J Child Fam Stud
DOI 10.1007/s10826-016-0419-4

ORIGINAL PAPER

Concurrent Validation of the Treatment Outcome Package (TOP) for Children and Adolescents

Elizabeth E. Baxter[1] · Pamela C. Alexander[1] · David R. Kraus[1] · Jordan H. Bentley[1] · James F. Boswell[2] · Louis G. Castonguay[3]

Abstract The Treatment Outcome Package for children and adolescents (TOP) is a behavioral health and well-being assessment used widely in clinical and child welfare populations. The primary purpose of this study was to examine the concurrent validity of the Child and Adolescent versions of the TOP with the Child Behavior Checklist (CBCL) and the Strengths and Difficulties Questionnaire (SDQ) with a community sample. Children and adolescents (N = 203) 3–18 years of age from a community sample completed the CBCL, the SDQ, and the TOP. Significant correlations emerged between the TOP and theoretically similar scales on the SDQ and the CBCL. Analyses demonstrated that the TOP has excellent concurrent validity on most subscales with both the CBCL and the SDQ. These results provide additional evidence for the TOP's utility and validity as a measure of psychological well-being and functioning.

Keywords Concurrent validity · Mental health · Children · Adolescents · Assessment

Introduction

The Treatment Outcome Package (TOP; Kraus et al. 2005) is a collection of assessment tools designed to measure the behavioral health, functioning, and well-being outcomes of children, adolescents, and adults. The TOP has been widely used for over 20 years in a behavioral health population to help therapists track their clients' psychotherapy outcomes. More recently, the TOP for children (Kraus et al. 2010) and adolescents has been used in child welfare populations throughout the United States.

The most recent version of the TOP for Children was developed in 2009 and includes 48 items that span thirteen domains including: ADHD symptoms, Assertiveness, Incontinence, Conduct, Depression, Psychosis, Separation Anxiety, Sleep, Resiliency, Suicide, Eating Problems, and Violence. The TOP for Adolescents consists of 11 domains derived from both the Child TOP and the Adult TOP including: ADHD symptoms, Conduct, Depression, Mania, Psychosis, Sleep, Suicide, Social Conflict, School Functioning, Violence, and Substance Abuse. Based on exploratory and confirmatory factor analytic methods, the construct validity of the scale and these domains was demonstrated to be excellent (Kraus et al. 2010), with the final factor model exceeding the fit statistics established for the adult version of TOP (Kraus et al. 2005).

The predictive validity of the TOP has also been established for all age versions in both commercial and public-sector health plans to predict future behavioral health hospitalizations (Beacon Health Strategies 2008; Kelly et al. 2008; Stelk and Berger 2009). Also, the Adult TOP has demonstrated an ability to predict the level of improvement in clients over time, both in relation to the client's characteristics and using the measured effectiveness of clinicians (Kraus et al. 2011, 2016). Finally, the predictive validity of the adolescent and child versions of the tool has been established in a child welfare sample (Alexander et al. 2016), using TOP data to predict placement disruptions.

✉ Elizabeth E. Baxter
lbaxter@outcomereferrals.com

[1] Outcome Referrals Inc., 1 Speen Street, Framingham, MA 01701, USA

[2] Department of Psychology, University at Albany, SUNY, Albany, NY, USA

[3] Department of Psychology, The Pennsylvania State University, University Park, PA, USA

