



## Whose Outcomes Come Out? Patterns of Caregiver- and Youth-reported Outcomes Based on Caregiver-youth Baseline Discrepancies

F. Tony Bonadio, Spencer C. Evans, Grace Y. Cho, Kelsey P. Callahan, Bruce F. Chorpita, the Research Network on Youth Mental Health & John R. Weisz

To cite this article: F. Tony Bonadio, Spencer C. Evans, Grace Y. Cho, Kelsey P. Callahan, Bruce F. Chorpita, the Research Network on Youth Mental Health & John R. Weisz (2021): Whose Outcomes Come Out? Patterns of Caregiver- and Youth-reported Outcomes Based on Caregiver-youth Baseline Discrepancies, *Journal of Clinical Child & Adolescent Psychology*, DOI: [10.1080/15374416.2021.1955367](https://doi.org/10.1080/15374416.2021.1955367)

To link to this article: <https://doi.org/10.1080/15374416.2021.1955367>



Published online: 23 Aug 2021.



Submit your article to this journal [↗](#)



Article views: 50



View related articles [↗](#)



View Crossmark data [↗](#)



## Whose Outcomes Come Out? Patterns of Caregiver- and Youth-reported Outcomes Based on Caregiver-youth Baseline Discrepancies

F. Tony Bonadio <sup>a</sup>, Spencer C. Evans<sup>b,c</sup>, Grace Y. Cho<sup>b,d</sup>, Kelsey P. Callahan<sup>b</sup>, Bruce F. Chorpita<sup>e</sup>, the Research Network on Youth Mental Health\*, and John R. Weisz <sup>b</sup>

<sup>a</sup>The Institute for Innovation & Implementation, University of Maryland School of Social Work; <sup>b</sup>Department of Psychology, Harvard University; <sup>c</sup>Department of Psychology, University of Miami; <sup>d</sup>Department of Psychological and Brain Sciences, University of Massachusetts Amherst; <sup>e</sup>Department of Psychology, University of California

### ABSTRACT

**Objective:** Discrepancies between caregiver and youth reports of emotional and behavioral symptoms are well-documented, with cross-informant correlations often falling in the low to moderate range. Studies have shown that caregiver-youth (dis)agreement in reporting of youth symptoms is related to treatment outcomes. However, commonly used methods for exploring reporter discrepancies (e.g., difference scores) are limited by their inability to assess discrepancies across multiple symptom domains simultaneously, and thus these previous findings do not explore multiple patterns of (dis)agreement.

**Method:** We used latent profile analysis (LPA) to identify subgroups of clinically referred youths based on patterns of caregiver- and youth-reported internalizing and externalizing symptoms for 174 caregiver-youth dyads. Longitudinal multilevel models were used to examine changes in weekly caregiver- and youth-reported internalizing symptoms, externalizing symptoms, and top problems for identified subgroups.

**Results:** The LPA identified four latent subgroups: (a) Caregiver Internalizing (9%), (b) Caregiver Internalizing-Externalizing (45%), (c) Youth Internalizing (7%), and (d) Caregiver-Youth Internalizing-Externalizing (39%). Clinical outcomes varied across informants and subgroups. Significant improvements in caregiver- and youth-reported outcome measures were documented within the Caregiver Internalizing, Caregiver Internalizing-Externalizing, and Caregiver-Youth Internalizing-Externalizing subgroups. However, only youth-reported improvements were detected in the Youth Internalizing subgroup. The results show differences in treatment outcomes across caregiver-youth informant subgroups.

**Conclusions:** These findings suggest how youth and caregiver baseline data could provide guidance for clinicians in interpreting discrepant reporting and its relevance to change during treatment.

### Introduction

Best practices for assessing youth emotional and behavioral symptoms include gathering information from multiple informants (De Los Reyes et al., 2015). In clinical practice, those reporters frequently include both caregivers and youths. Interpreting and using these assessment results can be straightforward when caregivers and youths agree but challenging when they disagree. Indeed, reporter discrepancies are very common in youth mental health, warranting further research and empirically based guidance for how clinicians should interpret and act on multi-informant data.

Reporter discrepancies can be particularly confounding when examining treatment outcomes and determining if treatment was effective. Discrepancies between caregivers and youth reports are common and well-documented (De Los Reyes et al., 2015; De Los Reyes & Kazdin, 2005). Research indicates that many factors contribute to these discrepancies, but less is known about whether these discrepancies provide useful information about how youths may respond to treatment. Understanding how patterns of discrepancies at the beginning of treatment predict clinical outcomes at the end of treatment could provide clinicians guidance for

**CONTACT** F. Tony Bonadio  [fbonadio@ssw.umaryland.edu](mailto:fbonadio@ssw.umaryland.edu)  The Institute for Innovation & Implementation, University of Maryland School of Social Work, 306 West Redwood Street, Baltimore, MD, USA

\*The Research Network on Youth Mental Health is a collaborative network funded by the John D. and Catherine T. MacArthur Foundation. Network Members at the time of the parent study were: John Weisz, Ph.D. (Network Director), Bruce Chorpita, Ph.D., Robert Gibbons, Ph.D., Charles Glisson, Ph.D., Evelyn Polk Green, M.A., Kimberly Hoagwood, Ph.D., Peter S. Jensen, M.D., Kelly Kelleher, M.D., John Landsverk, Ph.D., Stephen Mayberg, Ph.D., Jeanne Miranda, Ph.D., Lawrence Palinkas, Ph.D., and Sonja Schoenwald, Ph.D.

how treatment may progress. The aims of the current study were to identify latent subgroups of treatment-referred youths based on patterns of convergent and divergent reporting between caregivers and youths at levels of internalizing and externalizing problems at baseline, and to explore how multiple informant treatment outcomes may differ across those subgroups.

There is a well-established literature exploring the patterns of and mechanisms contributing to caregiver-youth discrepancies in reports of youth emotional and behavioral problems (for review, see De Los Reyes et al., 2015). Correlations between caregivers and youth reports consistently fall in the low to moderate range, with a mean of  $r = .28$  (De Los Reyes et al., 2015). These discrepancies occur even when symptoms are assessed with psychometrically sound measures, suggesting that caregiver-youth discrepancies reflect more than measurement error (Achenbach & Rescorla, 2004; Achenbach, 1991; De Los Reyes, 2011). In fact, reporter discrepancies may be influenced by many factors including but not limited to: type of problem being reported, context and setting of reporter observation, relationship of the reporters to the youth, and types of measures being used (De Los Reyes et al., 2015; De Los Reyes & Kazdin, 2005; Rescorla et al., 2017, 2013). Thus, caregiver-youth discrepancies could provide clinically meaningful information about youth, caregiver, and family functioning (De Los Reyes & Kazdin, 2005); however, it is difficult for clinicians to use these discrepancies to guide clinical decisions as less is known about how they relate to outcomes (Makol et al., 2019).

As patterns in informant discrepancies have emerged, researchers have developed frameworks to interpret discrepancies and inform youth mental health research and clinical practice (Dirks et al., 2012; De Los Reyes & Kazdin, 2005). The Attribution Bias Context (ABC) Model provides a framework for understanding caregiver-youth discrepancies as differences in attributions for youth behavioral and emotional problems (De Los Reyes & Kazdin, 2005). For example, caregivers frequently report more youth mental health symptoms than their children across multiple problem areas in samples of clinically referred youths (Goolsby et al., 2018; De Los Reyes et al., 2011; Salbach-Andrae et al., 2009). According to the ABC Model, caregivers are more likely to attribute the locus of youth mental health problems to the youths themselves; in contrast, children and adolescents often view their own difficult behaviors and unpleasant emotions as reactions to their environment. This difference in attribution suggests caregivers are more likely than their children to believe treatment is needed. Although the ABC framework helps explain why discrepancies may be occurring in a single

assessment occasion, it does not explicitly predict how discrepant reporting at baseline may predict response to treatment.

