

The Adolescent Psychotherapy Q-set (APQ): a validation study

Ana Calderon, PhD^{ab}, Celeste Schneider, PhD^c, Mary Target, Professor^{ab},
Nick Midgley, PhD^{ab}.

^a Department of Clinical, Educational, and Health Psychology,
University College London, London, UK.

^b Anna Freud National Centre for Children and Families, London, UK.

^c San Francisco Center for Psychoanalysis, San Francisco, California,
USA.

Ana Calderon is a Research Fellow at Evidence-Based Practice Unit
(EBPU) at Anna Freud National Centre for Children and Families.

Celeste Schneider is a Member and Faculty at the San Francisco Center
for Psychoanalysis and Faculty at the Psychoanalytic Institute of
Northern California

Mary Target is Professor of Psychoanalysis and Director of the MSc in
Theoretical Psychoanalytic Studies at University College London.

Nick Midgley is Director of the MSc in Developmental Psychology &
Clinical Practice, Academic Director of the PsychD Child and
Adolescent Psychotherapy, and Child and Adolescent Psychotherapist in
Family Support Services, at Anna Freud National Centre for Children
and Families

Contact information corresponding author:

12 Maresfield Gardens, London, NW3 5SU, UK,

a.calderon@ucl.ac.uk

Acknowledgements: This work was partly funded by the CONICYT

PAI/INDUSTRIA 79090016. The validation of the measure was
supported by many colleagues working on the IMPACT Study

(Goodyer, 2011), without whom this study could not have been conducted.

Abstract

Objective: This article reports the validation of the Adolescent Psychotherapy Q-set (APQ), a newly developed instrument, adapted from the well-established Psychotherapy Q-Set (PQS) and the Child Psychotherapy Q-set (CPQ). The APQ aims to describe the psychotherapy process in the treatment of adolescents in a form suitable for quantitative comparison and analysis.

Method: The validation was conducted with the ratings of seventy audio-recorded youth psychotherapy sessions from a range of therapists, patients, and treatment stages, using two therapeutic approaches (Short-Term Psychoanalytic Psychotherapy and Cognitive Behavioural Therapy). Data analysis included intraclass correlation coefficients, Q-factor analysis, non-parametric mean differences, and Pearson correlations.

Results: Results suggest that the APQ has good levels of interrater reliability, is able to identify differences and similarities of two therapeutic approaches, and good convergent and discriminant validity with a widely-used measure of therapist behaviours (the Comparative Psychotherapy Process Scale).

Conclusions: The APQ reported good levels of validity and reliability. It is hoped that it will contribute to new ways of investigating the mechanisms of therapeutic change for those working with adolescents.

Introduction

Despite the significant progress that has been made in outcome psychotherapy research, it is still not possible to provide a comprehensive, evidence-based explanation for *how* or *why* treatments produce change (Kazdin, 2007). One possible explanation for psychotherapy research's inability to identify consistent and strong correlations between process dimensions and treatment outcome could be

that most studies have attempted to find simple and direct association without considering other variables; for example, therapist activity and transference interpretations alone have not been consistent predictors of change, but have been able to predict change in interaction with certain patient qualities (Jones, Cumming, & Horowitz, 1988). Research supports a multiple factor view of psychotherapy effects that includes independent roles in the prediction of treatment outcome for patient, treatment, relationship, and patient-therapist matching variables, as well as an interaction of factors (Beutler et al., 2003).

It follows that traditional data analysis techniques that only include a few variables in the analyses are likely to give an incomplete picture of the process of psychotherapy. It is clear that more complex data analysis techniques (Beutler et al., 2003) and different methodologies (Kazdin, 2000) are needed. One of those methodologies is Q-methodology, which provides a holistic approach to the phenomena under study, i.e. it does not start by examining a few variables but explores how all the variables relate to each other by using Q-factor analysis (Watts & Stenner, 2012). Q-methodology differs from more traditional research in that it allows the discovery of associations among various aspects of the therapeutic process, instead of limiting the study to a particular dimension of presumed theoretical importance for the therapy process (Jones et al., 1988).

The Psychotherapy Process Q-Set (PQS; Jones, 1985) was developed to study process in psychotherapy of adults and has been used to examine process predictors of what works for whom (Jones et al., 1988), to track the treatment process over time (Jones, Parke, & Pulos, 1992), to compare the therapy process in different types of treatments (Jones & Pulos, 1993), to associate specific techniques with outcome (Price & Jones, 1998), and to study the adherence of different treatments to their theoretical orientations (Ablon & Jones, 1998). This has been achieved in a range of research designs: single-case studies, naturalistic studies, and large randomized controlled trials. The contributions to the adult psychotherapy process-outcome research of the PQS have been of immense value (Smith-Hansen, Levy,

Seybert, Erhardt, & Ablon, 2012), and excellent and detailed summaries of the last 25 years of contributions of the PQS can be found in Ablon, Levy, and Smith-Hansen (2011) and Smith-Hansen et al. (2012).