Published online: 09 April 2016

Data also exist on the concurrent validity of the TOP. In a sample of both general population and behavioral health participants, Adult TOP scores correlated strongly with the Beck Depression Inventory-II ($r = -.92$ to .71) (Beck et al. 1996), the Brief Symptom Inventory ($r = -.90$ to .66) (Derogatis 1975), the Minnesota Multiphasic Personality Inventory-II ($r = -.73$ to .60) (Butcher et al. 2001), the BASIS 32 ($r = -.86$ to .73) (Eisen et al. 1986), and the Short-Form Health Survey-36 ($r = -.68$ to .82) (McHorney et al. 1993) (Kraus et al. 2005; Kraus and Seligman 2004). Studies investigating the concurrent validity of previous versions of the Child TOP have yielded similar results, with strong positive correlations between Child TOP domains and the Behavioral Assessment System for Children (BASC; Reynolds and Kamphaus 1992), the Child and Adolescent Functional Assessment Scales (CAFAS; Hodges 2000), and the Child Behavior Checklist (CBCL; Achenbach and Rescorla 2001) (Kraus and Seligman 2004). Although these data were available in the user manual (Kraus and Seligman 2004), they were not published in a peer-reviewed source. The purpose of this study was to re-test and publish the concurrent validity of the most recent version of the Child and Adolescent TOP in a community sample of children and adolescents recruited from the general population. The validity measures that were used in this study were two commonly used instruments: the Child Behavior Checklist (CBCL; Achenbach and Rescorla 2001) and the Strengths and Difficulties Questionnaire (SDQ; Goodman 1997).

We predicted that the TOP Depression, ADHD, Psychosis, Social Conflict and Violence scales would be highly associated with theoretically similar scales on the SDQ (Hyperactivity, Emotional Problems, and Conduct) and the CBCL (Withdrawn/Depressed, Anxious/Depressed, Attention Problems, Thought Problems, Social Problems, and Aggressive Behavior). Furthermore, we expected that there would be moderately sized correlations between TOP domains and scales on the CBCL and SDQ scales that tap into distinct yet related constructs (i.e., TOP Sleep and CBCL Somatic Complaints). We did not, however, expect a strong correlation between the TOP Conduct domain and the SDQ Conduct scale because, despite some overlap, the TOP scale focuses on more serious conduct problems while the SDQ Conduct items reflect anger and interpersonal conflict more broadly.

Method

Participants

Participants consisted of 203 children and adolescents from a community convenience sample recruited through word of mouth. For example, packets were distributed to parents and youth at sporting events.

Parents of children 3–12 years of age ($N = 157$) completed the assessments on behalf of the child. Adolescents 13–18 years of age ($N = 46$) completed the assessments themselves. Although parents may have reported on multiple children, the anonymity of the study precluded tracking this information. In order to remain consistent with the age ranges defined by the assessment tools used in this study, participants were divided into the following age groups: preschool (3–5 years of age); school age (6–12 years of age); and adolescents (13–18 years of age). School age children and preschool children were collapsed into a larger "child" sample in order to compare the TOP with the SDQ. The preschool CBCL results were omitted from the analysis due to insufficient numbers ($N = 23$). In the child sample ($N = 157$), 53.5 % were male and the average age was 8.55 years ($SD = 3.41$). In the adolescent sample ($N = 46$), 46.7 % were male and the average age was 15.87 years ($SD = 1.80$).

Procedure

Parents of children and adolescents were either handed or mailed a sealed envelope containing a letter describing the study and instructions detailing how to anonymously complete the enclosed forms for the following assessments: The Child Behavior Checklist (CBCL), the Strengths and Difficulties Questionnaire (SDQ), and the Treatment Outcome Package for children and adolescents (TOP). Parents were instructed to fill out the measures on behalf of their children ages 3–12 years and to distribute the measures to their adolescent children ages 13–18 years for self-completion. Each form clearly indicated whether the measure was designed for children or for adolescents to minimize participant error. Additionally, a pre-stamped return envelope was included. In order to preserve their anonymity, participants were not required to sign and return the letter of consent, but instead were encouraged to call the lead researcher for information or referral to community resources, if needed. Participants were informed that their data would remain completely anonymous and, therefore, we would not be able to follow up with them regarding their responses. Participants were given $10 in the packet of forms as compensation for their time and assistance. Due to the nature of the data (which was de-identified throughout the duration of the study) and the complete anonymity of participants, this study met the criterion of exempt status with regard to an IRB review.

Measures

The Treatment Outcome Package for Children (TOP; Kraus et al. 2010) is comprised of 13 domains of psychological functioning for children and 11 domains of

functioning for adolescents. The TOP is written at a 5th grade reading level and takes approximately 15 min to complete in its entirety. TOP is scored based on weightings from confirmatory factor analytic studies (Kraus et al. 2005, 2010). Domain scores are then transformed into Z scores based on general population means and standard deviations.