Although the majority of research has investigated the nature and causes of informant discrepancies, several studies have explored how discrepancies at baseline are related to treatment engagement and outcomes. Higher levels of caregiver-youth disagreement at baseline predict poorer session attendance (Brookman-Frazee et al., 2008) and reduced caregiver involvement in treatment (Israel et al., 2007). Furthermore, higher levels of discrepancy at baseline also predicts poorer treatment outcomes. For example, greater caregiver-youth discrepancies in baseline youth PTSD symptoms (Humphreys et al., 2017), social phobia symptoms (De Los Reyes, Alfano et al., 2010), and anxiety symptoms (Becker-Haimes et al., 2018) have all been found to be related to poorer response to treatment. Early studies only examined if general agreement predicted outcomes (i.e., difference scores, agreement on treatment goals) and did not examine specific patterns of discrepant reporting (Brookman-Frazee et al., 2008; Israel et al., 2007). However, other studies examined differences in treatment response based on whether youth or caregiver reported higher symptoms. When looking within anxiety symptoms, there were poorer treatment outcomes when caregivers reported significantly more symptoms than their child (Becker-Haimes et al., 2018; De Los Reyes, Alfano et al., 2010). Interestingly, when looking across multiple symptom domains, the patterns of the discrepancies differentially impacted outcomes across symptom domains (Goolsby et al., 2018). Specifically, greater caregiver-youth discrepancy on measures of depression (higher caregiver rating), anxiety (higher caregiver rating), or atypicality (higher youth rating) predicted poorer outcomes, whereas caregiver-youth discrepancy (higher caregiver rating) on hyperactivity predicted better outcomes (Goolsby et al., 2018). Taken together, these studies show that baseline discrepancies predict treatment outcomes and that patterns of these discrepancies may be meaningful in how they predict treatment outcomes.

Although these studies have shown the predictive utility of reporter discrepancies, several limitations reduce the generalizability of these studies. One major limitation is that, although baseline assessments were gathered from two reporters, most studies only examined treatment outcomes from one reporter, either caregivers or independent evaluators (Becker-Haimes et al., 2018; Goolsby et al., 2018; De Los Reyes, Alfano et al., 2010). Using independent evaluators may not be feasible in typical outpatient treatment settings. Additionally, given the known discrepancies in reporting, treatment

outcomes may be different depending on whom (i.e., caregiver or youth) reports the outcomes (De Los Reyes & Kazdin, 2009; Weisz et al., 1995). In a meta-analysis of behavioral parent training and cognitive-behavioral therapy, studies using caregiver-reported outcomes had more robust findings for behavioral parent training, whereas studies using youth-reported outcomes had more robust findings for cognitive-behavioral therapy (De Los Reyes & Kazdin, 2009). Discrepancies have been shown to be consistent over time (De Los Reyes, Goodman et al., 2010), and can often persist following certain treatment (Guo & Slesnick, 2013; De Los Reyes, Alfano et al., 2010). Given these findings, it is important to take a multi-informant approach when assessing outcomes and determining if baseline discrepancies are related to treatment outcomes.

A second limitation of prior research is that studies often use difference scores to explore how discrepancies relate to distal treatment outcomes (e.g., De Los Reyes, Alfano et al., 2010). Several limitations in the use of difference scores have been identified (Laird & De Los Reyes, 2013), including issues with the reliability of difference scores (Edwards, 2003), concerns regarding interpretability due to differences in variance between measures used to create difference scores (Edwards, 2003), and mathematical constraints when analyzing directional differences (Laird & De Los Reyes, 2013). These mathematical constraints limit the incremental validity of difference scores by reducing their ability to identify nuanced relationships between patterns of discrepancies and outcomes, beyond the existing relationship that each component measure has with the outcome (for further discussion see Laird & De Los Reyes, 2013; Tackett et al., 2013). In addition to these limitations, using difference scores does not allow for the identification of patterns of discrepancy across multiple domains simultaneously. Given known differences in discrepancies across internalizing and externalizing symptoms, it is reasonable to believe that there are underlying patterns of discrepancies. If caregiver and youth both report on internalizing and externalizing symptoms, there is a possibility for multiple patterns of convergent and divergent reporting, including but not limited to: (a) caregiver and youth both report elevated internalizing and externalizing symptoms; (b) caregiver and youth both report low internalizing and externalizing symptoms; and (c) caregiver reports low internalizing and elevated externalizing but youth reports elevated internalizing and low externalizing. Families and youths displaying these different patterns of discrepancies may respond differentially to treatment. Difference scores

can be used to assess the amount of variance accounted for by discrepant reports on one domain, but they cannot show patterns of discrepant reporting across multiple domains simultaneously. Recently, there has been a shift toward statistical approaches that can tease out these specific patterns of discrepant reporting, such as latent profile analysis, for examining patterns of multi-informant (dis)agreement and their relations to outcomes.

Latent profile analysis (LPA) is a person-centered approach that can identify specific subgroups or classes based on patterns across multiple variables, and thus can be used to understand the patterns in discrepant reporting across multiple symptom domains (Lanza & Rhoades, 2013; Makol et al., 2019; Makol & Polo, 2018). Classes identified through LPA can be used in combination with longitudinal outcomes analyses via methods such as multilevel modeling (MLM) to explore differential outcomes across subgroups. In the case of reporter discrepancies, it can be used to identify subgroups based on patterns of discrepancies using scores from multiple domains from both caregiver and youth measures simultaneously. Additionally, LPA uses multiple continuous indicator variables, such as item or subscale scores from a measure, to determine subgroups that allow for a more dimensional approach to understanding discrepancies. Thus, researchers can use LPA to find varying levels of discrepant reporting (i.e., caregiver reporting high, youth reporting moderate levels of symptoms) rather than whether an informant reports symptom levels above a clinical cutoff. LPA provides a method for identifying the patterns of discrepancies that occur within a sample, and MLM then provides a strategy for exploring and comparing the individual longitudinal trajectories for groups based on these patterns of discrepancies. LPA has been applied to a broadband measure of symptoms to identify subgroups based on symptom profiles for caregiver- and youth-reported symptoms independently (Bonadio et al., 2016); however, only a few studies have applied these methods to explore patterns of reporter discrepancies (Makol et al., 2019).

Studies that have applied LPA to reporter discrepancies have found distinct subgroups based on the patterns of discrepant caregiver and youth reporting (De Los Reyes et al., 2009). Other studies have used LPA to identify subgroups based on reporting from multiple informants, including caregiver-teacher ratings (Lerner et al., 2017) and caregiver-researcher ratings (De Los Reyes et al., 2009). However, only one study has examined treatment characteristics and outcomes across latent subgroups (Makol et al., 2019). In that study,

which focused on inpatient treatment, it was found that treatment characteristics and outcomes (i.e., standing antipsychotics, more seclusion, longer hospital stays, and more intensive aftercare services) varied across four latent subgroups identified based on (dis)agreement in caregiver and youth-reported internalizing symptoms. Taken together, these studies demonstrate the promise of LPA for identifying valid subgroups based on reporter discrepancies, with some suggestion that these latent subgroups may respond differentially to treatment.

A nuanced understanding of how discrepant reporting at baseline predicts treatment response could provide clinically meaningful information to clinicians and researchers. By understanding how certain patterns of discrepancies relate to change in outcomes over time, clinicians could have benchmarks for determining if treatment is progressing as expected. Further, if certain patterns of discrepancy are related to poorer outcomes, it may be an indicator for more intensive treatment, greater focus on engagement with caregivers, or the need for motivational interviewing at the outset of treatment. Additionally, researchers may account for variance in outcomes based on known differences across discrepancies subgroups. Previous research has demonstrated that treatment characteristics and outcomes can vary across LPA-derived subgroups for youths with internalizing symptoms in inpatient settings (Makol et al., 2019). However, these methods have not been applied to community mental health settings or to youths with a broader array of symptoms (i.e., internalizing and externalizing symptoms).

The current study used LPA to investigate latent subgroups of treatment-referred youths based on caregiver and youth reports on broad dimensions of youth emotional and behavioral problems (i.e., internalizing and externalizing). Additionally, we examined youth- and caregiver-reported treatment outcomes across the identified subgroups. Because LPA is a sample-specific exploratory approach to identifying subgroups, specific hypotheses about subgroups are not appropriate. However, based on previous studies applying LPA to reporter discrepancies, we anticipated three to five subgroups based on multiple informant reporting patterns, with some variation in treatment outcomes across these subgroups.

## Methods

### Procedure

The current study used data collected from the Child STEPs multisite study, a cluster randomized

effectiveness trial that compared a modular treatment approach for depression, anxiety, trauma, and conduct problems in youth with standard evidence-based treatments (EBTs) and usual care (Chorpita et al., 2013; Weisz et al., 2012). The original study randomly assigned clinicians to be trained in the modular approach, standard EBT, or the usual care condition. Clinicians then delivered the assigned treatments to youths seeking mental health services in public mental health clinics across two states. A battery of self- and caregiver-report measures were collected from youths and caregivers at baseline, post-treatment, and at six planned intervals (3, 6, 9, 12, 18, and 24 months) for 2 years following baseline, regardless of when treatment was complete. Additionally, two measures were collected weekly during treatment from both youths and caregivers to assess changes in symptoms and problems over time and to that provide feedback to intervention consultants who worked directly with clinicians. Study procedures, including descriptions and outcomes for each treatment condition, are described in detail in the original article (Weisz et al., 2012). Approval for all procedures was obtained from the relevant institutional review boards. Parental consent and youth assent were obtained from all participants prior to data collection.