The Child Psychotherapy Q-Set (CPQ; Schneider & Jones, 2004), an adaptation of the PQS for the study of child play therapy process, has been used in similar ways to the PQS, i.e. in single-case and group designs, in adherence studies, in linking psychotherapy process to outcome, in clinical supervision, among others. The CPQ has been used, for example, by Schneider, Pruetzel-Thomas, and Midgley (2009), to study the differences and similarities between cognitive behavioural therapy (CBT) and psychodynamic treatments finding that children present themselves in similar ways in both CBT and psychodynamic treatments, but that therapists use different techniques depending on their theoretical background. Additionally it has been employed to examine interaction structures in the therapy of children with Asperger's Disorder (Goodman & Athey-Lloyd, 2011), and those with emerging borderline personality disorder (Goodman, 2015); to explore the distinct and overlapping features of CBT and psychodynamic therapy with children (Goodman, Midgley, & Schneider, 2015), and to help assess competence in the supervision of child therapy trainees (Goodman, 2010).

The PQS and CPQ cover the psychotherapeutic process of adults and children, respectively. In the interest of creating prototypes for adolescent treatment Bambery, Porcerelli, and Ablon (2007) modified the CPQ by changing the word "child" to "adolescent". As the developmental stage and the therapeutic process with adolescents has many distinct features (see for example Jacobson & Mufson, 2010; Verduyn, Rogers, & Wood, 2009), there is a need to construct instruments and adapt research designs specifically to this population. Research into the treatment of adolescents should account for issues that are unique to this age group, such as the emergence of sexual interest, the development of self-identity, the search for autonomy from parents, and the newly developed capacity for perspective-taking and abstract and logical thinking (Tolan & Titus, 2011). Therefore there is a need for an adolescent-specific

psychotherapy process measure, rather than using a child-specific version with adolescents, as described by Bambery et al. (2007).

Furthermore, although a number of process measures exist, they either focus on specific aspects of the therapeutic process, such as therapeutic alliance (e.g. McLeod & Weisz, 2005) or therapist techniques (Kronmüller et al., 2010; Weersing, Weisz, & Donenberg, 2002) or else they were developed for use with younger children in therapy, where the medium of communication is usually more play-based (Estrada & Russell, 1999; Kernberg, Chazan, & Normandin, 1998; Schneider & Jones, 2004). Hence, there is a need for a measure designed specifically for the psychotherapy process of adolescents that can address the complexity of an entire session, and that allows for comparisons between therapeutic modalities. The development and validation of such an instrument is presented in this article.

Development of the APQ

Description of the APQ

The APQ is an adaptation of the PQS and the CPQ. Like those instruments, the APQ is a Q-set composed of 100 items that describe three aspects of a psychotherapeutic process: (1) the *young person's* feelings, experience, behaviour, and attitudes (e.g. item 8: "Young person expresses feelings of vulnerability"); (2) the *therapist's* attitudes and actions (e.g. item 33: "Therapist adopts a psychoeducational stance"); and (3) the nature of the *interaction* of the dyad (e.g. item 38: "Therapist and young person demonstrate a shared understanding when referring to events or feelings"). In order to ensure interrater reliability, a coding manual details instructions for the rater and provides descriptions and examples for each of the items.

Items describe psychotherapeutic processes in terms of linguistic and behavioural cues, the absence or presence of which can be observed in the clinical material with minimal inference. In addition, items aim to describe psychotherapy processes avoiding theoretical jargon. The unit of observation is the entire session, not

just small segments. The method can be applied to verbatim transcripts, audiotapes or videotapes of the entire treatment session.

The rating procedure is straightforward. After studying the record of a psychotherapy session and the manual, raters order the 100 items into a row of nine categories. This can be done manually (using printed cards), or online, using a specially designed website (<http://www.homepages.ucl.ac.uk/~ucjtaca/>). At one end raters place those items believed to be the *most characteristic* with reference to the understanding of the material, while at the other end raters place those items believed to be *most uncharacteristic*. The number of items sorted into each pile ranges from 18 in the middle to five at the extremes, and form a quasi-normal distribution.

It is important to highlight that despite sharing the methodology and procedures of the PQS and CPQ, the APQ aims to capture what is characteristic and unique to the psychotherapy process of an *adolescent aged 12 to 18*.