As reviewed above, the psychometric properties of the TOP are well-documented. Construct validity of the TOP has been established for both the Adult and Child versions. Confirmatory factor analyses of the TOP's internal structure demonstrated excellent internal consistency and construct validity (Kraus et al. 2005, 2010). A study of test–retest reliability demonstrated the ability of adult TOP domain scores to remain stable and reliable over time (Kraus et al. 2005). In terms of predictive validity, the adult TOP is able to identify and predict the future performance of therapists and to provide evidence of their relative effectiveness on specific TOP domains (Kraus et al. 2011; Kraus et al. 2016). Furthermore, child TOP domain scores were able to significantly predict future placement disruptions in a child welfare sample (Alexander et al. 2016). As also noted, previous research has demonstrated concurrent validity for previous version of the Child TOP, yet a specific test of concurrent validity for the most recent versions of both the Child and Adolescent TOP is needed.

The Child Behavior Checklist (CBCL; Achenbach and Rescorla 2000, 2001) is widely considered the gold standard of psychological assessments for children. The CBCL for children (parent form) and adolescents (self-report) has excellent psychometric validity and reliability within both normative and clinical populations (Achenbach and Rescorla 2000, 2001). The child and adolescent forms consist of 120 items that can be summed to yield eight syndrome scale scores, and two broad domains of psychopathology (internalizing and externalizing). In the present study, parents of children 3–12 years of age completed either the Preschool CBCL or the School Age CBCL on behalf of their child. Adolescents 13–18 years of age completed the youth self-report form. Per the recommendation of Achenbach and Rescorla's (2001) CBCL manual, raw scores were used in order to account for the full range of score variation.

Strengths and Difficulties Questionnaire (SDQ; Goodman 1997) is a brief measure (25 items) of child and adolescent psychopathology. The validity and reliability of the SDQ have been established with acceptable results (Goodman 1997; Goodman et al. 1998; Goodman 2001). Of particular interest to this study and part of the rationale for choosing these measures is the strong concurrent validity between the CBCL and the SDQ (Goodman and Scott 1999; Muris et al. 2003), as well as the SDQ's frequent use in child welfare. In the present study, parents of children ages 4–12 years completed the SDQ on behalf of their child, and adolescents ages 13–18 years completed the self-report SDQ. The SDQ raw scores can be summed to create five domains of child and adolescent psychopathology (i.e. emotional problems, conduct problems, hyperactivity, peer problems, and a pro-social scale).

Statistical Analysis

A total of 13 correlations were derived between each of the Child or Adolescent TOP domains and the other scales (8 CBCL syndrome scales and 5 SDQ scales). Therefore, a Bonferroni-adjusted significance level of .00384 (.05/13) was calculated to minimize the likelihood of type-I error when multiple comparisons are analyzed. The concurrent validity of the TOP for children and adolescents was assessed through correlations between TOP domains and scales of the SDQ and CBCL, following the predictions outlined above. Correlations were conducted separately for children and adolescents.

Results

Descriptive results can be found in Table 1. The percentage of clinically elevated scores on at least one domain was calculated for each measure. Clinical elevations on the CBCL are defined as a *t* score of 70 or above calculated from the raw scores for the respective ages and genders (Achenbach and Rescorla 2001). Based on these criteria, 32.4 % of children and 20.0 % of adolescents had at least one clinically elevated score on any CBCL scale. Clinical elevation on the SDQ scales is defined by specific cutoff scores for parent report and youth self-report. Based on these criteria, 21.6 % of children and 56.5 % of adolescents had a clinically elevated score on at least one of the SDQ scales. Elevation on the Prosocial scale was not calculated as this scale does not indicate problems. Finally, clinical elevation on the TOP is defined as scores >2 standard deviations above general population norms (Kraus and Seligman 2004). Based on these criteria, 18.6 % of children and 8.7 % of adolescents had a clinically elevated score on at least one TOP domain.

