

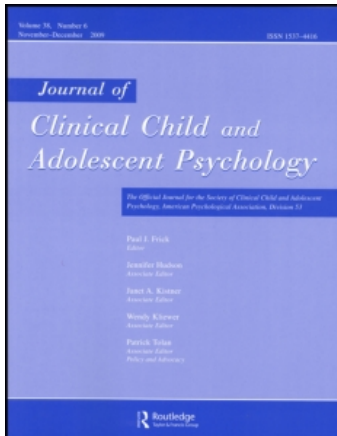
This article was downloaded by: [De Los Reyes, Andres][University of Maryland College Park]

On: 8 March 2011

Access details: Access Details: [subscription number 932222870]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Clinical Child & Adolescent Psychology

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t775648094>

Are Clinicians' Assessments of Improvements in Children's Functioning "Global"?

Andres De Los Reyes^a; Candice A. Alfano^b; Deborah C. Beidel^c

^a Department of Psychology, University of Maryland at College Park, ^b Center for Neuroscience Research, Children's National Medical Center, The George Washington University School of Medicine,

^c Department of Psychology, University of Central Florida,

Online publication date: 07 March 2011

To cite this Article De Los Reyes, Andres , Alfano, Candice A. and Beidel, Deborah C.(2011) 'Are Clinicians' Assessments of Improvements in Children's Functioning "Global"?', *Journal of Clinical Child & Adolescent Psychology*, 40: 2, 281 — 294

To link to this Article: DOI: 10.1080/15374416.2011.546043

URL: <http://dx.doi.org/10.1080/15374416.2011.546043>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Are Clinicians' Assessments of Improvements in Children's Functioning "Global"?

Andres De Los Reyes

Department of Psychology, University of Maryland at College Park

Candice A. Alfano

*Center for Neuroscience Research, Children's National Medical Center,
The George Washington University School of Medicine*

Deborah C. Beidel

Department of Psychology, University of Central Florida

In this study, the authors examined the relations among clinician ratings of treatment improvement and discrepancies between parent and blinded laboratory rater reports of child social functioning administered before and after treatment for social anxiety disorder. Participants included a clinic sample of 101 children (7–16 years old; $M = 11.67$, $SD = 2.57$; 51 girls, 81% Caucasian) receiving treatment as part of a two-site controlled trial. Overall, clinician ratings reflected lack of improvement when parents reported persistent (i.e., pre- to posttreatment) social functioning deficits not reported by blinded raters. However, when blinded raters reported persistent social skill deficits not reported by parents, we did not observe the same effect on clinician ratings as we did when the direction of discrepant reports was reversed. We replicated these observations in a subset of participants ($n = 81$) providing parent and child pre-post reports of social anxiety symptoms. These findings have implications for the interpretations of clinical ratings as "primary outcome measures" within controlled trials.

Clinical assessments commonly rely on multiple informants to assess a single person (e.g., self, significant other, clinician, and parents and teachers in the case of children). However, inconsistencies commonly arise across

multiple informants' reports of the same person (hereafter referred to as "informant discrepancies"; Achenbach, 2006). There is no definitive way to determine who is an "accurate" informant (De Los Reyes, 2011). Thus, it is important to understand the implications of informant discrepancies for drawing conclusions from research.

Informant discrepancies are particularly crucial to study when they occur within randomized controlled trials, which commonly rely on multiple informants' outcome reports (De Los Reyes & Kazdin, 2008; Weisz, Jensen Doss, & Hawley, 2005). For instance, recent meta-analytic work assessing the effects of psychological treatments for childhood depression indicates that, within studies using both parents and children as informants, effect sizes based on child-reported outcomes were three times larger than those based on parent-reported outcomes (Weisz, McCarty, & Valeri, 2006).

This research was supported in part by NIMH grant R01MH53703 to the third author. Lilly Corporation supplied the fluoxetine and matching placebo capsules. The authors acknowledge the efforts of Floyd R. Sallee, MD., PhD., and Robert T. Ammerman, PhD., PIs at the Cincinnati site, and Lori Crosby, Psy.D., project coordinator at the Cincinnati site. The authors also acknowledge Patricia Rao, Ph.D., Project Coordinator, Richard Gross, M.D., and Stephen Kwass, M.D., psychiatrists.

Clinical trial registration information URL: <http://www.clinicaltrials.gov>. Unique identifier: NCT00043537.

Correspondence should be addressed to Andres De Los Reyes, Comprehensive Assessment and Intervention Program, Department of Psychology, University of Maryland at College Park, Biology/Psychology Building, Room 3123H, College Park, MD 20742. E-mail: adelosreyes@psyc.umd.edu

These differences across studies translated into child-reported outcomes being significantly different from zero, whereas parent-reported outcomes were not significantly different from zero. There are profound clinical implications of such differences. Depending upon the informant chosen to determine treatment outcome, an intervention could be determined to be efficacious or nonefficacious. Therefore, these differences may influence whether (a) clinicians subsequently use the treatment in clinical practice and/or (b) future research is dedicated to studying the treatment. Although the data from Weisz and colleagues (2006) nicely illustrate this point, informant discrepancies are not specific to studies of treatments for childhood depression. In fact, they have long been found within studies of treatments for both children and adults (e.g., Casey & Berman, 1985; De Los Reyes & Kazdin, 2009; Koenig, De Los Reyes, Cicchetti, Scahill, & Klin, 2009; Lambert, Hatch, Kingston, & Edwards, 1986; Ogles, Lambert, Weight, & Payne, 1990). Despite frequent documentation of inconsistent outcomes within controlled trials, less clarity exists on how to interpret these inconsistencies.

Historically, informant discrepancies within controlled trials have typically been regarded as methodological nuisances that need to be “rectified” (e.g., Bird, Gould, & Staghezza, 1992; Piacentini, Cohen, & Cohen, 1992). Indeed, the ubiquitous presence of discrepancies has led many researchers to select one a priori measure to represent overall outcomes, deeming it a “primary outcome measure” (see Bowden et al., 2000; Hayward et al., 2000; Hazell & Stuart, 2003; Michelson et al., 2004; Wigal et al., 2004). In fact, when the results of a controlled trial are registered within a public database, a prerequisite for publication in journal outlets edited by members of the International Committee of Medical Journal Editors, definitions of the “primary” and “secondary” outcome measures must be included (De Angelis et al., 2004). Thus, use of methods to address informant discrepancies and in particular the “primary outcome measure” method has become a cornerstone of the design of controlled trials.

Use of methods to *account* for potentially inconsistent findings rests on the extent to which these methods may be affected by inconsistent reports. However, recent work suggests that even when a single informant is used to assess treatment outcome, the impressions of other informants may nonetheless affect the final judgment. For instance, “primary outcome measures” in treatment efficacy research tend to focus on whether treatment improved patients’ functioning and often rely on clinician ratings (e.g., Guy, 1976; Hamilton, 1960; Scahill et al., 1997; Shear et al., 2001; Young, Biggs, Ziegler, & Meyer, 1978). In addition, clinicians are typically blinded to treatment condition to ensure that their ratings are not unduly influenced by expectations of

improvement. Yet clinicians often base their ratings on interviews with patients and significant others, and on their informal observations of the patient. Therefore, they are not “blind” to the *informant(s)* on which they rely to determine treatment response.

The lack of clinician blindness to informants when determining outcome has crucial implications for how one interprets clinician ratings. Indeed, prior work suggests that clinicians commonly hold assumptions about the reliability of particular reports when assessing mental health concerns in children (Loeber, Green, & Lahey, 1990). In fact, parents and children rarely agree on what to target in treatment, and clinicians more often agree with the parent when the targeted problem deals with the child’s behavior (Hawley & Weisz, 2003).

