4 (100.00)	Yes ^e
Webster-Stratton (1984)	GBPT vs. WL	2	2	1	3	5	3 (60.00)	No
Webster-Stratton et al. (1988)	GBPT (VM) vs. WL	3	2	1	4	8	7 (87.50)	No
	GBPT (GD) vs. WL	3	2	1	4	8	4 (50.00)	Yes ^f
	IBPT vs. WL	3	2	1	4	8	5 (62.50)	No
Webster-Stratton (1990)	IBPT (VM) vs. WL	3	2	1	4	5	1 (20.00)	No
	IBPT (VM/TC) vs. WL	3	2	1	4	5	0	No
Webster-Stratton (1992)	IBPT vs. WL	3	2	1	4	7	5 (71.43)	No
Webster-Stratton & Hammond (1997)	GBPT vs. WL	3	2	1	5	7	5 (71.43)	Yes ^g
Webster-Stratton et al. (2004)	GBPT vs. WL	4	2	1	3	3	2 (66.67)	No

53 Note. ICBT = Individual CBT; GCBT; Group CBT; CM = Cognitive Modification; SD = Systematic Desensitization; IBPT = Individual
54 BPT; GBPT = Group BPT; WL = Waitlist; VM = Video Modeling; GD = Group Discussion; TC = Therapist Consultation; ^a Three
55 measures were significant between conditions pre-intervention; ^b Two measures were significant between conditions pre-intervention; ^c
56 one measure was significant between conditions pre-intervention; ^d The authors reported employing outcome measures prior to
57 intervention to identify anxious youths to participate in the study, but did not report pre-intervention scores. However, the authors did not
58 report significant pre-intervention differences between conditions; ^e One measure was significant between conditions pre-intervention; ^f
59 one measure was significant between conditions pre-intervention; ^g One measure was significant between conditions pre-intervention.

60