Concurrent Validity of the TOP with Related Constructs

The correlation matrix for the child sample can be found in Table 2. The TOP ADHD scale correlated significantly with both the Hyperactivity scale of the SDQ ($r = .75$, $p < .001$, $n = 116$) and the Attention Problems scale of the CBCL ($r = .77$, $p < .001$, $n = 127$). The TOP Conduct scale correlated moderately with the CBCL Social

Table 1 Descriptive results

Children			Adolescents		
Variable	M (SD)	Elevated (%)	Variable	M (SD)	Elevated (%)
TOP (N = 156)		18.6	TOP (N = 46)		8.7
ADHD	.15 (1.19)	8.3	ADHD	.80 (1.11)	19.6
Assertiveness	−.26 (1.03)	4.5	Conduct	.64 (2.96)	8.7
Incontinence	.17 (1.28	9.8	Depression	.17 (1.15)	8.7
Conduct	.20 (1.25)	5.8	Manic	−.20 (1.00)	2.2
Depression	.18 (1.34)	10.2	Psychosis	.25 (1.24)	4.3
Psychosis	−.04 (.75)	2.6	Sleep	−.02 (1.04)	4.3
Separation anxiety	.01 (1.06)	7.1	Suicide	.37 (1.47)	6.5
Sleep	.12 (1.04)	5.8	Substance abuse	1.01 (3.46)	17.4
Resiliency (lack of)	.27 (1.44)	14.1	Violence	.16 (1.29)	8.7
Suicide	.33 (1.55)	10.3	School functioning	.10 (1.21)	8.7
Eating problems	.05 (1.14)	7.8	Social conflict	.43 (1.07)	8.7
Violence	−.28 (1.52)	7.1			
Worrisome sexual behavior	.31 (1.00)	6.5			
SDQ (N = 116)		94.0	SDQ (N = 45)		100.0
Emotional problems	1.71 (1.91)	10.3	Emotional problems	2.98 (2.57)	26.1
Conduct problems	1.18 (1.32)	6.9	Conduct problems	1.93 (2.18)	21.7
Hyperactivity	3.06 (2.78)	10.3	Hyperactivity	3.09 (2.87)	10.9
Peer problems	1.04 (1.41)	6.0	Peer problems	1.98 (1.91)	21.7
Prosocial	8.41 (1.96)	89.7	Prosocial	8.20 (1.90)	87.0
CBCL (N = 131)		32.4	CBCL (N = 45)		20.0
Anxious/depressed	3.40 (3.49)	8.2	Anxious/depressed	4.60 (4.86)	13.6
Withdrawn/depressed	1.76 (2.49)	3.9	Withdrawn/depressed	2.73 (2.86)	0
Somatic complaints	1.65 (2.34)	11.1	Somatic complaints	2.73 (3.24)	6.5
Social problems	2.77 (3.34)	30.0	Social problems	2.91 (3.31)	6.7
Thought problems	2.03 (2.28)	17.2	Thought problems	3.22 (3.64)	4.3
Attention problems	3.82 (3.94)	16.0	Attention problems	4.67 (4.15)	10.9
Rule breaking behavior	2.18 (3.22)	28.1	Rule breaking behavior	3.89 (5.71)	6.5
Aggressive problems	6.13 (7.03)	39.4	Aggressive problems	4.87 (5.56)	6.5

Problems scale ($r = .51$, $p < .001$, $n = 126$). There were strong correlations between the TOP Depression scale and both of the CBCL depression scales (Withdrawn/Depressed, $r = .66$, $p < .001$, $n = 128$; Anxious/Depressed, $r = .59$, $p < .001$, $n = 126$). The TOP Psychosis scale correlated significantly with the CBCL Thought Problems scale ($r = .55$, $p < .001$, $n = 125$). Finally, the TOP Violence scale correlated with the CBCL Social Problems scale ($r = .51$, $p < .001$, $n = 127$) and the CBCL Aggressive Behavior scale ($r = .51$, $p < .001$, $n = 127$). Each of these correlations exceeded the Bonferroni-adjusted significance level.