It is important to note that when a clinical rating is a “primary outcome measure,” this typically implies that ratings gathered from this measure are treated as an index representing overall treatment outcome. That is, a primary outcome measure should represent treatment outcomes broadly construed, and not specific to particular contexts, situational constraints, or a single informant’s perspective. This is because data gathered from controlled trials, regardless of whether the method of outcomes evaluation is based on multiple reports or a primary outcome measure, are most often used to identify evidence-based treatments (De Los Reyes & Kazdin, 2006a). Thus, clinicians interview different informants because of their unique access to aspects of the child’s functioning (e.g., parents for behaviors expressed at home vs. teachers or children for behaviors expressed at school) and because of the well-accepted notion that the presence of informant discrepancies does not necessarily signify that any one of the informants’ reports is “right” or “wrong” (Achenbach, 2006; Comer & Kendall, 2004; Kraemer et al., 2003). Indeed, as mentioned previously to construct primary outcome measures clinicians make reports based on information collected in interviews with the parent and child (Birmaher et al., 2003; Compton et al., 2010). Yet, if clinicians nonetheless believe that certain informants are more reliable reporters than others, then clinician ratings might differentially reflect the reports of a specific informant. For instance, clinicians might base their reports more so on what the parent reports than what the child reports. Beyond the reliance on a single informant’s report, relying on parent reports relative to child reports has other disadvantages as well. That is, parent reports are not only often in disagreement with child reports; they often disagree with other adult informants and trained observer reports made under controlled conditions (e.g., Achenbach, McConaughy, & Howell, 1987). Ultimately, if clinician ratings represent only certain aspects of the child’s functioning, the claim that these ratings represent “global” changes in functioning is weakened. Thus, it is important

to examine whether clinician ratings reflect treatment response rates differentially or specific to particular informant's reports.

Controlled trials testing treatments for childhood social anxiety disorder are a prime context through which to address these issues. Indeed, researchers commonly observe high rates of informant discrepancies within clinic samples of anxious children at pretreatment, posttreatment, and follow-up assessments (Choudhury, Pimentel, & Kendall, 2003; De Los Reyes, Alfano, & Beidel, 2010; Grills & Ollendick, 2003; Rapee, Barrett, Dadds, & Evans, 1994; Safford, Kendall, Flannery-Schroeder, Webb, & Sommer, 2005). Most crucially, controlled trials vary widely in their use of multiple informants' reports, with clinician ratings often identified as a "primary outcome measure" within some trials (Beidel et al., 2007; Birmaher et al., 2003; Kendall, 1994; The Pediatric OCD Treatment Study Team, 2004; Silverman, Kurtines, Ginsburg, Weems, Lumpkin et al., 1999; Silverman, Kurtines, Ginsburg, Weems, Rabain et al., 1999; Wagner et al., 2004; Walkup et al., 2008; Weisz et al., 2005). As a result, one can test whether clinician ratings vary as a function of informant variation in views on treatment response.

In this study, we examined the relations among clinician ratings of improvement and informant discrepancies in reports of children's social functioning from pre- to posttreatment. We examined social functioning both to equate the valence of clinician ratings with that of informants' reports of children's behavior and because changes in "real-world" social functioning, and in particular social skill, is a core focus of treatments for childhood social anxiety disorder (see Beidel, Turner, & Morris, 2000). We addressed this question using data from a two-site controlled clinical trial of psychological and pharmacological treatments for childhood social anxiety disorder (Beidel et al., 2007). Recent work suggests that informant discrepancies are stable over the course of controlled trials and particularly so when clinicians identify a lack of treatment response (De Los Reyes et al., 2010). Of importance, clinicians within trials testing treatments for child anxiety base their ratings of treatment response solely on interviews with the parent and child (see Compton et al., 2010). This scenario may create opportunities for clinician ratings to represent some informants' impressions more so than other informants' impressions.

In this study, we predicted low-to-moderate levels of agreement across informants' reports over the course of treatment (Safford et al., 2005). Further, we predicted that clinician ratings would be more reflective of parent reports relative to other informants' reports (Hawley & Weisz, 2003). However, we also predicted that under some circumstances, parents and other informants would agree regarding treatment improvement (i.e.,

"agreement" informant dyads). Relative to "agreement" parent-observer dyads, we predicted that when parents persistently (both at pre- and posttreatment) reported deficits in children's social functioning not corroborated by blinded independent observers, clinicians would endorse less treatment improvement. Conversely, relative to "agreement" parent-observer dyads, when blinded independent observers reported persistent social deficits uncorroborated by the parent, we would not observe this same effect on clinicians' ratings of treatment improvement. Our primary analyses ($n = 101$) were based on parent and blind independent observers of children's social functioning in relation to clinician reports of improvements in global functioning. Further, we tested whether any observations made in relation to parent and independent observer reports could be replicated in relation to parent reports and reports of other informants. Thus, secondary analyses ($n = 81$) were based on a subset of participants providing parent and child reports of children's social anxiety. Last, we predicted these relations when taking into account pretreatment clinical severity as well as demographic characteristics, which sometimes relate to informant discrepancies (child age, gender, ethnicity/race; De Los Reyes & Kazdin, 2004, 2005, 2006b).

METHOD

Participants

Participants came from a sample of 139 parents and children who participated in a two-site controlled trial comparing pharmacological (fluoxetine) and behavioral (Social Effectiveness Therapy for Children; Beidel et al., 2000) treatments for childhood social anxiety disorder (see Beidel et al., 2007). Detailed information on demographic and clinical characteristics for the total sample has been reported elsewhere (Beidel et al., 2007). A flow chart detailing the participants included in the present study is presented in Figure 1. Briefly, we conducted the main tests of our hypotheses by examining a subsample of 122 children (7–16 years old; $M = 11.61$, $SD = 2.59$; 57 girls, 78% Caucasian, 1 participant did not provide ethnicity/race data). For the primary analyses detailed next (i.e., statistical modeling of parent-observer reporting discrepancies of children's social functioning) we examined 101 participants (7–16 years old; $M = 11.67$, $SD = 2.57$; 51 girls, 81% Caucasian, 1 participant did not provide ethnicity/race data). For the secondary analyses (i.e., statistical modeling of parent-child reporting discrepancies of children's social anxiety symptoms) we examined 81 participants (7–16 years old; $M = 11.75$, $SD = 2.57$; 39 girls, 70% Caucasian, 1 participant did not provide ethnicity/race data).

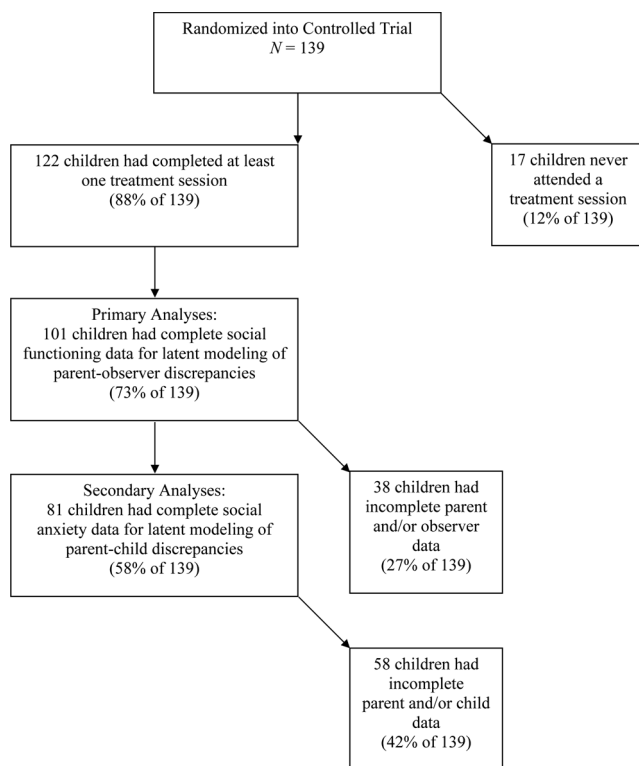


FIGURE 1 Flow chart of the total sample and the different subsamples employed in the statistical modeling of parent-observer pre-post reporting discrepancies of children's social functioning (Primary Analyses) and parent-child pre-post reporting discrepancies of children's social anxiety symptoms (Secondary Analyses). *Note:* Although 101 and 81 participants were included in the Primary and Secondary Analyses (respectively), 100 and 80 participants were included in tests of hypotheses based on Primary and Secondary Analyses samples because one participant did not provide ethnicity/race data.