Development iterations

The development of a Q-Set is an iterative process. An initial draft of the APQ was constructed between 2008 and 2009. A report on the early development of the APQ, face validity, and item coverage can be found in Bychkova, Hillman, Midgley, and Schneider (2011). During the following three years the APQ went through six iterations, which included analysis of experts' qualitative feedback, and the coding and analysis of 27 psychotherapy sessions from different therapists, young people, and therapeutic approaches (a detailed description of each of the iteration analysis can be found in the author's PhD thesis that can be accessed upon request).

Six principles guided the process of selection and creation of the APQ items in each of the iterations: (1) items had to be relevant for the psychotherapeutic process of an adolescent patient; (2) items had to be as theoretically neutral as possible (the wording of the items should not be solely related to one therapeutic modality but to a wide range of interventions, events, and processes that could be observed in several treatment orientations); (3) items had to describe

the therapeutic process without entailing a judgment as to whether what the therapist did or said was 'good' practice or not; (4) items had to describe a process rather than the content of the session; (5) items had to be written in the most specific way possible, in order to avoid a high level of inference from the raters in the rating procedure; (6) the items that the APQ shared with the PQS and/or CPQ had to be kept as similar as possible to the original item, unless there was a need for revision, based on the previous five principles.

As aforementioned, the APQ's items had to be relevant for the psychotherapeutic process of an adolescent, aged 12-18. Some examples of items that were created because they had been identified as potentially significant elements of youth therapy in our review of the literature on therapy with adolescents, and were not part of either the PQS nor the CPQ, included: "Young person's experience of his/her body is discussed" (item 79), "Young person feels rejected or abandoned" (item 41), and "Young person feels unfairly treated" (item 55). For the same reason, PQS and CPQ items that were kept because of their relevance were: "Self-image is a focus of the session" (item 35), and "Young person explores sexual feelings and experiences" (item 11).

In the end of the development process, the APQ shared 45 items with both the PQS and CPQ, 18 items only with the PQS, 4 items solely with the CPQ, and had 33 unique items; 40 items attempted to capture young person's feelings, experience, behaviour, and attitudes; 30 items therapist's attitudes and actions; and 30 items alluded to the nature of the interaction of the dyad.

Method

The validation study for the APQ had three aims: (1) an assessment of the level of consistency across independent raters; (2) an assessment of the ability of the APQ to identify differences and shared features of two different therapeutic approaches to working with adolescents; and (3) convergent and discriminant validity with a widely-used measure of therapist behaviours, the Comparative

Psychotherapy Process Scale (CPPS-ER; Hilsenroth, Blagys, Ackerman, Bonge, & Blais, 2005).

All the audio-recorded psychotherapy sessions for the development and validation of the APQ were provided by the IMPACT study (Improving Mood with Psychoanalytic and Cognitive Therapies; Goodyer et al., 2011). The IMPACT study is a multicentre randomized controlled trial that provides three therapeutic interventions (Short-Term Psychoanalytic Psychotherapy [STPP], Cognitive-Behavioural Therapy [CBT], and Specialist Clinical Care [SCC]) to adolescents with moderate to severe depression. Participants were recruited from clinical referrals to Child and Adolescent Mental Health Services in three UK regions. For more details about procedures followed to recruit participants, eligibility and exclusion criteria please refer to Goodyer et al. (2011).

Sample of recordings

Sample size

Data were analysed with Q-technique, which has important consequences for calculating the appropriate sample size. The ideal sample sizes for R and Q-factor analyses differ greatly. On the one hand, recommended absolute sample sizes for R-factor analysis vary from 100 to 1000, whilst participants to variable ratios vary from 3:1 to 20:1 (Mundfrom, Shaw, & Tian, 2005). On the other hand, in Q-factor analysis recommendations state that a sample of 40 to 50 participants is considered enough because it provides an adequate picture of the subject under study (Stainton Rogers, 1995), or that the ratio of participants to variables should be of 1:2 (Kline, 1994). This huge variation of the final number of participants for Q and R factor analysis is based on the difference in the structure of the data matrices: in R variables are in the columns and participants in the rows, whilst in Q variables are in the rows and participants in the columns. Thus, the rule for R sample size that "p cannot exceed N" (Velicer & Fava, 1998, p. 247) coincides with Watts and Stenner's (2012) suggestion to have fewer participants than the number of items in the Q-set.

In light of the above considerations, a total of 70 audio-recorded psychotherapy sessions were randomly sampled from the IMPACT study. This was considered adequate for a 100-item Q-set because it follows Q-recommendations for participants sample size.

Sampling strategy and sample selection.