### Participants

Participants were 174 youth-caregiver dyads of clinically referred youths ( $N = 121$  males,  $M_{\text{age}} = 10.59$ ) and their caregivers. As described by Weisz et al. (2012) and Chorpita et al. (2013) in the short-term and long-term outcome papers for this study, 44% of these youths identified as white, 32% as multiracial, 10% as Black, 6% as Latino, 3% as Asian, 3% as “other,” and 2% chose not to indicate a race/ethnicity. Total annual income was under 40,000 USD for 55% of the families, between 40,000 USD and 80,000 USD for 22% of the families, between 80,000 USD and 120,000 USD for 12% of the families, and over 120,000 USD for 8% of the families in the sample. Additionally, 49% of the participants were from single-caregiver households. To be eligible for the study, youth needed to have diagnoses related to anxiety, depression, or disruptive conduct, or have clinical elevations in any of these symptom areas on the Child Behavior Checklist (CBCL) or Youth Self-Report (YSR). Exclusionary criteria included intellectual disability, pervasive developmental disorder, psychotic symptoms, bipolar disorder, or a primary diagnosis related to inattention or hyperactivity.

## Measures

### Child Behavior Checklist

The Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2004) is a 113-item caregiver-report measure assessing common emotional and behavioral problems for youth. Caregivers report the frequency of a symptom over the past 6 months on a 3-point scale ranging from 0 (*not true*) to 2 (*very true or often true*). The CBCL contains multiple subscales, including two broad scales of internalizing and externalizing symptoms, which are of interest to the current study. Scores for the items contained within each scale are totaled, and t-scores are derived based on age and gender norms from a national sample. The CBCL has demonstrated strong reliability (Achenbach & Rescorla, 2004).

### Youth Self Report

The Youth Self Report (YSR; Achenbach & Rescorla, 2004) is the parallel form to the CBCL in which youth rate common behavioral and emotional problems. The 112 items are rated on a similar 3-point scale ranging from 0 (*not true*) to 2 (*very true or often true*) about the past 6 months. Similar to the CBCL, the YSR contains multiple subscales, including two broad scales of internalizing and externalizing symptoms, which are of interest to the current study. Based on totaled scores within each subscale, t-scores are derived from a national sample of age and gender norms.

### Brief Problem Checklist

Derived from the CBCL and the YSR, the Brief Problem Checklist (BPC) assesses the 12 most common internalizing and externalizing problems with both youth and caregiver versions. The 12 items are rated on a scale from 0 (*not true*) to 2 (*very true*) based on the frequency of each problem over the past week. The BPC provides an Internalizing score (six items), externalizing score (6 items), and a Total score (12 items). The BPC has demonstrated strong construct validity and internal reliability (Chorpita et al., 2010).

### Top Problem Assessment

The Top Problem Assessment is an idiographic measure of the top three problems youths and caregivers want to target in treatment (Weisz et al., 2011). At baseline, caregivers and youths were separately prompted to identify the problems they viewed as most important to be addressed during the youth's treatment. Each informant can provide up to three Top Problems. Caregivers and youths were then asked to rate the current severity of their respective Top Problems, in terms of "how big of a problem" it is on a scale from 0 (not at all) to 10 (very,

very much) (higher scores indicate greater problem severity). In this study, we describe the mean scores above 5 as "elevated." Top Problem severity was then reassessed on a weekly basis throughout treatment. Mean scores across the identified top problems for each respondent were computed at each time point and used for analysis.

To shed further light on the nature of these idiographic data, the content of each Top Problem was matched to the closest corresponding items on the YSR (for youths) and CBCL (for caregivers). That is, caregiver- and youth-reported top problems were classified as internalizing or externalizing based on whether the matched YSR/CBCL item loaded onto the broadband internalizing and externalizing scale. For example, Top Problems such as "worries that something bad might happen to parents" and "His sadness and crankiness has been a constant" were classified as internalizing, while Top Problems such as "I blow up" and "He doesn't listen and always wants to argue" were classified as externalizing. If the matched item did not load onto either broadband scale, it was classified as "other." Examples of other Top Problems include "Trouble being distracted" and "I have trouble getting organized." The methodology for matching the Top Problems to the YSR/CBCL items, plus intercoder reliability, is detailed in Weisz et al. (2011). The Top Problem Assessment has been shown to have strong reliability, validity, and sensitivity to change (Weisz et al., 2011).

### Analysis Plan

Before testing hypotheses, we tested assumptions for statistical analyses and examined correlations between all variables of interest. Latent profile analysis (LPA) was used to identify subgroups of youths and caregivers with similar patterns of discrepant reporting based on t-scores from the CBCL and YSR Internalizing and Externalizing subscales. LPA is a type of mixture modeling that identifies unobserved subgroups or "classes" based on a set of indicator variables (Nylund, Bellmore et al., 2007). It is an exploratory approach in which iterative models are run with sequentially increasing number of classes, starting with a 2-class model and continuing until model fit no longer improves or models no longer converge. To determine the best fitting model, multiple fit indices, including the loglikelihood, Bayesian information criteria (BIC), Akaike information criterion (AIC), entropy, and the Lo Mendell Rubin Test (LMRT; Lo et al., 2001) were examined. The LMRT allows the comparison of models with  $k$  number of classes and  $k-1$  number of classes. Furthermore, the

number of cases within each class were considered, with all classes having at least 5% of the total sample (Nylund, Asparouhov et al., 2007). In addition to these fit statistics, substantive theory was considered when determining the most appropriate number of classes. After a best-fitting model was identified, each dyad received a posterior probability for each class in the model, and dyads were then assigned a class membership based on the class for which they had the highest posterior probability. Classes were named by exploring the means on the indicator variables within each class.

After deriving the best-fitting solutions through LPA, we examined participants' latent class membership as a predictor of treatment outcomes. Generally, these analyses mirrored those performed by Weisz et al. (2012) and Chorpita et al. (2013) for post-treatment and long-term outcomes, respectively, except that we were primarily interested in LPA class membership as the predictor of clinical outcome trajectories (treatment condition was included as a covariate). Specifically, we estimated longitudinal multilevel models, with weekly/quarterly measures (Level 1) nested within participants (Level 2). Models were specified with random effects for intercept and time (natural logarithm of [days since baseline +1]); fixed effects included LPA class membership, covariates (age, gender, treatment condition), and their longitudinal product terms (i.e., class×time, age×time, gender×time, condition×time). Similar to the original analyses (Chorpita et al., 2013; Weisz et al., 2012), we were primarily interested in the omnibus and pairwise tests for the class×time interaction term, which indicates whether LPA class membership predicts significantly different trajectories of change over time. Covariates were all assessed at baseline and mean-centered for analyses; thus, results can be interpreted as the "average" across treatment conditions, gender, and age in this sample. To facilitate a clinically meaningful interpretation of the results, we also report (a) class-specific slopes, which indicate whether there was any improvement within a given class; and (b) model-implied change scores estimated for 1-year post-baseline on all measures. The reliable change index was used to determine if estimated change in scores from baseline to 1-year post-baseline follow-up was greater than would be expected by chance variation. Reliable change thresholds were calculated as  $SD * \sqrt{1-r} * 1.96$ , where SD was from the present sample and reliability coefficients (r) were from psychometric publications for the BPC (Chorpita et al., 2010) and Top Problems (Weisz et al., 2011). All MLMs were estimated in SAS version 9.4 using restricted maximum likelihood estimation.

## Results

### Descriptive Analysis

Table 1 shows t-scores for caregiver- and youth-reported internalizing and externalizing symptoms. Across the entire sample of 174 youths, caregivers reported significantly greater levels of internalizing and externalizing symptoms than their children. The full sample CBCL scores do show clinical elevations on the internalizing scale ( $t = 66.5$ ) and borderline elevations on the externalizing scale ( $t = 63.9$ ). These scores are consistent with those of other clinically referred samples (e.g., Eeren et al., 2018; Kolko et al., 2011). A total of 942 Top Problems were coded, 457 for youth and 485 for caregivers. For youth, 51% of the Top Problems were internalizing problems, 25% were externalizing problems, and 24% were other problems. For caregivers, 38% of the Top Problems were internalizing problems, 38% were externalizing problems, and 24% were other problems. Fourteen youth-caregiver dyads were missing coded data for the Top Problems.

### Latent Profile Analysis

In order to identify subgroups of youths and caregivers based on discrepant reporting, CBCL and YSR Internalizing and Externalizing t-scores for youth-caregiver dyads were entered as indicator variables in a latent profile analysis. Multiple models were run, beginning with a 2-class model and running sequential iterations until the models no longer converged.