We conducted exploratory analyses to examine whether the 122 participants in this study differed from the rest of the total sample at pretreatment as a function of demographic characteristics (child age, gender, ethnicity, socioeconomic status), pretreatment functioning (primary diagnosis type, clinical severity of primary diagnosis, global illness severity), or treatment condition. We conducted a large number of tests ($n = 8$) and did not have specific a priori hypotheses due to a lack of previous research. Thus, we set a predefined bonferroni-corrected p value threshold of .006 (i.e., $.05/8$). Across these eight tests, none of these factors evidenced a significant relation to inclusion/exclusion for this study. Similarly, running the eight tests comparing data used in statistical modeling of parent-observer discrepancies ($n = 101$) and parent-child discrepancies ($n = 81$) to the rest of the 122 participants yielded null effects.

Procedures

Study procedures were approved by the Institutional Review Boards at both universities. Parents gave

informed consent, and children gave informed assent. Recruitment occurred through media announcements (newspaper, radio) or through referrals from local mental health professionals. All children met criteria for a primary diagnosis of social anxiety disorder, generalized subtype (American Psychiatric Association, 2000) based on child and parent semistructured diagnostic interviews (Anxiety Disorders Interview Schedule for Children [ADIS-C/P]; Silverman & Albano, 1996). In the case of parent-child reporting discrepancies, clinicians arrived at a diagnosis through composite diagnostic procedures as outlined by Silverman and Albano (1996). In the present study, 97% of the parents who participated in the ADIS-P interview and completed all other assessments reported next were mothers and 3% were fathers.

Interviewers administering the ADIS-C/P were doctoral-level clinicians or advanced doctoral students who were trained in child anxiety disorders and administration of the ADIS-C/P. Interrater reliability for a social anxiety disorder diagnosis was strong ($\kappa = .78$). Other diagnoses were not diagnosed with sufficient frequency to allow ratings of interrater agreement. Interrater reliability for the ADIS-C/P social anxiety disorder clinical severity rating, which was used to identify relative impairment between primary and secondary diagnoses, was $r = .82$. Further information on procedures and treatments is available elsewhere (Beidel et al., 2007).

To ensure generalization of study findings, secondary comorbid diagnoses were allowed, with the exception of bipolar disorder, psychosis, conduct disorder, autism spectrum disorders, and mental retardation. Youth with moderate to severe depression who expressed active suicidal ideation or who had a previous unsuccessful trial of fluoxetine or behavior therapy were excluded. In the present sample, participants were included if parents and laboratory observers (primary analyses) and/or parents and children (secondary analyses) provided information at pretreatment and posttreatment on the child's social functioning and social anxiety symptoms (respectively). For all participants, clinicians needed to provide pretreatment clinical severity ratings and posttreatment clinical improvement ratings.

Measures

Parent reports of childhood social functioning, pre- and posttreatment. The Child Behavior Checklist (CBCL; Achenbach, 1991) consists of 118 items and was administered to assess parent reports of the child's social functioning. The CBCL Social Competence scale was used in this study and extensive evidence in support of the measure's reliability and validity has been reported elsewhere (Achenbach & Rescorla, 2001). For the current study, statistical modeling procedures described next necessitated use of dichotomized indices, and thus

we used the norm-referenced T score of 37, indicating that any score below 37 indicated clinically relevant problems with social competence. The CBCL authors recommend using scores below 37 to identify children in the clinical range of concerns with social competence (Achenbach & Rescorla, 2001). Specifically, at pretreatment those reports at or above the T score were coded 0 ($n = 70$) and below 37 were coded 1 ($n = 31$; scores based on 101 participants). At posttreatment, those reports at or above the T score were coded 0 ($n = 74$) and below 37 were coded 1 ($n = 27$; scores based on 101 participants).

Parent and child reports of childhood social anxiety symptoms, pre- and posttreatment. For the subset of the sample ($n = 81$) in which both parents and children provided pre- and posttreatment reports of children's social anxiety, both parents and children completed respective versions of the Social Phobia and Anxiety Inventory for Children (SPAIC; Beidel, Turner, & Morris, 1995). The scale consists of 26 items that assess a range of situations known to be anxiety provoking to children with social anxiety disorder. The SPAIC has acceptable levels of internal consistency and test-retest reliability (Silverman & Ollendick, 2005). At pretreatment, Cronbach's alpha coefficients were .95 for the parent-report items and .97 for the child-report items (alpha analyses based on complete item data for 77 parents and 79 children). At posttreatment, Cronbach's alpha coefficients were .96 for the parent-report items and .97 for the child-report items (alpha analyses based on complete item data for 75 parents and 78 children).

Observer reports of childhood social functioning, pre- and posttreatment. To assess social functioning, youths participated in five brief role-play interactions with a same-age peer trained to respond in a friendly but non-leading manner (Beidel, Turner, & Morris, 1999). Scene content included starting a conversation, offering help, giving a compliment, receiving a compliment, and responding assertively to inappropriate behavior. Undergraduate research assistants blinded to group status and phase of assessment rated the children's social skills. All raters were trained using practice tapes until they reached interrater reliability of $r > .80$. Raters were monitored throughout the course of the study to ensure that they stayed above $r > .80$. Raters made skill and anxiety ratings using 4-point Likert scales. Skill ratings ranged from 1 (*not effective at all*; "no response or one-word response, does not ask questions, mumbling, barely audible speech") to 4 (*effective*; "no awkwardness, carries part of the conversation, may self-disclose, voice strong and clear").

To determine interrater reliability, 25% of the assessments were rated by a second blind rater and yielded high

interrater reliability ($r = .84$). Assessments were conducted at both sites but all behavioral ratings were conducted at the clinic of the Principal Investigator (DCB). For the current study, we used a dichotomized version of the total score of the social skill/effectiveness rating and attempted to equate the frequencies of reports with those identified using the CBCL Social Competence scale. Specifically, at pretreatment those reports at or above the score of 2 were coded 0 ($n = 70$) and below 2 were coded 1 ($n = 31$) (scores based on 101 participants). At posttreatment, those reports at or above the score of 2 were coded 0 ($n = 82$) and below 2 were coded 1 ($n = 19$) (scores based on 101 participants).

Clinician ratings of pretreatment clinical severity and posttreatment global improvement. Clinician ratings of pretreatment clinical severity and posttreatment global improvement were assessed using the Clinical Global Impressions (CGI) Severity of Illness and Improvement Scale (Guy, 1976). The CGI was completed by an independent evaluator blinded to group status (12 weeks) and interrater reliability, conducted for 20% of the total sample, was $r = .86$. Like the ADIS-C/P interviewers, CGI raters were doctoral-level clinicians or advanced doctoral students who were trained in child anxiety disorders and in the administration of the CGI. Ratings were based on the evaluator's separate interviews with parent and child. Specifically, clinicians gave CGI ratings after conducting interviews with the parent and child. They did not have access to other information, such as ratings from the blinded social skill/effectiveness raters. This method is consistent with methods for attaining CGI ratings within controlled trials of child anxiety and with other clinical ratings of child anxiety used within controlled trials (e.g., Pediatric Anxiety Rating Scale; see Birmaher et al., 2003; Compton et al., 2010; Research Units on Pediatric Psychopharmacology Anxiety Study Group [RUPP], 2002; Wagner et al., 2004). For analyses reported below the continuous rating of pretreatment Severity of Illness ranges from 1 (*normal, not at all ill*) to 7 (*among the most extremely ill patients*). Similarly, the posttreatment Clinical Global Improvement scale ranges from 1 (*very much improved*) to 7 (*very much worse*).

Child demographic characteristics. Demographic variables (child age, gender, and ethnicity/race) were assessed with a parent-reported demographic packet and contact sheet.

Data Analyses

To identify parent-observer discrepancies in pre- to posttreatment reports of children's social functioning,

we conducted exploratory latent class analyses (LCAs; McCutcheon, 1987) on the four dichotomous parent and observer reports of social functioning.^{1,2} Like cluster analysis, LCA identifies groups of cases based on similar patterns of indicator variables. Like confirmatory factor analysis, LCA tests the absolute and relative fit of models yielding indices such as the Bayesian Information Criterion to compare whether a given model is a parsimonious solution to the data, with lower scores indicating greater parsimony (Raftery, 1986). Latent class analysis uses categorical or ordinal variables to produce classes within which there is local independence of indicators (i.e., indicator variables are statistically independent within levels of each latent class). Thus, LCA is a person-centered approach that allowed us to identify classes of children for whom there was conflicting information regarding social functioning over the course of treatment. Probabilities provided by a latent class solution may be used to assess the confidence with which cases are assigned (McCutcheon, 1987).