In August 2013 the national IMPACT dataset had audio-recordings for 80 STPP cases and 62 CBT cases¹. The first and last sessions were excluded in both treatment arms because it was considered that it would not be expected to see typical therapeutic process in the first or last CBT or STPP sessions. Sessions in which the parent was present were also excluded, as the APQ was not designed to capture the therapy process in groups or family sessions. A random selection of cases for CBT and STPP was conducted to reach the target of 70 recordings explained above (35 from each therapeutic modality). All recordings corresponded to different cases.

Characteristics of the sampled recordings.

Duration of recordings ranged from 22 to 94 minutes, with an average of 48.04 minutes ($SD = 12.52$). Separated by treatment arm, CBT recordings ranged from 26 to 94 minutes, and had an average length of 51.46 minutes ($SD = 15.19$). STPP recordings ranged from 22 to 54 minutes, with an average length of 44.63 minutes ($SD = 7.96$). Sessions came from all stages of therapy.

Characteristics of participating clinicians.

Cases were treated by 45 different therapists (24 STPP and 21 CBT), eight (17.78%) of whom were men and 37 (82.22%) women. The majority of the therapists treated only one patient (29 therapists or

¹ SCC was not included in the sample of tapes because although it involves a conversational approach just like STPP and CBT, it regularly includes the young people's parents and family member in the sessions, so cannot be considered an individual therapy.

64.44%), followed by 12 therapists (26.67%) who treated two patients, and three therapists (6.67%) who treated three patients. Only one therapist treated five patients. Therapists followed the manuals provided by IMPACT study for each therapeutic approach (Cregeen, Hughes, Midgley, Rhode, & Rustin, In press; IMPACT study CBT Sub-group, 2010).

Characteristics of participating young people.

Patient age at baseline averaged 15.9 years ($SD = 1.51$), and ranged from 11.8 to 17.9. Regarding gender, 21 (30%) were boys and 49 (70%) girls. All participants met criteria for Major Depressive Disorder.

Among the cases sampled for the validation study, the number of sessions attended ranged from 2 to 29, with a mean of 15.09 sessions ($SD = 7.73$). Patients in CBT treatment received a minimum of two and a maximum of 24 sessions, with a mean of 11.85 sessions ($SD = 6.01$). On the other hand, young people in STPP treatment had a minimum of 6 and a maximum of 29 sessions, with a mean of 18.55 sessions ($SD = 7.95$).

Measure

Comparative Psychotherapy Process Scale (CPPS)

The CPPS is a measure that was created to assess the degree to which a therapist uses techniques of psychodynamic-interpersonal (PI) and/or cognitive behaviour psychotherapy (CB) in an entire psychotherapy session. It was developed by Hilsenroth et al. (2005) based on two empirical reviews of the comparative psychotherapy process literature in adults (Blagys & Hilsenroth, 2000, 2002). The CPPS is composed of 20 items, 10 of which correspond to the PI scale and 10 to the CB scale. Items are rated on a 7-point Likert scale ranging from 0 (Not at all characteristic) to 6 (Extremely characteristic), and there are no reversed items. Although the CPPS was developed for use in studies of adult psychotherapy, it is currently being used to assess treatment adherence in the large randomized controlled trial of youth psychotherapy from which these recordings were sampled (IMPACT, see Goodyer et al., 2011). As no

similar measure developed specifically for adolescents was available, it was selected as an appropriate measure for assessing convergent validity with the therapist technique elements of the APQ.

The psychometric properties of the CPPS have been well established in psychotherapy with adults (R. E. Goldman, Hilsenroth, Owen, & Gold, 2013; Hilsenroth, 2007). Internal consistency of both scales has been good to excellent: Cronbach's α of .82 to .92 for the PI scale and .75 to .94 for the CB scale (R. E. Goldman et al., 2013; Hilsenroth et al., 2005). Interrater reliability has also been between good (ICC between .60 and .74) and excellent (ICC \geq .75) across multiple studies (G. A. Goldman & Gregory, 2009; R. E. Goldman et al., 2013; Hilsenroth et al., 2005; Stein, Pesale, Slavin, & Hilsenroth, 2010).

Internal consistency of the CB and PI scales in this study was excellent with a Cronbach's α of .91 for the CB scale and .87 for the PI scale. Agreement between raters was examined using the two-way random absolute agreement intra-class correlation (ICC). The mean ICC was .78, and a total of 49 sessions (70%) had an excellent ICC², 14 sessions (20%) had a good agreement, four sessions (5.7%) had a fair agreement, and three sessions (4.3%) had poor agreement. In order to ensure the best possible ratings for this study, and following a similar procedure to G. A. Goldman and Gregory (2009), a third rater was asked to independently rate the seven sessions with ICCs lower than .60. Then, the two ratings that agreed best were combined.