As can be seen in Table 3, the Adolescent TOP ADHD scale was highly correlated with both the SDQ Hyperactivity Scale ($r = .82$, $p < .001$, $n = 46$) and the CBCL Attention Problems scale ($r = .85$, $p < .001$, $n = 46$). The Adolescent TOP Depression scale correlated significantly with the SDQ Emotional Problems scale ($r = .75$, $p < .001$, $n = 46$) and both of the CBCL depression scales (Anxious/Depressed, $r = .79$, $p < .001$, $n = 46$; Withdrawn/Depressed, $r = .78$, $p < .001$, $n = 46$). Finally, the TOP Violence scale was highly correlated with the SDQ Conduct scale ($r = .68$, $p < .001$, $n = 46$), the CBCL Rule-breaking scale ($r = .68$, $p < .001$, $n = 46$), and the CBCL Aggressive Behavior scale ($r = .60$, $p < .001$, $n = 46$). Again, each of these correlations exceeded the Bonferroni-adjusted significance level.

Concurrent Validity of the TOP with Distinct Constructs

Several of the Child TOP domains (i.e., Sleep, Separation Anxiety, Incontinence, and Eating Problems) are distinct constructs that are not directly measured by either the

Table 2 Child correlations

	ADHD	Asrtv.	Incont.	Cndct.	Deprs.	Psyc.	Sepax.	Sleep	Resil.	Suicd.	Eat.	Violn.	Sexwr.
SDQ	N = 116	N = 116	N = 115	N = 116	N = 116	N = 116	N = 116	N = 116	N = 116	N = 116	N = 116	N = 116	N = 116
Emotional	.42**	.60**	.03	.21	.34**	.23	.16	−.02	.29*	.06	.24	.01	−.03
Conduct	.62**	.18	.10	.24	.40**	.21	.32*	.31*	.59**	.24*	.14	.49**	.32
Hyperactivity	.75**	.36**	.08	.28*	.37**	.29*	.27*	.22	.64**	.23	.18	.33**	.22
Peer	.36**	.45**	−.16	.08	.37**	.19	.16	.03	.33*	−.04	.27*	−.02	.00
Prosocial	−.42**	−.17	−.13	−.14	−.41**	.04	−.17	−.22	−.42**	−.10	−.13	−.09	−.14
CBCL	N = 127	N = 127	N = 124	N = 126	N = 131	N = 125	N = 127	N = 129	N = 127	N = 127	N = 125	N = 127	N = 126
Anxious/depressed	.44**	.40**	.05	.31*	.59**	.33**	.20*	.11	.38**	.34**	.12	.29*	.25*
Withdrawn/depressed	.44**	.45**	.04	.25*	.66**	.16	.15	.21	.41**	.21	.18	.18	.20
Somatic complaints	.36**	.20	.26*	.21	.36**	.14	.27*	.46**	.21	.13	.16	.09	.02
Social problems	.70**	.39**	.18	.51**	.58**	.23	.35**	.21	.51**	.25*	.02	.51**	.41**
Thought problems	.59**	.20	.08	.16	.54**	.55**	.23*	.18	.39**	.19	.08	.21	.39**
Attention problems	.77**	.36**	.19	.41**	.49**	.34**	.22	.20	.61**	.23	.15	.35**	.30*
Rule breaking behavior	.59**	.15	.08	.43**	.52**	.17	.13	.11	.62**	.38**	−.15	.48**	.61**
Aggressive behavior	.71**	.17	.15	.41**	.60**	.16	.16	.21	.69**	.23*	.01	.51**	.40**
Other problems	.56**	.24*	.49**	.36**	.41**	.17	.30*	.31*	.42**	.18	.15	.38**	.17
Total depression	.48**	.46**	.05	.31**	.68**	.28*	.20	.17	.43**	.31**	.16	.27*	.25*

Asrtv. assertiveness, *Incont.* incontinence, *Cndct.* conduct, *Deprs.* depression, *Psyc.* psychosis, *Sepax.* separation anxiety, *Sexwr.* sexual worrisome behavior, *Resil.* resiliency (lack of), *Suicd.* suicide, *Eat.* eating problems, *Violn.* violence