We expected that the LCA would identify the following latent classes of parent and observer reports: (a) low likelihood of social functioning deficits from pre- to posttreatment based on parent and observer reports, (b) low likelihood of social functioning deficits from pre- to posttreatment based on parent report but high likelihood of social functioning deficits from pre- to posttreatment based on observer report, and (c) low

likelihood of social functioning deficits from pre- to posttreatment based on observer report but high likelihood of social functioning deficits from pre- to posttreatment based on parent report. We tested one-through four-class solutions, evaluating the fit and interpretability of each.

A three-class solution fit the data best, $\chi^2(1) = 4.94$, $L^2 = 5.10$; and $\chi^2(1) = 2.62$, $L^2 = 4.29$. Although some of these statistics were still under the $p < .05$ threshold when the three-class solution was tested, estimating the four-class solution produced negative degrees of freedom. This signified that any model above three classes resulted in more parameter estimations than possible to estimate. In Table 1 we report the frequencies, percentages, and the mean probabilities of latent class assignment for the three-class solution. We also report the latent class probabilities and the conditional probabilities for each indicator. Mean probabilities of latent class assignment relate to the assignment probabilities for each parent–observer dyad within a particular latent class group; this probability is used as the key criterion by which a participant dyad is assigned to a particular latent class group (i.e., the highest observed probability rating across the three classes dictates a participant dyad’s class assignment). Thus, a key criterion by which researchers assess the suitability of model fit is the mean participant probability of assignment within each group; with a common metric being mean values of assignment probability above .70 (see Nagin, 2006). Across the three groups the mean probability of assignment was above .80, and the mean assignment probabilities within each group were all above .80. This suggests superb model fit in that latent class assignments were made with a great degree of confidence that dyads were assigned to groups within which they were (a) maximally similar in response patterns to other dyads assigned to their group and (b) maximally different in these response patterns from dyads assigned to the other groups.

Consistent with our hypotheses, we identified the following profiles: (a) low likelihood of social functioning deficits from pre- to posttreatment based on parent and observer reports (P+/O+), (b) low likelihood of social functioning deficits from pre- to posttreatment based on parent report but high likelihood of social functioning deficits from pre- to posttreatment based on observer report (P+/O-), and (c) low likelihood of social functioning deficits from pre- to posttreatment based on observer report but high likelihood of social functioning deficits from pre- to posttreatment based on parent report (P-/O+).

Following our identification of the latent class solutions, we tested their relations to clinician ratings of children’s global improvement. Specifically, using analyses of covariance, CGI Global Improvement score was entered as the dependent variable, and child age,

¹One potential concern with statistically modeling parent and observer reports of children’s social functioning is that the response formats across these reports were not parallel. The concern is that the nonparallel formats might introduce increased methodological error when studying the discrepancies between the reports. However, the correlation between pretreatment and posttreatment observer reports of children’s social functioning ($r = .53$) was similar in magnitude to the correlation between pretreatment and posttreatment parent reports of children’s social functioning ($r = .59$). Further, we calculated correlations between posttreatment observer and parent reports of children’s social functioning and posttreatment clinician ratings of global improvement (i.e., the assessment point at which one would observe the greatest variability among children across these variables). Both posttreatment observer reports and parent reports of children’s social functioning were significantly related to posttreatment clinician reports of global improvement ($r_s = -.25$ and $-.29$, respectively; both $p_s < .05$). These findings suggest that methodological error did not play a significant role in the findings that we report next.

²We considered use of a continuous form of latent modeling such as latent profile analyses (Bartholomew et al., 2002). However, our parent- and observer-reported indices varied widely in both scale variance and number of response options. Even if the variables were standardized before conducting latent profile analyses, there would have still been a substantial imbalance in the distribution of continuous scores for all of the measures used. Thus, these features of the data made LCA the most viable choice for statistical modeling. However, as we report below, for secondary analyses of parent–child pre–post reporting discrepancies we did have access to parallel parent- and child-report measures of childhood social anxiety symptoms. Thus, we conducted latent profile analyses on these data.

TABLE 1
Latent Class Solution of Parent-Observer Reporting Discrepancies of Child Social Functioning Deficits

<i>Latent Class</i>	<i>N</i>	<i>%</i>	<i>M Assignment Probability</i>
P+/O+	53	52.5	.83
P+/O-	28	27.7	.92
P-/O+	20	19.8	.88
Total	101	100	.86

<i>Measured Variable</i>	<i>Latent Classes</i>		
	<i>P+/O+</i>	<i>P+/O-</i>	<i>P-/O+</i>
	Parent-Report Social Functioning Deficits, Pre	Yes Deficits: .41 No Deficits: .49	Yes Deficits: 0 No Deficits: .48
Parent-Report Social Functioning Deficits, Post	Yes Deficits: .25 No Deficits: .55	Yes Deficits: .06 No Deficits: .43	Yes Deficits: .68 No Deficits: .01
Observer Report Social Functioning Deficits, Pre	Yes Deficits: .05 No Deficits: .65	Yes Deficits: .66 No Deficits: .19	Yes Deficits: .28 No Deficits: .15
Observer Report Social Functioning Deficits, Post	Yes Deficits: .01 No Deficits: .57	Yes Deficits: .74 No Deficits: .24	Yes Deficits: .24 No Deficits: .18

Note: $n = 101$. Conditional probabilities are to be interpreted across the row of a given indicator and within each value; probabilities sum to 100% in each row (e.g., on the indicator Observer Report Social Functioning Deficits, Post, "Yes Deficits" values across the row of three latent class "Yes Deficits" values total approximately 100% probability). P+/O+ = low probability of children's social functioning deficits across parent and observer reports; P+/O- = high probability of children's social functioning deficits based on observer and not parent report; P-/O+ = high probability of children's social functioning deficits based on parent and not observer report.

gender, and a dichotomous ethnicity/race variable (0 = Caucasian, $n = 81$; 1 = all other ethnicities, $n = 19$; 1 participant did not provide information for this variable) were entered as covariates. We also entered as a covariate the pretreatment CGI Severity of Illness score.

The key independent variable was the codes for latent class assignments for parent-observer pre- to posttreatment reporting discrepancies. The variable was coded as follows: P+/O+ = 1, P+/O- = 2, and P-/O+ = 3. Further, we conducted a priori group contrasts to examine whether the P+/O+ class evidenced greater improvement scores (i.e., lower CGI Global Improvement scores) relative to the classes typified by parent-observer discrepancies (P+/O-; P-/O+). Here, if

clinicians weigh all informants' reports equally, children in the classes typified by parent-observer discrepancy should each be rated as exhibiting significantly fewer improvements (higher CGI Global Improvement scores), relative to children in the P+/O+ class.

RESULTS

Preliminary Analyses

Frequency distributions for all variables did not reveal any deviations from normality. Means and standard deviations for the measures are presented in Table 2.

TABLE 2
Means and Standard Deviations of Measures for Primary and Secondary Analyses

<i>Assessment Point</i>	<i>Informant</i>		
	<i>Parent</i>	<i>Observer</i>	<i>Clinician</i>
Primary Analyses ^a			
Pretreatment	CBCL Social Competence: 40.11 (7.92)	Role-Play Social Effectiveness: 2.10 (.82)	CGI Severity of Illness: 4.71 (.64)
Posttreatment	CBCL Social Competence: 41.82 (8.38)	Role-Play Social Effectiveness: 2.33 (.81)	CGI Global Improvement: 2.34 (1.14)
Secondary Analyses ^b			
Pretreatment	SPAIC Parent Version: 34.57 (8.50)	SPAIC: 27.75 (8.89)	CGI Severity of Illness: 4.72 (.63)
Posttreatment	SPAIC Parent Version: 24.69 (10.43)	SPAIC: 16.49 (10.84)	CGI Global Improvement: 2.07 (1.19)

Note: CBCL = Child Behavior Checklist; CGI = Clinical Global Impressions Severity of Illness and Improvement Scale; SPAIC = Social Phobia and Anxiety Inventory for Children.

^a $n = 101$.

^b $n = 81$.

Kappa coefficients were calculated between the dichotomous variables used in the parent–observer LCA at pre- and posttreatment. Consistent with prior work (De Los Reyes & Kazdin, 2004, 2005, 2006b), tests revealed low levels of agreement between parents and observers at both assessment points: pretreatment $\kappa = -.02$; posttreatment $\kappa = .05$.