Training of raters

A total of seven research assistants were trained in the use of the CPPS over the course of four months. During the training, a total of 12 sessions were coded, three of which were IMPACT sessions. All

² ICC level interpretations were based on Fleiss (1981): excellent agreement (ICC \geq .75); good agreement (ICC between .60 and .74), fair agreement (ICC between .40 and .59), and poor agreement (ICC $<$.40).

raters achieved a sufficient interrater reliability to code on their own (i.e., ICC of .70 or above). Raters completed the ratings over the course of a nine-month period, with ongoing monitoring and feedback to avoid rater drift.

A total of six child and adolescent psychotherapists were trained in the use of the APQ over the course of two months. During the training a total of 10 sessions were coded, all of which were IMPACT sessions. The six raters achieved a sufficient interrater reliability to code on their own (i.e., ICC of .70 or above). Raters completed the ratings over the course of a nine-month period, with ongoing monitoring and feedback to avoid rater drift.

Procedure and Data Analysis

In order to assess the level of consistency across independent raters, a total of 33 audio-recorded sessions (47 percent of the total sample of recordings) were double-coded by one of the article's authors (Author's initials) and a total of six trained child and adolescent psychotherapists. The author rated all the 33 sessions with the APQ, three raters coded seven CBT recordings (20 percent out of the total CBT sample), and five raters coded 26 STPP recordings (74 percent out of the total STPP sample). The latter percentage was higher because the STPP sessions were coded with the APQ for another independent study and, hence, there were more ratings available to compare.

ICCs were calculated for each session with pairs of ratings using the two-way random consistency model (Shrout & Fleiss, 1979). The ICC for each of the APQ items was not calculated because it was considered not appropriate for this measure for two reasons. Firstly, because of the forced distribution 50 items are always placed in the middle piles of the distribution (i.e. piles 4, 5 or 6); and low variation of scores might distort ICCs (Lahey, Downey, & Saal, 1983). Secondly, it is not expected that raters will place the items in exactly the same pile as other raters but rather that there should be a *consistency* in what was considered characteristic, uncharacteristic, or neutral in the sessions. It is more relevant to this measure to

calculate which items have the biggest discrepancies. Hence, differences for each item in each pair of sessions' ratings was calculated and were summed (the total CBT comparisons were 7, and total STPP comparisons were 42 because some sessions had three ratings).

In order to assess the convergent and discriminant validity of the APQ with the CPPS, a Q-factor analysis was firstly conducted, which is a data reduction technique that groups sessions instead of variables. The 70 complete Q-sorts were analysed with a Centroid Factor Analysis and varimax rotation. The resulting groups of sessions (Q-factors) were used to explore the APQ's convergent and discriminant validity.

As all the sessions had two CPPS ratings, the first step was to calculate a composite score for each session. Next, the CPPS ratings in the groups created with the Q-factor analysis reported in the previous chapter were examined with Wilcoxon signed-rank tests for the within groups effects, and with Kruskal-Wallis tests for the between group effects (with Mann-Whitney tests for post-hoc tests). Non-parametric tests were used to examine the differences in the PI and CB scores within and between Q-groups because scores were not normally distributed and the small sample size of some of the factors. In cases like this, when normality cannot be assumed, non-parametric tests are recommended (Field, 2009). Bonferroni correction was applied when appropriate in order to control for the familywise error due to multiple significance testing.

Then, in order to examine whether the therapists' techniques observed in the CPPS scales in the Q-factors were also captured by the APQ, Pearson correlation coefficients were calculated between the factor loadings of the 60 sessions that had significant loadings in the Q-factor analysis and the composite raw scores on the two CPPS scales. If the APQ is a valid instrument, factors composed of therapists using principally psychodynamic techniques should correlate positively and highly with the PI scale (convergent validity), and negatively and highly with the CB scale (discriminant validity). And vice versa, factors composed of therapists using principally

cognitive-behavioural techniques should correlate positively and highly with the CB scale (convergent validity), and negatively and highly with the PI scale (discriminant validity).

Q-factor analysis was conducted using the software PQMethod, version 2.33 (Schmolck, 2002), which provides optimal support for entering and factor-analysing Q-sort data having been purpose-built for this kind of analysis (Watts & Stenner, 2012). SPSS version 22 (IBM SPSS Statistics, Hampshire, UK) was used for the correlations.

Ethical Considerations

Ethical approval was granted as part of the ethics for the overall IMPACT study (Goodyer et al., 2011). Confidentiality of the material was ensured by several means: sessions were anonymized; all recordings were encrypted using TrueCrypt® (<http://www.truecrypt.org/>); raters had access to only the sessions they were coding; no rater belonged to a service in which either therapist or young person was known; and there was no personal contact with either therapists or young people.