* $p < .01$; ** $p < .001$

Table 3 Adolescent correlations

	ADHD	Cndct.	Deprs.	Manic.	Psyc.	Sleep	Suicd.	SA	Violn.	Schoolf.	Sconf.
SDQ	N = 46	N = 46	N = 46	N = 46	N = 46	N = 46	N = 46	N = 46	N = 46	N = 46	N = 46
Emotional	.26	.28	.75**	.21	.51**	.48**	.35	.18	.22	.39	.67**
Conduct	.47*	.70**	.53*	.57**	.39*	.28	.61**	.54**	.68**	.78**	.75**
Hyperactivity	.82**	.34	.54**	.40	.28	.25	.25	.44*	.30	.57**	.48*
Peer	.14	.25	.48**	.25	.21	.40*	.50*	.16	.33	.26	.33
Prosocial	−.46*	−.39*	−.24	−.40	−.23	−.12	−.32	−.51**	−.41*	−.48*	−.27
CBCL	N = 46	N = 46	N = 46	N = 46	N = 46	N = 46	N = 46	N = 46	N = 46	N = 46	N = 46
Anxious/depressed	.31	.65**	.79**	.48*	.69**	.56**	.67**	.45*	.64**	.57**	.80**
Withdrawn/depressed	.32	.53**	.78**	.49*	.63**	.54**	.48*	.39	.43*	.48*	.69**
Somatic complaints	.17	.44*	.71**	.36	.47*	.65**	.62**	.19	.49**	.39*	.53**
Social problems	.35	.64**	.66**	.38*	.52**	.47*	.66**	.41*	.61**	.52**	.62**
Thought problems	.25	.41*	.72**	.38*	.62**	.63**	.51**	.34	.39*	.27	.45
Attention problems	.85**	.52**	.54**	.48**	.30	.27	.33	.44	.43	.66**	.52**
Rule breaking behavior	.59**	.78**	.49*	.79**	.50**	.32	.53**	.77**	.68**	.78**	.68**
Aggressive behavior	.56**	.64**	.47*	.66**	.43*	.16	.51**	.59**	.60**	.72**	.69**
Other problems	.47*	.37	.49*	.44*	.32	.43	.37	.16	.28	.39*	.35
Total depression	.33	.65**	.83**	.51**	.71**	.59**	.63**	.46*	.59**	.57**	.80**

Cndct. conduct, *Deprs.* depression, *Psyc.* psychosis, *SA* substance abuse, *Sconf.* social conflict, *Suicd.* suicide, *Violn.* violence, *Schoolf* school functioning

* $p < .01$; ** $p < .001$

CBCL or the SDQ; therefore, we did not expect that they would correlate strongly with any of the CBCL or SDQ scales. However, because several of these scales do tap into related constructs, there were some moderately sized correlations between scales reflecting distinct but nonetheless related constructs. Namely, the TOP Sleep scale correlated significantly with the Somatic Complaints scale of the CBCL for both children ($r = .46$, $p < .001$, $n = 129$) and adolescents ($r = .65$, $p < .001$, $n = 46$), and the TOP Child Separation Anxiety scale correlated significantly with the CBCL Social Problems scale ($r = .35$, $p < .001$, $n = 127$).

Discussion

The purpose of the present study was to re-test the concurrent validity of the most recent version of the TOP for children and adolescents in a community sample by analyzing correlations between TOP domain scores and the SDQ and the CBCL. Results demonstrated that the TOP for children and adolescents has strong concurrent validity and measures constructs that are similar to those measured by well-established and widely used psychological assessments.

Correlations between the TOP, the SDQ, and the CBCL were as predicted. That is, strong and significant correlations emerged between TOP domains and scales on the CBCL and the SDQ that measure theoretically similar psychological constructs. For example, the TOP Child domains of ADHD, Conduct, Depression, Psychosis, and Violence were significantly associated with related scales on the SDQ and CBCL. Similar findings emerged for adolescents, with significant and strong correlations between the TOP ADHD, Depression, and Violence domains and related constructs on the CBCL and SDQ. The Child TOP Depression domain was only moderately correlated with the CBCL depression domains, perhaps because the TOP Depression domain captures both the anxiety and withdrawal aspects of depression measured by the two CBCL depression scales. To confirm this theory, when the CBCL depression scales were combined into one total depression score, there was a stronger association between the TOP Depression domain and the CBCL Depression scale for both children and adolescents.