Based on the fit statistics and substantive theory, the 4-class model was determined to be the best fitting model. The LMRT did not distinguish between any models as being a significantly better fit than the models with  $k - 1$  classes, and thus other factors had to be considered when selecting the best fitting model. The 4-class model was selected based on the number of cases in each class (i.e., no classes lower than 5% of the total sample), the improved entropy and fit (i.e., AIC and BIC) compared to the 3-class model, and the meaningful interpretability of the classes. Table 2 presents fit statistics for other models considered.

The mean t-scores for the four indicator variables – CBCL Internalizing, CBCL Externalizing, YSR Internalizing, YSR Externalizing – were examined across each of the four classes to compare and name the classes (Table 1). Class 1 (9%) was identified as “**Caregiver Internalizing**” as this group contained caregivers who reported high internalizing symptoms while youths reported low internalizing and externalizing symptoms. Class 2 (45%) was identified as “**Caregiver Internalizing**

and Externalizing” as this group contained caregivers who reported high internalizing and externalizing symptoms while youths reported low internalizing and externalizing symptoms. Class 3 (7%) was identified as “Youth Internalizing” as this group contained youths who reported high internalizing symptoms while caregivers reported low internalizing and externalizing symptoms. Class 4 (39%) was identified as “Caregiver and Youth Internalizing and Externalizing” as this group contained youths and caregivers who both reported high internalizing and externalizing symptoms. There were no significant differences among the classes (Table 1) across demographics (i.e., age, sex, and race/ethnicity).

Overall, Classes 1 through 3 included dyads of caregivers and youths with discrepant reporting. Classes 1 and 2 included dyads in which caregivers reported significantly more symptoms and Class 3 included dyads in which youths reported significantly more symptoms. Class 4 included dyads who agreed about elevated symptoms.

### Treatment Outcomes

Clinical outcome trajectories for each LPA class – controlling for treatment condition, age, and gender – are presented in Table 3 and Figure 1. Specifically, Table 3 presents the parameters of interest from longitudinal multilevel models estimating outcomes for each latent class, while Figure 1 plots these trajectories from baseline through 1 year after starting treatment. Emphasis is placed on within-group findings, and two sets of parameters in particular: (a) the coefficients and *p*-values of the log-linear slope terms, which test whether one group’s average rate of improvement was significantly greater than 0 (i.e., whether they showed *any* change);

and (b) the model-implied 1-year score and change estimates, evaluated in terms of their original-scale properties (e.g., the Top Problems 0–10 scale representing “not a problem” to “a very big problem”) and in terms of reliable change (i.e., whether the change exceeds the standard error of the measure). Though not of primary interest, we also report between-group contrasts regarding which latent classes showed faster change than other classes across all outcome measures (see Table 3) and interpret these results where they are relevant. Below, we discuss the results for each class in turn.

### Class 1 (“Caregiver Internalizing”)

perhaps unsurprisingly, began treatment with a very high level of caregiver-reported BPC Internalizing Problems at baseline. Class 1’s youth-reported internalizing problems were low at baseline, as were their externalizing problems per both informants. Of note, Class 1’s baseline Top Problems scores were elevated by *both* caregiver- and youth-reports, even though these youths had not reported these as elevated problems from a nomothetic perspective (i.e., YSR/BPC Internalizing or Externalizing Problems). The majority of caregiver-reported Top Problems in this class were internalizing (66%); this was consistent with elevations in the nomothetic measure. Interestingly, youth also identified their top problems as primarily internalizing problems (68%) on the Top Problems assessment, as did their caregivers.

In terms of change over time, the weekly scales that were most elevated at baseline for Class 1 were also the ones that showed pronounced longitudinal improvement: statistically significant and clinically reliable change was found for caregiver-reported internalizing problems and for caregiver- and youth-reported Top

Table 1. Sample characteristics.

Characteristic	Full Sample	Class 1 Caregiver Int	Class 2 Caregiver Int and Ext	Class 3 Youth Int	Class 4 Caregiver and Youth Int and Ext
Total <i>N</i> (%)	174	16 (9%)	78 (45%)	12 (7%)	68 (39%)
Sex, female, <i>n</i> (%) <sup>*</sup>	53 (30.5)	4 (25)	27 (34.6)	3 (25)	19 (27.9)
Age, years, <i>m</i> (SD)	10.13 (1.78)	10.20 (1.66)	9.78 (1.76)	10.58 (1.73)	10.48 (1.76)
Race/Ethnicity <sup>*</sup>					
Black, <i>n</i> (%)	17 (9.8)	1 (3)	7 (9.0)	3 (25)	6 (9)
Asian, <i>n</i> (%)	6 (3.4)	0	3 (3.8)	1 (8.3)	2 (3.0)
Latino, <i>n</i> (%)	10 (5.7)	2 (12.5)	2 (2.6)	1 (8.3)	5 (7.5)
Multi-racial, <i>n</i> (%)	55 (32.2)	5 (31.3)	21 (26.9)	2 (16.7)	27 (40.3)
White, <i>n</i> (%)	77 (44.3)	6 (37.5)	44 (56.4)	4 (33.3)	23 (34.3)
Other, <i>n</i> (%)	6 (3.4)	1 (6.3)	1 (6.3)	1 (8.3)	3 (4.5)
Informant Gender, female, <i>n</i> (%)	160 (92.5)	14 (87.5)	73 (93.6)	11 (91.7)	62 (92.5)
CBCL Internalizing, <i>m</i> (SD)	66.51 (9.58)	69.27 (5.28)	69.32 (7.58)	47.33 (7.38)	66.07 (8.89)
CBCL Externalizing, <i>m</i> (SD)	63.87 (9.91)	45.33 (5.91)	65.74 (6.06)	49.25 (10.04)	68.75 (6.15)
YSR Internalizing, <i>m</i> (SD)	57.19 (11.74)	54.13 (9.48)	50.49 (9.68)	64.25 (11.19)	64.31 (9.66)
YSR Externalizing, <i>m</i> (SD)	50.83 (12.28)	37.13 (6.38)	42.54 (6.88)	58.75 (9.09)	62.16 (7.03)

Int = Internalizing, Ext = Externalizing. <sup>\*</sup>Race/ethnicity and sex are missing for one youth.

**Table 2.** Latent Profile Analysis fit statistics.

Fit Statistic	2-Class	3-Class	4-Class	5-Class
Log-likelihood	-2590.81	-2575.02	-2561.67	-2548.64
BIC	5248.60	5186.04	5169.34	5153.28
AIC	5207.61	5242.80	5241.87	5241.58
Entropy	0.702	0.75	0.79	0.81
LMRT ( $p$ value)	-	30.39 (0.211)	25.70 (0.261)	25.08 (0.099)

BIC = Bayesian Information Criterion, AIC = Akaike Information Criterion, LMRT = Lo, Mendell, Rubin Test.

Problems; all other outcome trajectories were nonsignificant and did not exceed reliable change thresholds, most likely because these scores were already relatively low at baseline. In sum, Class 1's results reveal a group of youths characterized at baseline by only one type of symptom elevation (internalizing) by only one informant (caregivers), but nevertheless showed significant elevations and subsequent improvements in their Top Problems areas by both informants, as well as substantial improvements in caregiver-reported internalizing problems.

### **Class 2 ("Caregiver Internalizing and Externalizing")**

began treatment with elevated levels of caregiver-reported BPC Internalizing and Externalizing