Results

Interrater reliability

Interrater reliability of the APQ ratings was good, with a mean ICC of .73 for the CBT sessions (ranging from .65 to .81), and a mean ICC of .72 for the STPP sessions (ranging from .44 to .88). Out of the 26 STPP sessions, eight sessions or 24 percent were in the excellent range, 17 sessions or 52 percent were in the good range, one session was in the fair agreement range, and none was in the poor agreement range. In relation to the seven CBT sessions, three were in the excellent range, four were in the good range, and no session was either in the fair agreement or the poor agreement range.

Q-factor analysis, first step to assess therapist's techniques and convergent and discriminant validity.

In order to assess the ability of the APQ to identify differences and shared features of two different therapeutic

approaches to working with adolescents and to assess the APQ's convergent and discriminant validity with the CPPS, the first step was to conduct a Q-factor with the APQ codings of the 70 sessions.

Using Watt and Stenner's (2012) criteria, a four factor model was used. The four factors accounted for 49.98% of the variance, which is higher than the 35-40% that is considered as a sound solution in factor analysis (Watts & Stenner, 2012).

The next step was to identify the Q-sorts that had significant loadings on each factor. This resulted in a total of 60 sessions: 19 sessions flagged for Factor 1, 25 for Factor 2, 10 for Factor 3, and 6 for Factor 4. Z-scores based on factor estimates were, then, calculated for each factor (Watts & Stenner, 2012).

Table 1 presents the most and least characteristic items of each Factor (items that in the factor array of each factor were in pile 1,2,8, or 9). Although all the items have Z scores in all the factors, only the most and least characteristic items are presented in the table in order to make it easier to read.

Factor 1 had an EV of 10.57, accounted for 15.1% of the variance, and had an excellent internal consistency (Cronbach's $\alpha = .93$). It was composed of 17 STPP and two CBT sessions. The sessions in this factor were characterised by therapists who were not directly reassuring (-1.26)³, but focused the discussion on the therapy relationship (1.32), made links to situations in young people's past (0.61), and paid attention to young people's feelings about breaks and interruptions of the therapy process (0.71).

Factor 2 had an EV of 14.4, accounted for 20.57% of the variance, and had an excellent internal consistency (Cronbach's $\alpha = .96$). It included 25 sessions: 23 CBT and two STPP. Therapists in these sessions actively structured the sessions and asked questions (1.83; 2.01), expressed their opinion either implicitly or explicitly (-2.15), shared their emotions with the young person (0.93), and offered explicit advice and guidance (0.96). Also, therapists in this

³ Indicates the Z-score for specific item(s).

group of sessions provided psycho-education (1.34), actively encouraged the young people to reflect on their symptoms (1.44), and discussed specific activities or tasks for the young people to attempt outside of session, which mostly included homework (1.31).

Factor 3 had an EV of 5.91, accounted for 8.44% of the variance, and had a good internal consistency (Cronbach's $\alpha = .86$). It was composed of 10 sessions: 5 CBT and 5 STPP. In this factor, therapists actively structured the sessions (1.75), asked for more information or elaboration (2.33), provided psychoeducation (1), reflected on symptoms (0.95), and discussed specific tasks for the young person to conduct outside the session (0.74). One important therapists' activity was the rephrasing of young people's communication (1.33).

Finally, Factor 4 had an EV of 4.11, accounted for 5.87% of the variance, and had a good internal consistency (Cronbach's $\alpha = .72$). Six STPP sessions were exemplary of this factor. The therapists in this factor employed techniques that are associated to the psychoanalytic model of work more frequently than in any of the other factors: they focused in the therapy relationship, connected it to other relationships, paid attention to breaks and interruptions in therapy, and drew attention to young people's non-verbal behaviour (2.2; 0.84; 2.7; 0.7). In addition, they actively avoided techniques that are associated with a CBT model: they did not focus the discussion on the goals of the therapy (-0.8), refrained from providing explicit advice and guidance (-1.24), did not encourage the young person to behave differently with others (-0.73), and instead of adopting a psycho-educational stance these therapists explored the young people's concerns about their symptoms (-1.04). Also, therapists actively challenged young people's views (0.94; 1.39; 1.42) and drew attention to what young people considered as unacceptable feelings (1.57).

Therapist's techniques by Q-groups

Within-groups

Results showed that in Factor 1, which was composed of 89% of STPP sessions, therapists used significantly more PI techniques (*Mdn* =

2.65) than CB techniques ($Mdn = 0.4$), $T = -190$, $p < .001$, $r = -.62$ (see Table 2). The opposite was true for the second factor, which was composed by 92% of CBT sessions: therapists used significantly more CB techniques ($Mdn = 2.55$) than PI techniques ($Mdn = 1.3$), $T = 279.5$, $p = .002$, $r = -.45$. In the third factor, in which there was an equal amount of CBT and STPP sessions, therapists did not use techniques from one modality significantly more than techniques associated with the other (CB $Mdn = 1.4$ and PI $Mdn = 2$, $T = -32.5$, $p = .61$, $r = -.11$). Finally, the fourth factor, which was composed of only STPP sessions, failed to be significant after the Bonferroni correction. However, descriptively, therapists did use more PI than CB techniques (CB $Mdn = 0.57$ and PI $Mdn = 2.65$, $T = -21$, $p = .028$, $r = -.64$).