To examine differences in pretreatment clinical severity, we compared the latent classes via one-way analysis of variance on pretreatment CGI Severity of Illness scores. There were nonsignificant group differences for the latent class groups ($p > .09$). Similarly, chi-square analyses revealed nonsignificant relations between the latent class groups and treatment condition assignment ($p > .55$). This suggests that the predictive findings presented next resulted from neither pretreatment differences among the latent class groups on children's clinical severity nor type of treatment.

Primary Analyses

Relative to the P+/O+ class, we hypothesized that parent–observer reporting discrepancies of children's social functioning would predict variance in clinicians' ratings of posttreatment global improvement. In particular, we hypothesized that clinician ratings of children in the P-/O+ class would be higher than the clinician ratings of children in the P+/O+ class. For these analyses, we accounted for child age, gender, and ethnicity/race as well as clinician ratings of pretreatment clinical severity via the CGI. Results are presented in Table 3. Only one covariate (CGI pretreatment Severity of Illness score) significantly predicted clinician ratings of global improvement at posttreatment. Consistent with our hypotheses, latent classes of parent–observer discrepancies in reports of children's social functioning deficits predicted variance in clinician ratings of global improvement at posttreatment over and above the control variables.

To examine the direction of observed effects, we compared the marginal means (i.e., means when accounting for all covariates) of the CGI Global Improvement scores at posttreatment, using the P+/O+ class as an a priori contrast class. As shown in Table 3 and consistent with our hypotheses, the P-/O+ class evidenced significantly greater CGI Global Improvement scores at posttreatment (i.e., less improvement), relative to the P+/O+ class. However, CGI Global Improvement scores at posttreatment for children in the P+/O- class were not significantly different from scores for children in the P+/O+ class.

Secondary Analyses

In a subset of participants for whom complete parent and child data were available on children's social anxiety symptoms ($n = 81$), we tested whether we could (a)

replicate a similar three-group solution to that observed with parent–observer reports and (b) replicate the findings observed with parent–observer pre–post reporting discrepancies. A substantial reduction in the number of participants included in secondary analyses (see Figure 1) severely limited our power to detect significant effects. Further, the hypotheses we tested were directional and based on our primary analyses. Thus, we interpreted secondary analyses based on one-tailed significance tests.

We modeled parent–child reporting discrepancies by conducting a latent profile analysis (LPA) on parent and child pre- and posttreatment children's social anxiety reports (Bartholomew, Steele, Moustaki, & Galbraith, 2002). Latent profile analysis focuses on continuous indicators; these procedures are a generalization of the LCA procedure used to model parent–observer discrepancies, which uses categorical or ordinal variables

TABLE 3
Univariate Analysis of Covariance Contrasting Profile Groups of Parent–Observer Reporting Discrepancies on Levels of Children's Social Functioning Deficits at Posttreatment

Variable	CGI Global Improvement, Post				Observed Power (.05)	
	df	Partial η^2	F	p		
Total Model	6	.12	2.16	.054		
Child Age	1	.01	.97	.32		
Child Gender	1	0	.16	.69		
Child Ethnicity/Race	1	.01	1.35	.25		
CGI Severity of Illness, Pre	1	.05	5.04	.02*		
Profile Group of Parent–Observer Discrepancies	2	.07	3.65	.03*	.66	
<i>Follow-Up Group Contrasts</i>						
	M	SE	CE	SE	p	[95% CI]
P-/O+	2.84	.25				
vs.						
P+/O+	2.07	.15				
Contrast Statistics			.77	.30	.01*	[-.17, 1.37]
P+/O-	2.48	.22				
vs.						
P+/O+	2.07	.15				
Contrast Statistics			.41	.27	.13	[-.13, .96]

Note: $n = 100$. P+/O+ = low probability of reports of children's social functioning deficits across parent and observer reports; P+/O- = high probability of children's social functioning deficits based on observer and not parent report; P-/O+ = high probability of children's social functioning deficits based on parent and not observer report; p -values reported in table based on two-tailed tests; Follow-up group contrasts based on estimated marginal means and standard errors; CE = Contrast Estimate; SE = Standard Error; CI = Confidence Interval; Child gender was coded as 0 = male, 1 = female. Child ethnicity was coded as 0 = Caucasian, 1 = all other ethnicities; CGI = Clinical Global Impressions Severity of Illness and Improvement Scale.

* $p < .05$.

(McCutcheon, 1987). Here, we ran a three-profile LPA and assessed model fit by evaluating whether the mean participant probabilities of assignment within each group were above .70 (see Nagin, 2006).

As seen in Table 4, the mean probability of assignment was above .90, and the mean assignment probabilities within each group were all above .90. This suggests excellent model fit, consistent with modeling for parent–observer pre–post reporting discrepancies. Although the general structure of the LPA for parent–child reporting discrepancies differed from the structure of the LCA for parent–observer reporting discrepancies, a pattern typified by discrepant parent and child reports nonetheless

emerged. Specifically, three groups were identified: (a) pre- to posttreatment improvements in children’s social anxiety symptoms based on parent and child reports (P+/C+), (b) pre- to posttreatment improvements in children’s social anxiety symptoms based on parent report but not child report (P+/C–), and (c) pre- to posttreatment improvements in children’s social anxiety symptoms based on child report but not parent report (P–/C+).

Correlations computed between child and parent total SPAIC scores at pretreatment ($r = .49, p < .001$) and posttreatment ($r = .30, p < .01$) revealed significant but low-to-moderate correspondence between parent and child ratings, consistent with previous investigations (Achenbach, 2006; De Los Reyes & Kazdin, 2005). Similar to the latent classes observed for parent–observer

TABLE 4

Latent Class Solution of Parent-Child Reporting Discrepancies of Children's Social Anxiety

Latent Class	N	%	M Assignment Probability
P+/C+	62	76.5	.99
P+/C–	11	13.6	.98
P–/C+	8	9.9	.96
Total	81	100	.99

Conditional Probabilities for Measured Variables				
Measured Variable	Conditional Value	Latent Classes		
		P+/C+	P+/C–	P–/C+
Parent Social Anxiety, Pre	10.87–27.67	.50	0	.49
	28.27–33.22	.92	.06	.01
	33.50–37.02	.94	.05	0
	37.83–42.40	.75	.24	0
	42.42–51.00	.68	.31	0
Parent Social Anxiety, Post	2.00–14.00	.85	.06	.08
	14.33–21.67	.65	.05	.29
	21.73–27.97	.81	.11	.07
	28.08–33.53	.75	.19	.05
	34.43–47.00	.75	.24	0
Child Social Anxiety, Pre	10.12–19.07	.50	0	.50
	20.27–25.83	.92	.06	.01
	26.20–29.53	.99	0	0
	29.65–35.98	.93	.06	0
	36.35–48.87	.44	.56	0
Child Social Anxiety, Post	0–7.733	.69	0	.30
	8.033–11.67	.85	0	.14
	11.73–16.55	.94	0	.06
	17.60–24.67	.99	0	0
	24.73–43.35	.31	.68	0

Note: $n = 81$. Conditional probabilities are to be interpreted across the row of a given indicator and within each value; probabilities sum to 100% in each row. Within each cell, the highest observed conditional probabilities are bolded for emphasis (standard bold = P+/C+; underline and bold = P+/C–; italics and bold = P–/C+). P+/C+ = high probability of reductions in reports of children’s social anxiety across parent and child reports; P+/C– = high probability of parent report of children’s social anxiety evidencing improvement and not child report; P–/C+ = high probability of child report of children’s social anxiety evidencing improvement and not parent report.