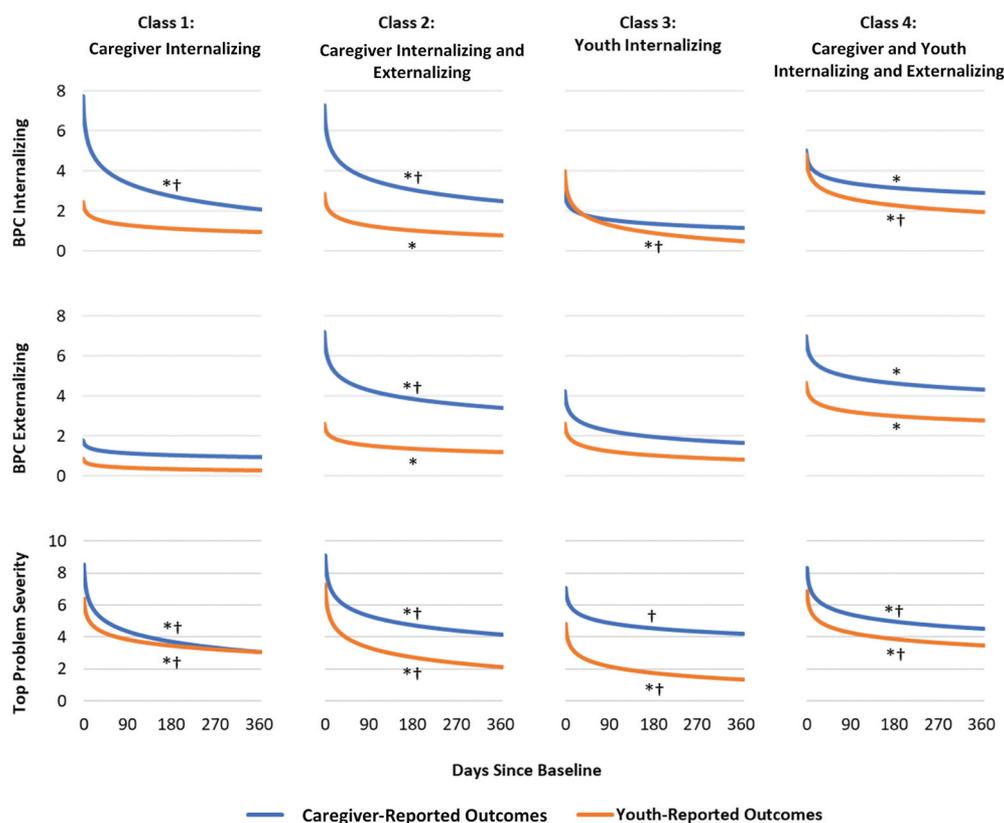
Problems. Additionally, caregiver-reported Top Problems were elevated at baseline. Interestingly, although youths in Class 2 reported relatively low baseline levels of internalizing and externalizing problems, youth-reported Top Problems scores were elevated at baseline. Some 42% of caregiver reported Top Problems were coded as internalizing problems, with 36% coded as externalizing. Over half of the Top Problems for youth in this class were coded as internalizing (53%), and fewer than a quarter were coded as externalizing (23%).

In terms of change over time, caregiver-reported BPC Internalizing and Externalizing Problems as well as Top Problems showed statistically significant and clinically reliable change over time, indicating significant

**Table 3.** Intercept, slope, and 1-year model-implied score estimates for all outcome measures.

	Intercept (Score at Baseline)			Linear Slope (Change Per Log-Day)			1-Year Estimate		
	Estimate	$p$	Higher Than	Estimate	$p$	Faster Than	Score	Change	
<b>Caregiver-Reported Outcomes</b>									
<b>BPC Int Problems</b>									
Class 1: Caregiver Int	7.74	<.001	C3	-0.96	<.001	C4	2.08	-5.65	†
Class 2: Caregiver Int-Ext	7.30	<.001	C3, C4	-0.82	<.001	C4	2.48	-4.81	†
Class 3: Youth Int	2.89	.051	.	-0.29	.292	.	1.18	-1.72	
Class 4: Caregiver-Youth Int-Ext	5.04	<.001	.	-0.36	.002	.	2.91	-2.13	
<b>BPC Ext Problems</b>									
Class 1: Caregiver Int	1.80	.123	.	-0.14	.520	.	0.95	-0.85	
Class 2: Caregiver Int-Ext	7.22	<.001	C1, C3	-0.64	<.001	C1	3.42	-3.80	†
Class 3: Youth Int	4.26	.002	.	-0.44	.104	.	1.66	-2.60	
Class 4: Caregiver-Youth Int-Ext	6.99	<.001	C1	-0.45	<.001	.	4.33	-2.66	
<b>Top Problems</b>									
Class 1: Caregiver Int	8.54	<.001	.	-0.93	<.001	.	3.06	-5.48	†
Class 2: Caregiver Int-Ext	9.12	<.001	.	-0.84	<.001	.	4.16	-4.97	†
Class 3: Youth Int	7.06	<.001	.	-0.48	.068	.	4.21	-2.85	†
Class 4: Caregiver-Youth Int-Ext	8.33	<.001	.	-0.65	<.001	.	4.50	-3.84	†
<b>Youth-Reported Outcomes</b>									
<b>BPC Int Problems</b>									
Class 1: Caregiver Int	2.47	.027	.	-0.26	.201	.	0.94	-1.53	
Class 2: Caregiver Int-Ext	2.85	<.001	.	-0.35	<.001	.	0.80	-2.04	
Class 3: Youth Int	3.97	.002	.	-0.59	.015	.	0.48	-3.50	†
Class 4: Caregiver-Youth Int-Ext	4.81	<.001	C2	-0.48	<.001	.	1.96	-2.85	†
<b>BPC Ext Problems</b>									
Class 1: Caregiver Int	0.85	.318	.	-0.09	.586	.	0.30	-0.55	
Class 2: Caregiver Int-Ext	2.62	<.001	.	-0.24	.003	.	1.22	-1.40	
Class 3: Youth Int	2.62	.011	.	-0.31	.148	.	0.82	-1.81	
Class 4: Caregiver-Youth Int-Ext	4.66	<.001	C1, C2	-0.32	<.001	.	2.76	-1.89	
<b>Top Problems</b>									
Class 1: Caregiver Int	6.40	<.001	.	-0.57	.020	.	3.05	-3.35	†
Class 2: Caregiver Int-Ext	7.30	<.001	.	-0.88	<.001	C4	2.12	-5.19	†
Class 3: Youth Int	4.82	<.001	.	-0.59	.030	.	1.36	-3.47	†
Class 4: Caregiver-Youth Int-Ext	6.88	<.001	.	-0.58	<.001	.	3.48	-3.40	†

Int = Internalizing, Ext = Externalizing. †Clinically meaningful as defined by reliable change index (i.e., improvement is significantly greater than the standard error of the measure)



**Figure 1.** Trajectories of change in BPC Internalizing Problems, BPC Externalizing Problems, and mean Top Problem severity from baseline through 12-months post-baseline. \*Log-linear slope is statistically significant at  $p < .05$ . †1-year change estimate is clinically significant based on reliable change index.

longitudinal improvement across multiple domains. However, for youth-reported symptoms, only the Top Problems showed both statistically significant and clinically reliable change over time. Further, youths in Class 2 reported significantly more improvement on their self-identified Top Problems compared with youths in Class 4, a class identified by elevated levels of youth-reported internalizing and externalizing problems. In sum, the results for Class 2 show that a group of youths characterized by discrepant caregiver and youth report at baseline (i.e., caregivers reporting elevated symptoms but not youths) nevertheless showed significant elevations and subsequent improvements in their Top Problems areas per both informants.

### **Class 3 (“Youth Internalizing”)**

began treatment with mildly elevated levels of caregiver-reported BPC Externalizing Problems and elevated levels of Top Problems at baseline; notably, this group did not have elevated levels of caregiver-reported BPC Internalizing Problems. Not surprisingly, Class 3’s youth-reported internalizing problems were elevated at baseline, but their self-reported Top Problems were not elevated.

Overall, for youths in Class 3, caregivers reported elevated Externalizing and Top Problems while youths primarily reported internalizing problems at baseline. Interestingly, fewer than half of the Top Problems identified by youths in this class were coded as internalizing (43%), with an additional 26% coded as externalizing. Caregivers’ reported Top Problems were about equally coded as internalizing (32%) and externalizing (36%).

Looking at change over time, all caregiver-reported outcome trajectories were nonsignificant, with only caregiver-reported Top Problems crossing the threshold for reliable change. For youth-reported outcomes, BPC Internalization showed a statistically significant and reliable decrease over time, which would be expected given the youth-reported elevations at baseline. Surprisingly, youth-reported Top Problems also showed statistically significant and reliable decreases over time despite being only mildly elevated at baseline. Overall, Class 3, characterized by baseline elevations in youth-reported internalization, showed significant elevations and subsequent improvements in youth-reported internalization problems and top problems, but no substantial improvements for any caregiver-reported outcomes.

#### **Class 4 (“Caregiver and Youth Internalizing and Externalizing”)**

Perhaps unsurprisingly, they had elevated baseline levels of all symptoms for both caregiver- and youth-reported outcomes. Generally, although youths in this class were characterized by elevations across both internalizing and externalizing for both caregiver- and youth-report, the overall levels of those elevations were not significantly different from the other classes. In other words, although caregivers and youths agreed on elevation of symptoms, the overall severity of those elevated symptoms was not higher than other classes that have elevated symptoms. Any significant differences in baseline scores is consistent with expected direction of those differences given the characteristics of the class. For example, Class 4’s (“Caregiver and Youth Internalizing and Externalizing”) caregiver-reported externalizing is significantly higher than Class 1’s (“Caregiver Internalizing”) caregiver-reported externalizing. Notably, youths in Class 4 did not report significantly higher Top Problems compared to the other classes. For youths in Class 4, 45% of identified Top Problems were classified as internalizing, and 33% were classified as externalizing. However, for caregivers in Class 4, only 27% of The Top Problems were classified as internalizing, with 48% classified as externalizing.