Between-groups

There was a significant difference in the CB subscale scores in the four groups ($H(3) = 37.41$, $p < .001$). Mann-Whitney tests were used to follow up this finding. A Bonferroni correction was applied and so all effects are reported at a .008 significance level ($.05/6 = .008$). CB scores were significantly higher in the factor composed mostly of CBT sessions (Factor 2) compared to the two factors with most STPP sessions (with Factor 1 $U = 4$, $p < .001$, $r = -.59$, and with the Factor 4 $U = 0$, $p < .001$, $r = -.48$). Scores in the CB scale failed to be significant after the Bonferroni correction between Factor 2 and Factor 3 ($U = 62.5$, $p = .022$, $r = -.27$), and were not significant for the rest of the comparisons.

In addition, there was a significant difference in the PI scores in the four groups ($H(3) = 17.545$, $p = .001$). Again, Mann-Whitney tests were used and the same Bonferroni correction was applied because of multiple testing. The only significant difference after Bonferroni correction in the PI scores was between the factor composed mostly of STPP sessions (Factor 1) and the factor composed mostly of CBT sessions (Factor 2; $U = 81$, $p < .001$, $r = -.40$). Differences in the PI scale between the two factors with mostly STPP sessions (Factor 1 and Factor 4) failed to be significant after Bonferroni correction ($U =$

27.5, $p = .017$, $r = -.30$). The rest of the comparisons were not significant.

Convergent and discriminant validity of APQ with therapists' techniques.

The four Q-factors were also used to assess the convergent and discriminant validity of APQ with therapists' techniques. As results in Table 2 show, the two factors in which STPP sessions had higher factor loadings (Factor 1 and Factor 4) were positively and significantly correlated with the PI scale on the CPPS, whilst they were negatively and significantly correlated with the CB scale. In addition, the opposite was true for the factor where CBT sessions had higher factor loadings (Factor 2): it was positively and significantly correlated with the CB scale, and negatively and significantly correlated with the PI scale. Finally, the factor that was composed of roughly the same number of STPP and CBT sessions (Factor 3) presented low and non-significant correlations with both the PI and CB sub-scales on the CPPS.

Discussion

The results of these studies provide empirical support for the psychometric properties of the APQ. Inter-rater reliability was achieved when the APQ was applied to rating the process of psychotherapy of a young person. Additionally, the results demonstrated the capacity of the APQ to capture and differentiate between the techniques used by CBT and STPP therapists.

Overall, APQ ratings presented good levels of interrater agreement (i.e. ICCs of .70 and above) and only one session had a low agreement between the author's ratings (author's initials) and the other two raters (despite those two raters having a good level of agreement between them). After re-listening the session it was noticed that it included a young person who was silent and wanted to play or draw instead of talking. It is possible, then, that the APQ works best in sessions where the young person communicates with words instead of playing or drawing, and it might be worth considering the use of the

CPQ when the sessions are more play-based, independently of the chronological age of the young person.

With the aim of exploring whether therapist's techniques varied across the four Q-groups, the CPPS subscales were analysed in the four Q-groups with non-parametric tests. The sessions in Factor 1 (composed mostly of STPP sessions) used significantly more PI than CB techniques, and the sessions in the Factor 2 (composed mostly of CBT sessions) used significantly more CB than PI techniques. In addition, in the between-group analyses Factor 1 and Factor 2 differed significantly in both scales. These results coincide with the APQ factor description made of the therapists' techniques in Factor 1 and Factor 2 (i.e. that in Factor 1 therapists used mostly STPP techniques, and in Factor 2 therapists used mostly CBT techniques). They also indicate that the APQ was able to identify the same trend that was distinguished by the CPPS analyses, providing evidence of the APQ's ability to identify and differentiate between the techniques that therapists use in different therapeutic modalities. In future research, when IMPACT outcome data becomes available, it would be interesting to link these results to outcome. One previous study (Owen, Hilsenroth, & Rodolfa, 2013) found that therapies that had high levels of PI scores and low levels of CB scores in the CPPS and a good working relationship were related to high levels of post-session gains; whilst, therapies that had high levels of CB scores and low levels of PI scores in the CPPS and good working relationships were not associated with post-session gains.