TABLE 5

Univariate Analysis of Covariance Contrasting Profile Groups of Parent-Child Reporting Discrepancies on Levels of Children's Social Anxiety at Posttreatment

Variable	df	CGI Global Improvement, Post			Observed Power (.10)
		Partial η^2	F	p	
Total Model	6	.08	1.12	.17	
Child Age	1	0	.01	.46	
Child Gender	1	0	.28	.29	
Child Ethnicity/ Race	1	0	.40	.26	
CGI Severity of Illness, Pre	1	.03	2.53	.055	
Profile Group of Parent–Child Discrepancies	2	.05	2.04	.065	.54

Follow-Up Group Contrasts	M	SE	CE	SE	p	[90% CI]
vs. P+/C+	1.93	.15				
Contrast Statistics			.83	.45	.03*	[.08, 1.58]
P+/C– vs. P+/C+	2.41	.40				
Contrast Statistics			.47	.44	.14	[–.25, 1.20]

Note: $n = 80$. P+/C+ = high probability of reductions in reports of children’s social anxiety across parent and child reports; P+/C– = high probability of parent report of children’s social anxiety evidencing improvement and not child report; P–/C+ = high probability of child report of children’s social anxiety evidencing improvement and not parent report; p-values reported in table based on one-tailed tests; Follow-up group contrasts based on estimated marginal means and standard errors; CE = Contrast Estimate; SE = Standard Error; CI = Confidence Interval. Child gender was coded as 0 = male, 1 = female. Child ethnicity was coded as 0 = Caucasian, 1 = all other ethnicities; CGI = Clinical Global Impressions Severity of Illness and Improvement Scale.

* $p < .05$.

pre-post discrepancies, there were nonsignificant group differences for the latent profile groups on pretreatment clinical severity ratings from the CGI ($p > .09$). Similarly, chi-square analyses revealed nonsignificant relations between the latent class groups and treatment condition assignment ($p > .25$).

The analysis of covariance tests for secondary analyses were conducted similarly to those conducted for the primary analyses. One difference is that the independent variable was the codes for latent profile assignments for parent-child pre- to posttreatment reporting discrepancies. The variable was coded as follows: P+/C+=1, P+/C-=2, and P-/C+=3. Further, we conducted a priori group contrasts to examine whether the P+/C+ group evidenced greater improvement scores (i.e., lower CGI Global Improvement scores) relative to the P+/C- and the P-/C+ groups.

As seen in Table 5, results for parent-child pre-post reporting discrepancies were consistent with those for parent-observer pre-post reporting discrepancies. There were significant trends for both the CGI pretreatment Severity of Illness score and the latent profile grouping of parent-child discrepancies for predicting variance in clinician ratings of global improvement at posttreatment. In addition, consistent with the primary analyses the P-/C+ profile evidenced significantly greater CGI Global Improvement scores at posttreatment, relative to the P+/C+ profile. However, CGI Global Improvement scores at posttreatment for children in the P+/C- profile were not significantly different from scores for children in the P+/C+ profile.

DISCUSSION

Based on data from a two-site controlled treatment trial for childhood social anxiety disorder, there were three main findings. First, we identified distinct patterns of pre-post reporting discrepancies between parent and independent observer reports of children's social functioning. Second, when parents reported persistent deficits in children's social functioning following treatment that were not corroborated by other informants' reports, clinician ratings most closely reflected parent reports. Yet, the reverse was not true: When laboratory observers reported persistent deficits in children's social functioning following treatment that were not corroborated by parents, clinician ratings were again more in line with parent reports. Third, these observations were replicated when examining parent reports in relation to child reports and when the behavior upon which clinician reports were based represented the primary target of treatment (social anxiety symptoms). All hypothesis tests controlled for child demographic characteristics and the child's pretreatment clinical severity.

These findings are important for three reasons. First, the latent modeling groups representing the discrepancies between parent and child and parent and observer reports did not significantly differ from each other in terms of children's pretreatment clinical severity or the treatment condition to which they were assigned. Second, the findings are in line with prior work showing that clinicians hold particular views as to which informants are "more reliable" informants of child behavior (Loeber et al., 1990), with clinicians agreeing more with parents than the child himself or herself when parent and child reports disagree (Hawley & Weisz, 2003). Taken together, the findings suggest that clinician ratings of children's improvement more reliably reflect parents' perspectives of their child's functioning relative to other informants.³

Because clinicians did not have access to the observer ratings, our findings regarding parent-observer discrepancies are not entirely surprising. However, clinician ratings were more representative of parent reports, even as compared to the reports of children who were interviewed as part of the evaluation. As mentioned previously, clinician ratings like the CGI are often treated as a measure of "global" functioning and used as "primary outcome measures" in controlled trials. By construction, such ratings are not global reports because the clinician relies only on a limited number of informants to create their ratings (e.g., parent and child interviews). In addition, even if trained clinicians' ratings of children are based only on limited numbers of informants, it is certainly reasonable to expect that if these clinicians' ratings of children are truly "global" and incorporate their own clinical impressions, then these ratings should be associated with other blinded, trained observers' ratings of the same children being rated by the clinicians.

³We considered the possibility that, given the nature of the social anxiety disorder diagnosis, clinicians may be more reluctant to rely on children to make clinical decisions. This would imply that the effects of clinicians preferring parent report over child report would be particularly strong for child social anxiety concerns over other child concerns. However, clinicians agree more with parent report over child report across numerous child psychological concerns (Hawley & Weisz, 2003; Loeber et al., 1990). That is, there does not appear to be a differential effect of clinician-parent versus clinician-child agreement on reports of child social anxiety versus other child psychological concerns. Further, this possibility would not be consistent with the plethora of psychometric work in the clinical literature documenting that children are reliable and valid reporters of their own social anxiety concerns (Silverman & Ollendick, 2005). In addition, in our secondary analysis subsample ($n=81$) we found that children's pretreatment self-reports on the SPAIC correlated just as highly with the pretreatment ADIS-C/P clinician severity ratings of the child's social anxiety disorder diagnosis as parents' pretreatment reports on the SPAIC (both $r_s = .35$, $p_s < .01$). Thus, prior work and our own data do not suggest that this is a probable explanation of the findings.

To our knowledge, this is the first study to empirically demonstrate that in practice, the clinician systematically relies on the parent report to form their evaluations of treatment response. We demonstrated this reliance by examining parent reports relative to the other informant who the clinician interviewed (child) as well as relative to a trained observer's report to which the clinician did not have access (independent rater). Thus, the availability of behavioral observations aside from reports provided by the parent and child is a major strength of this study.

Limitations

There are limitations to the present study. First, although a key strength was how we collected information on children's behavior from parents, children, and laboratory observers, the response formats were not parallel across measures taken from parents and laboratory observers (although they were parallel across parent and child reports). This could have impacted the results of dichotomizing scores used in latent class modeling of parent and laboratory observer data (however, see footnote 1). Prior work suggests that these informants' reports reflect treatment response and condition differences in outcome (i.e., differences between treatment and control conditions; see Beidel et al., 2007). At the same time, future research should replicate and extend this research using measures with consistent response formats.

Second, we did not directly assess informants' perceptions of treatment improvements independently from clinician ratings. As such, beyond observing changes in scores from pre- to posttreatment, we were unable to account for whether direct assessments of improvements from the perspectives of parents, children, and independent observers would have resulted in similar conclusions. We encourage future work to replicate and extend our findings using multiple informants' direct assessments of treatment improvement.

Third, we did not assess factors that may have explained the relation between clinician ratings of global improvement and its links to parent reports versus reports of other informants. In particular, one often studied factor is parents' levels of emotional distress (De Los Reyes & Kazdin, 2005). Researchers often posit that parental emotional distress, such as depressed mood, may lead to parents attending to, encoding, and remembering negative child behaviors (as opposed to positive or neutral behaviors), and predominantly using these remembrances to provide reports (Richters, 1992; Youngstrom, Izard, & Ackerman, 1999). In turn, if high levels of parental emotional distress lead to parents reporting more problems than other informants, parental mood, and not the child's behavior, might lead to higher clinician-parent agreement. However, existing

data do not support this assumption. In fact, whether clinicians agree more with parents relative to children on problems that ought to be targeted in child therapy does not relate to a parents' level of psychopathology (Hawley & Weisz, 2003). Furthermore, when examining latent patterns of parent-child reporting discrepancies revealed across multiple reports of children's behavior and emotional problems, there is no relation between these discrepancies and parental mood levels (depressive and manic symptoms), when considering other characteristics previously identified as correlates of these discrepancies (e.g., child mood levels, family functioning, child clinical severity, child demographic characteristics; De Los Reyes et al., 2011). In any event, future research seeking to replicate and extend our findings should examine potential mechanisms explaining our observed effects.