As predicted, several TOP domains did not yield significant correlations. For example, the Child TOP Conduct scale was not significantly associated with the SDQ Conduct scale. This is, in part, because the TOP Conduct scale taps into more serious conduct problems whereas the SDQ Conduct Problems scale is focused more on interpersonal conflict, temper tantrums, and minor theft offenses. On the other hand, there was a strong correlation between the Adolescent TOP Social Conflict scale and the SDQ

Conduct scale as both of these scales reflect interpersonal struggles more than police involvement and arrests.

Overall, the correlations between the adolescent self-reports on the TOP scales and both the CBCL and SDQ scales were visibly stronger than were the correlations for parent reports of their children. There is substantial research suggesting that adolescents tend to rate themselves differently and presumably more accurately than would their parents. For example, youth rate themselves higher on measures of both internalizing and externalizing disorders than do parents and teachers and the discrepancies for ratings of internalizing disorders increase with the age of the child (van der Ende et al. 2012). On the other hand, both self-ratings and parent-ratings are informative and provide unique information (Connor and Rueter 2009; Shakoor et al. 2011; Zou et al. 2013), suggesting the advantages of administering the TOP to multiple informants for maximal benefit.

Given the prevalent use of the CBCL in the child mental health field, the rationale for the use of the TOP must include more than its demonstrated similarity to the CBCL. One important advantage of the TOP is the infrastructure which accompanies its use, consisting of graphic depictions of a child's change over time (as rated by multiple informants), areas of agreement and disagreement between those multiple informants, and crisis alerts of suicidal and violent behavior. In addition, aggregate reports are available to mental health providers that display the initial profiles and outcomes of all clients served by a particular provider. By comparing these outcomes to risk-adjusted predictions, as we have done with adults (cf., Kraus et al. 2016), the effectiveness of clinicians in a particular TOP domain can be measured.

As opposed to child mental health, the child welfare field has tended to use the Child and Adolescent Needs and Strengths (CANS; Lyons, 2009), which takes a "communimetric" rather than psychometric approach to assessment. That is, it relies upon the caseworker's synthesis of multiple perspectives; in contrast, the TOP's approach is to solicit multiple respondents (including the child him/herself) to provide independent and private assessments of the child's behavior. In addition to the advantages of the infrastructure described above, the use of the TOP allows an empirical evaluation of the effectiveness of interventions and providers within child welfare by assessing children's actual current behavior rather than their presumed behavior if they were not currently in treatment, as required by the CANS. While the TOP has these distinct advantages over the CANS, a future comparison of these two measures is clearly warranted.

Results of the current study should be considered in light of several limitations. First, the sample size is relatively small, particularly for the adolescent sample. Second, demographics on the sample were not collected in order to maintain the privacy of participants. Third, participants were recruited through a community sample in the general population, rather than a treatment population. However, a sizeable portion of our sample had clinical elevations on each measure which indicates that even our community sample exhibited evidence of clinical severity and psychopathology. Furthermore, another advantage of a community sample is its greater applicability to a wider range of populations, including children in schools, primary care and child welfare. Finally, future research should be conducted with a purely treatment-seeking population and, in direct comparison to the CANS, a child welfare population.

In conclusion, this study offers support for the TOP's concurrent validity with two well-established measures of psychological health and functioning and provides additional evidence for the use of the TOP to measure the psychological well-being of children and adolescents in community settings.

Acknowledgments This research was funded in part by the Annie E. Casey Foundation and the Duke Endowment. We thank them for their support but acknowledge that the findings and conclusions presented in this report are those of the authors alone, and do not necessarily reflect the opinions of these foundations.

Funding This research was funded in part by the Annie E. Casey Foundation (Grant No. 212.0369) and the Duke Endowment (Grant No. 1935-SP).

Compliance with Ethical Standards

Conflict of interest Author David R. Kraus is the developer and owner of the Treatment Outcome Package.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the American Psychological Association and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent All data were de-identified throughout the duration of the study and participants remained completely anonymous. Therefore, this study met the criterion of exempt status with regard to an IRB review and did not require informed consent from participants.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Alexander, P.C., Bentley, J.H., Kraus, D.R., Baxter, E.E., Boswell, J.F., & Castonguay, L.G. (2016). Placement disruptions and the predictive validity of the Treatment Outcome Package (TOP). Manuscript submitted for publication.