In terms of change over time, statistically significant improvement with clinically reliable change was found for both caregiver- and youth-reported Top Problems and for youth-reported BPC Internalizing problems; all other outcome trajectories were nonsignificant and did not exceed reliable change thresholds. Overall, results for Class 4, a group characterized by youth and caregiver agreement on elevated baseline symptoms, showed reliable improvements in multi-informant top problems and youth-reported internalization problems, despite having baseline elevations on internalizing and externalizing problems per both informants.

**Patterns in Treatment Outcomes.** Although there were specific patterns of treatment outcomes within each class, overarching patterns across classes could be identified (Table 4). Specifically, when there was disagreement about internalizing symptoms at baseline with either caregivers reporting higher symptoms (Classes 1 and 2) or youths reporting higher symptoms (Class 3), clinically significant changes were only found for the reporter who reported higher symptoms. In other words, if a youth reported elevated internalizing symptoms at baseline but their caregiver did not; then, only the youth’s report would indicate clinically reliable change in internalizing symptoms over time. There were clinically reliable changes for youth-reported Top Problems for all four classes and for three of the four

classes for caregiver-reported Top Problems. There was no clinically significant change in youth-reported externalizing symptoms found for any class. This was the only outcome to show no clinically significant change in any class.

#### **Discussion**

Previous research has established that caregivers and their children often report different levels of youth mental health symptoms (De Los Reyes et al., 2015), but less is known about what these discrepancies might predict about treatment outcomes. The present study uses latent profile analysis (LPA) to explore if patterns of discrepant reporting between caregivers and youths predicted treatment outcomes. To date, our study is one of the first to use the LPA-approach to examine the relationship between patterns in discrepant reporting of internalizing and externalizing symptoms and treatment outcomes in an outpatient youth sample. In addition, our study is unique as we explore differential outcomes using both caregiver and youth rating on idiographic (i.e., Top Problems Assessment) and nomothetic (i.e., internalizing and externalizing) methods of symptom measurement. Overall, findings demonstrated that treatment outcomes reported by both caregivers and youths differed across subgroups and across measurement types. These findings provided interesting implications for clinical interpretation of measures and methods for symptom measurement.

When applying LPA to baseline caregiver- and youth-rated internalizing and externalizing symptoms, a four-class model emerged as the best fitting model for the caregiver-youth dyads. The classes were defined by those who reported elevations in internalizing and externalizing symptoms. The four subgroups identified were as follows: Class 1 “Caregiver Internalizing,” Class 2 “Caregiver Internalizing and Externalizing,” Class 3 “Youth Internalizing,” and Class 4 “Caregiver and Youth Internalizing and Externalizing.”

We found that youth and caregiver-reported outcomes varied across classes. For youths in Class 1 (“Caregiver Internalizing”) and Class 2 (“Caregiver Internalizing and Externalizing”), symptom improvements corresponded to baseline elevations. For example, Class 1 caregivers reported elevated internalizing symptoms and Top Problems at baseline and improvements in these areas over time. Interestingly, youths in both classes reported elevated baseline Top Problem scores despite reporting non-elevated levels of internalizing and externalizing problems on the nomothetic assessments, and they reported significant decreases in these Top Problem scores throughout treatment.

**Table 4.** Summary of treatment outcome findings by class.

	<b>Class 1</b> Caregiver Internalizing	<b>Class 2</b> Caregiver Internalizing and Externalizing	<b>Class 3</b> Youth Internalizing	<b>Class 4</b> Caregiver and Youth Internalizing and Externalizing
Caregiver Internalizing	●	●		
Caregiver Externalizing		●		
Caregiver Top Problems	●	●		●
Youth Internalizing			●	●
Youth Externalizing				●
Youth Top Problems	●	●	●	●

Shaded cells indicate elevated baseline scores on CBCL/YSR internalizing and externalizing scales (not applicable for Top Problems); ● indicates statistically significant and clinically reliable change on Top Problems and on Brief Problems Checklist internalizing and externalizing scales.

Interestingly, in both classes, over half of the youth-reported Top Problems were internalizing problems. This may suggest that despite the lack of clinical elevations on nomothetic measures of internalizing symptoms, youths may be experiencing specific internalizing problems that are not being detected as clinically elevated by nomothetic measures. Findings for Class 1 and Class 2 highlight the importance of using both idiographic and nomothetic approaches for symptom measurement. Thus, both approaches can offer clinically useful information, with sensitivity for detecting levels of, and changes in, youth emotional and behavioral problems.

Class 3 (“Youth Internalizing”) was characterized by elevated youth-reported internalizing problems at baseline which decreased over time. Although Class 3 youths reported internalizing problems at baseline, they reported lower Top Problem scores than youths from all other classes. It is possible that youths with internalizing problems view their self-reported Top Problems through a lens distorted by depression or anxiety (e.g., “My problems aren’t as big as others”). However, it is also possible that the relative severity of problems for youth in this class was actually lower than for their peers, despite the clinical elevation of internalizing symptoms. Interestingly, this is the only class in which youth had higher ratings on the standardized measures than their caregivers. This may indicate that the problems that these youths face are isolated to one context (e.g., social anxiety at school, but not at home), specifically a context in which caregivers are not present (Deros et al., 2018; De Los Reyes et al., 2015). Given that these problems could be isolated to one context, they may not be as pervasive or disruptive. Thus, youths in Class 3 report lower Top Problems compared to youth in other classes who may experience problems across multiple contexts. Despite only mild baseline elevations, youth Top Problem scores significantly decreased over time. This finding underscores the importance of monitoring progress in outcome areas that are clinically of interest even when they are not significantly elevated at baseline,

particularly when the clinician has a hypothesis as to why a youth might be underreporting.

Also, contrary to expectations, Class 3 caregivers did not report elevated internalizing symptoms for their children, but instead reported mildly elevated externalizing symptoms. One explanation for this discrepancy is that these caregivers may have interpreted behavioral outcomes of emotional problems (e.g., irritability) as externalizing problems (e.g., aggression, disobedience). However, we might then expect decreases in caregiver-reported externalizing symptoms as youth-reported internalizing problems decrease, which we did not find. Alternatively, this discrepancy may also relate to parents only observing youth in one context (i.e., home), whereas the youth is reporting on experiences across multiple contexts (i.e., home, school, extra-curricular). In this case, some externalizing behaviors at home may elevate the parent reports, but if these problems are not consistent across contexts the youth may not report elevated externalizing behaviors (De Los Reyes et al., 2015). Further, if treatment was focused on the elevated internalizing symptoms reported by youth in Class 3, the mildly elevated externalized problems reported by caregivers may not have been addressed and thus persist after treatment. Overall, findings for Class 3 suggest that when youths report elevated internalizing symptoms but caregivers do not track youth-reported symptoms, in addition to caregiver-reported symptoms, is essential for monitoring treatment progress.

Results for Class 4 (“Caregiver and Youth Internalizing and Externalizing”) were also unexpected. While caregiver-youth agreement generally predicts better treatment outcomes (Humphreys et al., 2017; Israel et al., 2007), in the present study, the group characterized by high agreement fared the worst during treatment. This difference in findings in the present study may be partially due to the marked variation in youth presenting concerns, problems targeted by treatment, and treatment procedures used; in some previous studies all youth were selected for having similar presenting problems (e.g., PTSD) and received the same treatment

(see Humphreys et al., 2017). Additionally, previous studies examined only the level of caregiver-youth agreement but not the direction of the agreement (i.e., youth-high, parent-high vs. youth-low, parent-low). One previous study using similar methods found that a group of youths and caregivers who agreed on high internalizing symptoms had less favorable outcomes than youths and caregivers who agreed on low internalizing symptoms (Makol et al., 2019). Thus, this subgroup of caregivers and youths who agree on elevated symptoms may represent a unique population for whom more intensive services are needed. Of note, although there was agreement about symptom elevations in all areas, Class 4 caregivers and youths generally did not report greater symptom *severity* than other classes with elevated symptoms, so poorer treatment response should not be attributed to Class 4 youths having more acute symptoms at baseline. Further, although results from other classes in which elevated symptoms at baseline improve over time may be interpreted as regression to the mean, the lack of a decrease in elevated scores for Class 4 suggests that findings may represent more than regression to the mean.

Another interesting pattern emerged when looking at results across all classes. Youth-reported externalizing symptoms were the only outcome to show no clinical improvement across any class. This may be partly attributable to the low baseline scores on youth-reported externalizing symptoms for three of the four classes. These low baseline scores may limit the ability to detect changes in externalizing symptoms over time. However, youths in Class 4 were the exception, as they reported significantly higher baseline externalizing symptoms compared to youths in other classes, but they also showed no clinically significant change. Interestingly, youths in Class 4 were the only youths with self-reported comorbid internalizing problems, which has previously been shown to enhance treatment outcomes (Ollendick et al., 2016). Also, as noted discussed above, Class 4 was the only class that demonstrated youth-caregiver agreement in symptom reporting, which may be an indicator of less favorable outcomes for some youths (Makol et al., 2019). Youths in Class 4 did show clinically significant changes in youth-reported internalizing symptoms which may indicate that their treatment focused on these symptoms rather than their externalizing symptoms. Further research is needed, particularly into effects of different treatments, for this unique group of youth with comorbid symptoms.