Fourth, we tested our hypotheses with a clinician rating scale that assesses global improvements in functioning, broadly construed (CGI; Guy, 1976). These findings may not replicate with clinician measures that assess for improvements in functioning or symptom reduction for specific clinical symptoms (see Hamilton, 1960; Scahill et al., 1997). Symptom-specific scales may be better equipped to reflect multiple informants' reports over the course of treatment, particularly if each of the informants' reports were taken as changes in the same condition. Future research should address these questions using a variety of clinician rating scales that themselves differ in whether they were designed to assess improvements in global functioning versus specific to particular diagnostic presentations.

Fifth, researchers have a number of different methods of information gathering to complete ratings on the CGI. For instance, the version of the CGI used in this study (based solely on information from parent and child interviews) is the key method used to gauge treatment response using the CGI in controlled trials of child anxiety (see Compton et al., 2010), as well as within treatment literatures for adult disorders (e.g., Spearing, Post, Leverich, Brandt, & Nolen, 1997; Zaider, Heimberg, Fresco, Schneier, & Liebowitz, 2003). In addition, this method is the one used for forming ratings for other clinician ratings of child anxiety used in controlled trials (RUPP, 2002). At the same time, the CGI as originally developed allows for the collection and use of information gathered outside a clinical interview, and one can identify examples in the literature of researchers supplementing patient interviews with collateral information (e.g., Guy, 1976; Knopman, Knapp, Gracon, & Davis, 1994; Niederhofer, Staffen, & Mair, 2003). Therefore, it is unclear whether our findings would have changed had we used another method by which to collect information on which to base the CGI rating.

Last, we included participants in the sample regardless of the number of treatment sessions that they

completed, so long as they completed at least one session. This was consistent with the intent-to-treat analytic methodology used in the original controlled trial from which we conducted this study (i.e., if participant drops out with missing data, last data point recorded carried forward to the next assessment point analyzed; Beidel et al., 2007). We considered it possible that this variability in participants' treatment attendance played a role in our observed effects, but not probable. Indeed, in our analyses we controlled for children's pretreatment clinical severity. Presumably, children attending fewer sessions would have been less likely to evidence responses to treatment. That is, if this was a factor that accounted for variability in outcomes, then our analyses would have been biased against observing variability among informants' reports of changes over the course of treatment. Further, as reported in the original trial less than 13% ($n=16$) of the 122 children in the intent-to-treat sample did not complete treatment (Beidel et al., 2007); this figure is well below the expected attrition rate for child therapy (Nock & Kazdin, 2005). Nevertheless, future research in samples that have higher rates of attrition should examine whether attrition moderates the effects that we observed in this sample.

Implications for Research, Policy, and Practice

Our findings have important implications for interpreting clinical outcomes. Indeed, multiple informants are commonly used within controlled treatment trials, with inconsistent findings being the rule rather than the exception. In response, controlled trials have varied widely in their use of multiple informants, with some studies basing determinations of treatment outcomes on single or "primary outcome measures." In many cases, these measures are based on clinician reports, which in turn are based on information from one or more informants. Controlled trials take great strides to ensure clinician blindness to treatment characteristics. Yet what is unknown is how assessments of treatment response are impacted by the fact that clinicians are not blind to the informants upon which they rely to make these evaluations. This is a crucial issue, because prior work suggests that clinicians' decisions are influenced by their beliefs about the veracity of different informants and their reports (Hawley & Weisz, 2003; Loeber et al., 1990). It is important to note that this is not an issue specific to research settings because practitioners often observe the same kinds of informant discrepancies (Yeh & Weisz, 2001).

In line with prior work, the findings suggest that when informants disagree about changes in a patient's functioning, clinician ratings may reflect persistent deficits (or lack of improvements) in the child's

functioning when based on certain informants' reports (i.e., parents). If clinician reports of outcome are to be interpreted specific to the sources of information upon which they are based, then this might not be of great concern. However, these ratings, particularly when interpreted as "primary outcome measures" within controlled trials, are often used as the *sole basis* by which to establish a treatment as efficacious. Further, treatment efficacy is rarely qualified with regard to the nature of the outcome evidence (De Los Reyes & Kazdin, 2006a, 2008). Thus, when clinicians base their ratings on a subset of all available informants, such ratings, by definition, may not represent "global improvements in functioning."

The issues raised by clinician ratings in relation to informants' reports might be a function of the information to which clinicians have access when forming their impressions of treatment response. That is, would clinicians' ratings still differentially relate to parent reports versus other informants if clinicians had systematic access to (a) multi-informant reports apart from the CGI interview, and (b) evidence that these other reports yield reliable and valid information? For example, investigators might instruct clinicians to conduct the interviews that typically form the basis for CGI ratings (i.e., parent and child interviews). However, investigators might also provide clinicians with the outcomes of an assessment battery containing multiple informants' reports (e.g., parent, child, independent observer, official records). To accompany these outcomes, the investigators might include a summary document (1 page) of research demonstrating the reliability and validity of the reports in the battery. Investigators could counterbalance these assessment components (i.e., CGI first vs. battery outcomes first) and instruct clinicians to provide CGI ratings only after the entire assessment process is complete. In comparison with the CGI as traditionally administered (i.e., single interview with parent and child and no other information), investigators could examine whether this modified version of the CGI reduces the differential relations between clinician ratings and parent reports relative to reports of other informants. Indeed, this modified procedure would be consistent with theoretical and empirical work indicating that multiple informants' reports and, in particular, the discrepancies between their reports yield valuable information on the circumstances in which children express specific behaviors (Achenbach, 2006; De Los Reyes, 2011; De Los Reyes, Henry, Tolan, & Wakschlag, 2009; Kraemer et al., 2003). In sum, we encourage both researchers and practitioners to systematically examine whether such alterations to existing practices in making ratings of patients' improvements reduce the impact that informant discrepancies have on clinical ratings.

REFERENCES

- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist 14-18 and 1991 profile*. Burlington: University of Vermont.
- Achenbach, T. M. (2006). As others see us: Clinical and research implications of cross-informant correlations for psychopathology. *Current Directions in Psychological Science*, 15, 94-98.
- 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.
- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington: University of Vermont, Research Center for Children, Youth, & Families.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2002). *The analysis and interpretation of multivariate data for social scientists*. Philadelphia: Chapman & Hall/CRC.
- Beidel, D. C., Turner, S. M., & Morris, T. L. (1995). A new inventory to assess childhood social anxiety and phobia: The Social Phobia and Anxiety Inventory for Children. *Psychological Assessment*, 7, 73-79.
- Beidel, D. C., Turner, S. M., & Morris, T. L. (1999). Psychopathology of childhood social phobia. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 643-650.
- Beidel, D. C., Turner, S. M., & Morris, T. L. (2000). Behavioral treatment of childhood social phobia. *Journal of Consulting and Clinical Psychology*, 68, 1072-1080.
- Beidel, D. C., Turner, S. M., Sallee, F. R., Ammerman, R. T., Crosby, L. A., & Pathak, S. (2007). SET-C versus fluoxetine in the treatment of childhood social phobia. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46, 1622-1632.
- 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.
- Birmaher, B., Axelson, D. A., Monk, K., Kalas, C., Clark, D. B., Ehmann, M., et al. (2003). Fluoxetine for the treatment of childhood anxiety disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42, 415-423.
- 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.
- Casey, R. J., & Berman, J. S. (1985). The outcomes of psychotherapy with children. *Psychological Bulletin*, 98, 388-400.
- Choudhury, M. S., Pimentel, S. S., & Kendall, P. C. (2003). Childhood anxiety disorders: Parent-child (dis)agreement using a structured interview for the DSM-IV. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42, 957-964.
- Comer, J. S., & Kendall, P. C. (2004). A symptom-level examination of parent-child agreement in the diagnosis of anxious youth. *Journal of the American Academy of Child and Adolescent Psychiatry*, 43, 878-886.
- Compton, S. N., Walkup, J. T., Albano, A. M., Piacentini, J. C., Birmaher, B., Sherill, J. T., et al. (2010). Child/Adolescent Anxiety Multimodal Study (CAMS): Rationale, design, and methods. *Child and Adolescent Psychiatry and Mental Health*, 4, 1-15.
- De Angelis, C., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., et al. (2004). Clinical trial registration: A statement from the International Committee of Medical Journal Editors. *New England Journal of Medicine*, 351, 1250-1251.
- De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 40, 1-9.
- De Los Reyes, A., Alfano, C. A., & Beidel, D. C. (2010). The relations among measurements of informant discrepancies within a multisite trial of treatments for childhood social phobia. *Journal of Abnormal Child Psychology*, 38, 395-404.
- De Los Reyes, A., Henry, D. B., Tolan, P. H., & Wakschlag, L. S. (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology*, 37, 637-652.
- 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.
- De Los Reyes, A., & Kazdin, A. E. (2006a). 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. (2006b). Informant discrepancies in assessing child dysfunction relate to dysfunction within mother-child interactions. *Journal of Child and Family Studies*, 15, 643-661.
- 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.
- De Los Reyes, A., & Kazdin, A. E. (2009). Identifying evidence-based interventions for children and adolescents using the range of possible changes model: A meta-analytic illustration. *Behavior Modification*, 33, 583-617.
- De Los Reyes, A., Youngstrom, E. A., Pabón, S. C., Youngstrom, J. K., Feeny, N. C., & Findling, R. L. (2011). Internal consistency and associated characteristics of informant discrepancies in clinic referred youths age 11 to 17 years. *Journal of Clinical Child and Adolescent Psychology*, 40, 36-53.
- Grills, A. E., & Ollendick, T. H. (2003). Multiple informant agreement and the Anxiety Disorders Interview Schedule for Parents and Children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42, 30-40.
- Guy, W. (1976). *ECDEU assessment manual for psychopharmacology*. Washington, DC: DHEW.
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, & Psychiatry*, 23, 56-62.
- Hawley, K. M., & Weisz, J. R. (2003). Child, parent, and therapist (dis)agreement on target problems in outpatient therapy: The therapist's dilemma and its implications. *Journal of Consulting and Clinical Psychology*, 71, 62-70.
- 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.
- Kendall, P. C. (1994). Treating anxiety disorders in children: Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 62, 100-110.
- Koenig, K., De Los Reyes, A., Cicchetti, D., Scahill, L., & Klin, A. (2009). Group intervention to promote social skills in school-age children with pervasive developmental disorders: Reconsidering efficacy. *Journal of Autism and Developmental Disorders*, 39, 1163-1172.