Beacon Health Strategies (2008). Alert score analysis report. Neighborhood Health Plan. Unpublished report.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck depression inventory-II*. San Antonio, TX: Psychological Corporation.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., & Dahlstrom, W. G. (2001). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring* (rev ed.). Minneapolis, MN: University of Minnesota Press.

Connor, J., & Rueter, M. (2009). Predicting adolescent suicidality: Comparing multiple informants and assessment techniques. *Journal of Adolescence, 32*, 619–631.

Derogatis, L. R. (1975). *Brief symptom inventory*. Baltimore: Clinical Psychometrics Research.

Eisen, S. V., Grob, M. C., & Klein, A. A. (1986). BASIS: The development of a self-report measure of psychiatric inpatient evaluation. *Psychiatric Hospital, 17*, 165–171.

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*, 581–586.

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire (SDQ). *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 1337–1345

Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry, 7*, 125–130.

Goodman, R., & Scott, S. (1999) Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology, 27*, 17–24.

Hodges, K. (2000). *Child and adolescent functional assessment scale*. Ypsilanti, MI: Eastern Michigan University.

Kelly, K., O'Donnell, R., Pelletier, A., & Simmons, J. (2008). *Behavioral health outcomes program update.* Blue Cross and Blue Shield of Massachusetts. Unpublished report.

Kraus, D.R., Bentley, J.H., Alexander, P.C., Boswell, J.F., Constantino, M., Baxter, E.E., & Castonguay, L.G. (2016). Predicting therapist effectiveness from their own practice-based evidence. *Journal of Consulting and Clinical Psychology*. Advance online publication.

Kraus, D. R., Boswell, J. F., Wright, A. G. C., Castonguay, L. G., & Pincus, A. L. (2010). Factor structure of the treatment outcome package for children. *Journal of Clinical Psychology, 66*, 1–14.

Kraus, D. R., Castonguay, L., Boswell, J. F., Nordberg, S. S., & Hayes, J. A. (2011). Therapist effectiveness: Implications for accountability and client care. *Psychotherapy Research, 21*, 267–276.

Kraus, D. R. & Seligman, D. A. (2004). *TOP user's manual*. Ashland, MA: Behavioral Health Laboratories.

Kraus, D. R., Seligman, D. A., & Jordan, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The Treatment Outcome Package. *Journal of Clinical Psychology, 61*, 285–314.

Lyons, J. S. (2009). *Communimetrics: A theory of measurement for human service enterprises*. New York: Springer.

McHorney, C. A., Ware, J. J., & Raczek, A. E. (1993). The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care, 31*, 247–263.

Muris, P., Meesters, C., & van den Berg, F. (2003). The Strengths and Difficulties Questionnaire (SDQ): Further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European Child and Adolescent Psychiatry, 12*, 1–8. doi:10.1007/s00787-003-0298-2.

Reynolds, C. R., & Kamphaus, R. W. (1992). *Manual for the BASC: Behavior assessment system for children*. Circle Pines, MN: American Guidance Service.

Shakoor, S., Jaffee, S. R., Andreou, P., Bowes, L., Ambler, A. P., Caspi, A., et al. (2011). Mothers and children as informants of bullying victimization: Results from an epidemiological cohort of children. *Journal of Abnormal Child Psychology, 39*, 379–387.

Stelk, W., & M. Berger (2009, June). Predictive modeling: Using TOP clinical domain items to identify adult Medicaid recipients at risk for high utilisation of behavioral health services in a managed care provider network. *Paper presented at the 40th annual meeting of the Society for Psychotherapy Research*, Santiago, Chile, 2009.

van der Ende, J., Verhulst, F. C., & Tiemeier, H. (2012). Agreement of informants on emotional and behavioral problems from childhood to adulthood. *Psychological Assessment, 24*, 293–300.

Zou, C., Schimmack, U., & Gere, J. (2013). The validity of well-being measures: A multiple-indicator—multiple-rater model. *Psychological Assessment, 25*, 1247–1254.