Beyond class differences in outcomes, our study highlights the importance of using both nomothetic and idiographic measures. Differences in elevation scores and types of problems identified (i.e., internalizing vs. externalizing)

between our idiographic measure (Top Problems) and our nomothetic measure (Brief Problem Checklist) indicate the utility of including various types of assessment measures to evaluate behavioral and emotional issues. For example, although youth in Class 1 (Caregiver Internalizing) and Class 2 (Caregiver Internalizing and Caregiver Externalizing) may not have reached clinical levels based on their responses to items on nomothetic measures, they endorsed elevated levels when rating self-identified problems via the Top Problem Assessments. Some problems that were especially important to the youth may not have been adequately captured via the nomothetic measures. For example, youths in Class 3 (Youth Internalizing) reported clinically elevated internalizing problems on the nomothetic measures, but fewer than half of their self-identified top problems were internalizing. These self-identified problems likely aided in identifying targets for treatment and provided a means of tracking treatment progress for youths without baseline clinical elevations on nomothetic measures. Idiographic measures not only provide a different way of capturing clinically relevant information but may also elicit responses that fall outside of what nomothetic measures may identify as a clinical need. Additionally, for three of the four classes, both caregivers and youth indicated significant improvement on their Top Problems. This suggests that although there are discrepant reports from a nomothetic measure of symptoms, there may be significantly more agreement when using more idiographic measures of youth behavioral and emotional problems. Future research is required to delineate what contributes to discrepant reporting between idiographic and nomothetic measures.

Although the present study contributes to an emerging field in the youth treatment literature, several limitations exist. First, due to small within-group sample sizes, we had limited power to detect differences between groups (e.g., demographics) or to investigate questions concerning which treatment conditions worked best for whom. Replicating these analyses with large, diverse samples would improve the generalizability of our findings. Additionally, the idiographic Top Problem Assessment may pull in for high baseline ratings due to the nature of asking treatment-seeking families which issues they would like to work on at intake. Thus, top problems may artificially inflate the probability of change over time. However, idiographic measures may also identify problems not captured on nomothetic measures and the personalized nature of the Top Problems measure fits current recommendations for evidence-based assessment of emotional and behavioral functioning (Youngstrom et al., 2017). Finally, we show that there are some limitations with respect to change scores and the associated reliable change index for a measuring change over time. Change scores face some of the same interpretative issues that arise for difference scores (see,

Cronbach & Furby, 1970). Thus, it is useful to note that our use of this approach to assess estimated changes in symptoms from baseline to 1-year is complemented by slope estimates and related significance testing, in addition to multiple informants and measures. Of note, the consistency between reliable change indicators and significant slopes in our results underscores the clinical and statistical significance of the observed patterns of change.

Despite these limitations, the present study expands on the extant literature with important implications for both researchers and clinicians. Although previous literature suggests that discrepant reporting between caregivers and adolescents in inpatient settings may predict negative treatment characteristics such as longer periods of hospitalization and prescription of antipsychotics (Makol et al., 2019), our study suggests that in clinical samples certain patterns of discrepant reporting between caregivers and youths can predict positive treatment outcomes. Knowing that these patterns of discrepancy predict treatment outcome, clinicians could have a better understanding of anticipated response to treatment which they could monitor over time to understand whether treatment is progressing as expected. Additionally, it may provide them with insight into which youths and families may need more intensive engagement and services to ensure both youths and families are benefiting from treatment. Future research should identify these patterns of discrepancy at baseline and apply various engagement strategies (e.g., psychoeducation, expectation setting; Lindsey et al., 2014) or varying intensity of services to determine if treatment outcomes can be improved for youths and families who may be at higher risk of poor response to treatment. Future research should also examine reasons for discrepant reporting and whether those reasons are addressed in treatment. Convergence of discrepant reporting over time could be explored to determine if it may indicate positive impacts from treatment. Replication of these findings is key to disentangle these differences and ensure that our results are not sample-specific and can be generalized.

## Acknowledgments

We thank the youths, caregivers, staff, and administrators who participated in this research.

## Disclosure Statement

The Modular Approach to Treatment of Children with Anxiety, Depression, or Conduct Problems (MATCH) protocol tested in this study was a precursor to a revised and expanded version for which Bruce F. Chorpita and John R.

Weisz receive income. The other authors declare that they have no conflicts of interest.

## Funding

We gratefully acknowledge the support from the John D. and Catherine T. MacArthur Foundation [Grant 83423, to John R. Weisz].