- Knopman, D. S., Knapp, M. J., Gracon, S. I., & Davis, C. S. (1994). The Clinician Interview-Based Impression (CIBI): A clinician's global change rating scale in Alzheimer's disease. *Neurology*, *44*, 2315–2321.
- 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.
- 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.
- Loeber, R., Green, S. M., & Lahey, B. B. (1990). Mental health professionals' perception of the utility of children, mothers, and teachers as informants of childhood psychopathology. *Journal of Clinical Child Psychology*, *19*, 136–143.
- McCutcheon, A. L. (1987). *Latent class analysis*. Newbury Park, CA: Sage.
- 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.
- Nagin, D. S. (2006). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.
- Niederhofer, H., Staffen, W., & Mair, A. (2003). A placebo-controlled study of lofexidine in the treatment of children with tic disorders and attention deficit hyperactivity disorder. *Journal of Psychopharmacology*, *17*, 113–119.
- Nock, M. K., & Kazdin, A. E. (2005). Randomized controlled trial of a brief intervention for increasing participation in parent management training. *Journal of Consulting and Clinical Psychology*, *73*, 872–879.
- 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.
- The Pediatric OCD Treatment Study Team. (2004). Cognitive-behavior therapy, sertraline, and their combination for children and adolescents with obsessive-compulsive disorder: The pediatric OCD treatment study (POTS) randomized controlled trial. *Journal of the American Medical Association*, *292*, 1969–1976.
- Piacentini, J. C., Cohen, P., & Cohen, J. (1992). Combining discrepant information from multiple sources: Are complex algorithms better than simple ones?. *Journal of Abnormal Child Psychology*, *20*, 51–63.
- Raftery, A. E. (1986). Choosing models for cross-classifications. *American Sociological Review*, *51*, 145–146.
- Rapee, R. M., Barrett, P. M., Dadds, M. R., & Evans, L. (1994). Reliability of the *DSM-III-R* childhood anxiety disorders using structured interview: Interrater and parent-child agreement. *Journal of the American Academy of Child and Adolescent Psychiatry*, *33*, 984–992.
- Research Units on Pediatric Psychopharmacology Anxiety Study Group. (2002). The Pediatric Anxiety Rating Scale (PARS): Development and psychometric properties. *Journal of the American Academy of Child and Adolescent Psychiatry*, *41*, 1061–1069.
- Richters, J. E. (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. *Psychological Bulletin*, *112*, 485–499.
- Safford, S. M., Kendall, P. C., Flannery-Schroeder, E., Webb, A., & Sommer, H. (2005). A longitudinal look at parent-child diagnostic agreement in youth treated for anxiety disorders. *Journal of Clinical Child and Adolescent Psychology*, *34*, 747–757.
- Scahill, L., Riddle, M. A., McSwiggin-Hardin, M., Ort, S. I., King, R. A., Goodman, W. K., et al. (1997). Children's Yale-Brown Obsessive Compulsive Scale: Reliability and validity. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*, 844–852.
- Shear, M. K., Rucci, P., Williams, J., Frank, E., Grochocinski, V., Bilt, J. V., et al. (2001). Reliability and validity of the Panic Disorder Severity Scale: Replication and extension. *Journal of Psychiatric Research*, *35*, 293–296.
- Silverman, W. K., & Albano, A. M. (1996). *The Anxiety Disorders Interview Schedule for Children*. San Antonio, TX: Psychological Corporation.
- Silverman, W. K., Kurtines, W. M., Ginsburg, G. S., Weems, C. F., Lumpkin, P. W., & Carmichael, D. H. (1999). Treating anxiety disorders in children with group cognitive-behavioral therapy: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *67*, 995–1003.
- Silverman, W. K., Kurtines, W. M., Ginsburg, G. S., Weems, C. F., Rabian, B., & Serafini, L. T. (1999). Contingency management, self-control, and education support in the treatment of childhood phobic disorders: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *67*, 675–687.
- Silverman, W. K., & Ollendick, T. H. (2005). Evidence-based assessment of anxiety and its disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *34*, 380–411.
- Spearing, M. K., Post, R. M., Leverich, G. S., Brandt, D., & Nolen, W. (1997). Modification of the Clinical Global Impressions (CGI) Scale for use in bipolar illness (BP): The CGI-BP. *Psychiatry Research*, *73*, 159–171.
- Wagner, K. D., Berard, R., Stein, M. B., Wetherhold, E., Carpenter, D. J., Perera, P., et al. (2004). A multicenter, randomized, double-blind, placebo-controlled trial of paroxetine in children and adolescents with social anxiety disorder. *Archives of General Psychiatry*, *61*, 1153–1162.
- Walkup, J. T., Albano, A. M., Piacentini, J., Birmaher, B., Compton, S. N., Sherill, J. T., et al. (2008). Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. *New England Journal of Medicine*, *359*, 2753–2766.
- Weisz, J. R., Jensen Doss, A., & Hawley, K. M. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology*, *56*, 337–363.
- Weisz, J. R., McCarty, C. A., & Valeri, S. M. (2006). Effects of psychotherapy for depression in children and adolescents: A meta-analysis. *Psychological Bulletin*, *132*, 132–149.
- 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 dexamethylphenidate 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.
- Yeh, M., & Weisz, J. R. (2001). Why are we here at the clinic? Parent-child (dis)agreement on referral problems at outpatient treatment entry. *Journal of Consulting and Clinical Psychology*, *69*, 1018–1025.
- Young, R. C., Biggs, J. T., Ziegler, V. E., & Meyer, D. A. (1978). A rating scale for mania: Reliability, validity, and sensitivity. *British Journal of Psychiatry*, *133*, 429–435.
- Youngstrom, E. A., IZard, C., & Ackerman, B. (1999). Dysphoria-related bias in maternal ratings of children. *Journal of Consulting and Clinical Psychology*, *67*, 905–916.
- Zaider, T. I., Heimberg, R. G., Fresco, D. M., Schneier, F. R., & Liebowitz, M. R. (2003). Evaluation of the clinical global impression scale among individuals with social anxiety disorder. *Psychological Medicine*, *33*, 611–622.