## ORCID

F. Tony Bonadio  <http://orcid.org/0000-0002-7248-3246>  
John R. Weisz  <http://orcid.org/0000-0002-5560-6814>

## References

- Achenbach, T. M. (1991). *Manual for the CBCL/4-18 and 1991 Profile*. Burlington, VT: University of Vermont Department of Psychiatry.
- Achenbach, T. M., & Rescorla, L. A. (2004). The Achenbach System of Empirically Based Assessment (ASEBA) for ages 1.5 to 18 years. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (Vol. 2, pp. 179–213). Routledge.
- Becker-Haimes, E. M., Jensen-Doss, A., Birmaher, B., Kendall, P. C., & Ginsburg, G. S. (2018). Parent–youth informant disagreement: Implications for youth anxiety treatment. *Clinical Child Psychology and Psychiatry*, 23(1), 42–56. <https://doi.org/10.1177/1359104516689586>
- Bonadio, F. T., Dynes, M., Lackey, J., Tompsett, C., & Amrhein, K. (2016). Grouping youth with similar symptoms: A person-centered approach to transdiagnostic subgroups. *Journal of Clinical Psychology*, 72(7), 676–688. <https://doi.org/10.1002/jclp.22274>
- Brookman-Frazee, L., Haine, R. A., Gabayan, E. N., & Garland, A. F. (2008). Predicting frequency of treatment visits in community-based youth psychotherapy. *Psychological Services*, 5(2), 126–138. <https://doi.org/10.1037/1541-1559.5.2.126>
- Chorpita, B. F., Reise, S., Weisz, J. R., Grubbs, K., Becker, K. D., & Krull, J. L. (2010). Evaluation of the brief problem checklist: Child and caregiver interviews to measure clinical progress. *Journal of Consulting and Clinical Psychology*, 78(4), 526–536. <https://doi.org/10.1037/a0019602>
- Chorpita, B. F., Weisz, J. R., Daleiden, E. L., Schoenwald, S. K., Palinkas, L. A., Miranda, J., Higa-mcmillan, C. K., Nakamura, B. J., Austin, A. A., Borntreger, C. F., Ward, A., Wells, K. C., & Gibbons, R. D., & Research Network on Youth Mental Health. (2013). Long-term outcomes for the Child STEPs randomized effectiveness trial: A comparison of modular and standard treatment designs with usual care. *Journal of Consulting and Clinical Psychology*, 81(6), 999–1009. <https://doi.org/10.1037/a0034200>
- Cronbach, L. J., & Furby, L. (1970). How we should measure “change”: Or should we? *Psychological Bulletin*, 74(1), 68–80. <https://doi.org/10.1037/h0029382>
- 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 & Adolescent Psychology*, 40(1), 1–9. <https://doi.org/10.1080/15374416.2011.533405>
- 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(3), 395–404. <https://doi.org/10.1007/s10802-009-9373-6>
- De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, 141(4), 858–900. <https://doi.org/10.1037/a0038498>
- De Los Reyes, A., Goodman, K. L., Kliewer, W., & Reid-Quinones, K. (2010). The longitudinal consistency of mother–child reporting discrepancies of parental monitoring and their ability to predict child delinquent behaviors two years later. *Journal of Youth and Adolescence*, 39(12), 1417–1430. <https://doi.org/10.1007/s10964-009-9496-7>
- 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(5), 637–652. <https://doi.org/10.1007/s10802-009-9307-3>
- 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(4), 483–509. <https://doi.org/10.1037/0033-2909.131.4.483>
- 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(5), 583–617. <https://doi.org/10.1177/0145445509343203>
- 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 & Adolescent Psychology*, 40(1), 36–53. <https://doi.org/10.1080/15374416.2011.533402>
- Deros, D. E., Racz, S. J., Lipton, M. F., Augenstein, T. M., Karp, J. N., Keeley, L. M., Qasmieh, N., Grewe, B. I., Aldao, A., & De Los Reyes, A. (2018). Multi-informant assessments of adolescent social anxiety: Adding clarity by leveraging reports from unfamiliar peer confederates. *Behavior Therapy*, 49(1), 84–98. <https://doi.org/10.1016/j.beth.2017.05.001>
- Dirks, M. A., Reyes, A. D. L., Briggs-Gowan, M., Cella, D., & Wakschlag, L. S. (2012). Annual Research Review: Embracing not erasing contextual variability in children’s behavior – Theory and utility in the selection and use of methods and informants in developmental psychopathology. *Journal of Child Psychology and Psychiatry*, 53(5), 558–574. <https://doi.org/10.1111/j.1469-7610.2012.02537.x>
- Edwards, J. R. (2003). *Alternatives to difference scores: Polynomial regression and response surface methodology: (576892011-020) [Data set]*. San Francisco, CA: Jossey-Bass. <https://doi.org/10.1037/e576892011-020>
- Eeren, H. V., Goossens, L. M. A., Scholte, R. H. J., Busschbach, J. J. V., & van der Rijcken, R. E. A. (2018). Multisystemic therapy and functional family therapy compared on their effectiveness using the propensity score method. *Journal of Abnormal Child Psychology*, 46(5), 1037–1050. <https://doi.org/10.1007/s10802-017-0392-4>
- Goolsby, J., Rich, B. A., Hinnant, B., Habayeb, S., Berghorst, L., Reyes, A. D. L., & Alvord, M. K. (2018). Parent–child informant discrepancy is associated with poorer treatment outcome. *Journal of Child and Family Studies*, 27(4), 1228–1241. <https://doi.org/10.1007/s10826-017-0946-7>
- Guo, X., & Slesnick, N. (2013). Family versus individual therapy: Impact on discrepancies between parents’ and adolescents’ perceptions over time. *Journal of Marital and Family Therapy*, 39(2), 182–194. <https://doi.org/10.1111/j.1752-0606.2012.00301.x>
- Humphreys, K. L., Weems, C. F., & Scheeringa, M. S. (2017). The role of anxiety control and treatment implications of informant agreement on child PTSD symptoms. *Journal of Clinical Child & Adolescent Psychology*, 46(6), 903–914. <https://doi.org/10.1080/15374416.2015.1094739>
- Israel, P., Thomsen, P. H., Langeveld, J. H., & Stormark, K. M. (2007). Parent–youth discrepancy in the assessment and treatment of youth in usual clinical care setting: Consequences to parent involvement. *European Child & Adolescent Psychiatry*, 16(2), 138–148. <https://doi.org/10.1007/s00787-006-0583-y>
- Kolko, D. J., Iselin, A.-M. R., & Gully, K. J. (2011). Evaluation of the sustainability and clinical outcome of Alternatives for Families: A Cognitive-Behavioral Therapy (AF-CBT) in a child protection center. *Child Abuse & Neglect*, 35(2), 105–116. <https://doi.org/10.1016/j.chiabu.2010.09.004>
- Laird, R. D., & De Los Reyes, A. (2013). Testing informant discrepancies as predictors of early adolescent psychopathology: Why difference scores cannot tell you what you want to know and how polynomial regression may. *Journal of Abnormal Child Psychology*, 41(1), 1–14. <https://doi.org/10.1007/s10802-012-9659-y>
- Lanza, S. T., & Rhoades, B. L. (2013). Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science*, 14(2), 157–168. <https://doi.org/10.1007/s11121-011-0201-1>
- Lerner, M. D., De Los Reyes, A., Drabick, D. A. G., Gerber, A. H., & Gadow, K. D. (2017). Informant discrepancy defines discrete, clinically useful autism spectrum disorder subgroups. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 58(7), 829–839. <https://doi.org/10.1111/jcpp.12730>
- Lindsey, M. A., Brandt, N. E., Becker, K. D., Lee, B. R., Barth, R. P., Daleiden, E. L., & Chorpita, B. F. (2014). Identifying the common elements of treatment engagement interventions in children’s mental health services. *Clinical Child and Family Psychology Review*, 17(3), 283–298. <https://doi.org/10.1007/s10567-013-0163-x>
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767–778. <https://doi.org/10.1093/biomet/88.3.767>
- Makol, B. A., De Los Reyes, A., Ostrander, R. S., & Reynolds, E. K. (2019). Parent-youth divergence (and convergence) in reports of youth internalizing problems in psychiatric inpatient care. *Journal of Abnormal Child Psychology*, 47(10), 1677–1689. <https://doi.org/10.1007/s10802-019-00540-7>
- Makol, B. A., & Polo, A. J. (2018). Parent-child endorsement discrepancies among youth at chronic-risk for depression.

- Journal of Abnormal Child Psychology*, 46(5), 1077–1088. <https://doi.org/10.1007/s10802-017-0360-z>
- Nylund, K., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535–569. <https://doi.org/10.1080/10705510701575396>
- Nylund, K., Bellmore, A., Nishina, A., & Graham, S. (2007). Subtypes, severity, and structural stability of peer victimization: What does latent class analysis say? *Child Development*, 78(6), 1706–1722. <https://doi.org/10.1111/j.1467-8624.2007.01097.x>
- Ollendick, T. H., Greene, R. W., Austin, K. E., Fraire, M. G., Halldorsdottir, T., Allen, K. B., Jarrett, M. A., Lewis, K. M., Whitmore Smith, M., Cunningham, N. R., Noguchi, R. J. P., Canavera, K., & Wolff, J. C. (2016). Parent management training and collaborative & proactive solutions: A randomized control trial for oppositional youth. *Journal of Clinical Child & Adolescent Psychology*, 45(5), 591–604. <https://doi.org/10.1080/15374416.2015.1004681>
- Rescorla, L. A., Ewing, G., Ivanova, M. Y., Aebi, M., Bilenberg, N., Dieleman, G. C., Döpfner, M., Kajokiene, I., Leung, P. W. L., Plück, J., Steinhausen, H.-C., Metzke, C. W., Zukauskienė, R., & Verhulst, F. C. (2017). Parent–adolescent cross-informant agreement in clinically referred samples: Findings from seven societies. *Journal of Clinical Child & Adolescent Psychology*, 46(1), 74–87. <https://doi.org/10.1080/15374416.2016.1266642>
- Rescorla, L. A., Ginzburg, S., Achenbach, T. M., Ivanova, M. Y., Almqvist, F., Begovac, I., Bilenberg, N., Bird, H., Chahed, M., Dobrean, A., Döpfner, M., Erol, N., Hannesdottir, H., Kanbayashi, Y., Lambert, M. C., Leung, P. W. L., Minaei, A., Novik, T. S., Oh, K.-J., Petot, J.-M., ... Verhulst, F. C. (2013). Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *Journal of Clinical Child & Adolescent Psychology*, 42(2), 262–273. <https://doi.org/10.1080/15374416.2012.717870>
- Salbach-Andrae, H., Klinkowski, N., Lenz, K., & Lehmkuhl, U. (2009). Agreement between youth-reported and parent-reported psychopathology in a referred sample. *European Child & Adolescent Psychiatry*, 18(3), 136–143. <https://doi.org/10.1007/s00787-008-0710-z>
- Tackett, J. L., Herzhoff, K., Reardon, K. W., Smack, A. J., & Kushner, S. C. (2013). The relevance of informant discrepancies for the assessment of adolescent personality pathology. *Clinical Psychology: Science and Practice*, 20(4), 378–392. <https://doi.org/10.1111/cpsp.12048>
- Weisz, J. R., Chorpita, B. F., Frye, A., Ng, M. Y., Lau, N., Bearman, S. K., Ugueto, A. M., Langer, D. A., & Hoagwood, K. E., & Research Network on Youth Mental Health. (2011). Youth top problems: Using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy. *Journal of Consulting and Clinical Psychology*, 79(3), 369–380. <https://doi.org/10.1037/a0023307>
- Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., Daleiden, E. L., Ugueto, A. M., Ho, A., Martin, J., Gray, J., Alleyne, A., Langer, D. A., Southam-Gerow, M. A., & Gibbons, R. D., & Health, and the R. N. on Y. M. (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth: A randomized effectiveness trial. *Archives of General Psychiatry*, 69(3), 274–282. <https://doi.org/10.1001/archgenpsychiatry.2011.147>
- Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin*, 117(3), 450–468. <https://doi.org/10.1037/0033-2909.117.3.450>
- Youngstrom, E. A., Meter, A. V., Frazier, T. W., Hunsley, J., Prinstein, M. J., Ong, M.-L., & Youngstrom, J. K. (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice*, 24(4), 331–363. <https://doi.org/10.1111/cpsp.12207>