Interestingly, in Factor 3 therapists did not significantly use any set of modality-specific techniques more than the other. Based on the APQ item configuration, however, Factor 3 appeared to have active therapists who used techniques associated with CBT (such as the provision of psychoeducation), irrespective of the therapists' original theoretical orientation. Descriptive statistics of the CPPS showed that CB scores were indeed lower than PI scores, but at the same time both scores were very low implying that therapists did not frequently employ either set of techniques during the sessions (both means were lower than 1.8). This might be highlighting an important

distinction between the CPPS and the APQ. In the CPPS scores represent the average amount of techniques used in the treatment (Owen et al., 2013); whilst in the APQ items' ratings are related to the different techniques used and their relevance for the entire session. Hence, for example, if the therapist only provided psychoeducation on one occasion the CPPS final score would be low, while the same could have a higher rating in the APQ if that psychoeducation had an impact on the development of the session. Consequently, these results might be reflecting a methodological difference in the instruments rather than a contradiction.

In addition, the between-group analysis showed that there was a significant difference in the amount of CB techniques that therapists used in Factor 2 and Factor 4. Thus, although the CB scores might have been raised because of the therapists' levels of activity (their active confrontation), this did not imply that these STPP therapists were using the same amount of CB techniques as the therapists in Factor 2. Again, regarding the APQ's ability to capture and differentiate the therapists' techniques, these results might be indicating that the APQ is not only able to differentiate between CBT and STPP techniques in the larger groups of sessions, but also more subtle and complex variations of therapists' techniques. The APQ's convergent and discriminant validity were further examined by correlating the factor loadings of the sessions that loaded significantly on any of the resulting factors ($n = 60$) with the sessions' mean scores in another well-validated measures (CPPS, examining the convergent validity of the APQ as a measure of therapist technique).

Results supported the APQ's convergent and discriminant validity as the factors in which STPP sessions had the higher loadings (Factor 1 and Factor 4) had a significant and positive correlation with the PI CPPS scale, and a significant and negative correlation with the CB CPPS scale. In the same line, Factor 2 (in which CBT sessions had the highest factor loadings), had a significant positive correlation with the CB CPPS scale and a significant negative correlation with the PI CPPS scale. This indicates that in the group of sessions where

therapists employed PI techniques and less frequently CB techniques (Factor 1 and Factor 4), the APQ also presented a configuration of items in which STPP techniques were more characteristic and CBT techniques less characteristic; whilst the opposite was true for Factor 2.

Further evidence of the APQ's convergent and discriminant validity was provided by the correlations with Factor 3. Neither the PI nor the CB CPPS scales correlated significantly with the factor loadings of this factor, and CPPS descriptive statistics showed that scores in both scales were very low, implying that therapists did not frequently employ either set of techniques during these sessions.

In summary, results provided support to the APQ's good levels of interrater reliability, showed convergent and discriminant validity with a well-validated instrument that measures and differentiates therapists' techniques in psychodynamic and cognitive-behavioural therapies (the CPPS).

Limitations

Despite the promising results of these studies, there were several limitations that need to be mentioned. Firstly, the APQ is composed of many constructs that are roughly grouped in three categories, and the convergent and discriminant validity of the APQ was only examined regarding the therapists' techniques. Furthermore, the APQ includes items that may not be as relevant for the techniques used in this sample, but could potentially be relevant for other therapeutic approaches (e.g. Interpersonal Therapy for Adolescents, IPT-A; Mufson et al., 2004). Thus, as the APQ is a complex instrument with many different interrelated constructs, not all of them could be validated in these studies.

In addition, although measures were taken to attain raters' blindness to the sessions' therapeutic approach, true blindness was not possible to achieve as most of the sessions of the two therapeutic approaches presented the distinctive features of their respective manuals that made them easy to recognize even within the first few minutes of the session. For example, most of CBT sessions started with

a therapist establishing an agenda and most of the STPP sessions started with a relatively silent therapist who allowed the young person to take the lead of the session. Unfortunately, this might be an inherent bias both in the ratings of these studies and in the general approach because no more measures could have been taken to ensure blindness. However, this limitation is not unique to the APQ or to the other psychotherapy Q-sets, and there is no reason to believe that the study of the psychotherapy process with the APQ is more biased than with other instruments that are coded after listening to the whole session.

A further limitation is that these studies were carried out using audio-tapes of only two therapeutic approaches (CBT and STPP) involving adolescents all of whom had been diagnosed with Major Depressive Disorder. Hence, future research will be needed to test whether results are generalizable to other therapeutic approaches and/or young people with other diagnoses.

Unfortunately, IMPACT outcome data was not available to be analysed and, hence, questions such as which Q-factors were associated with better outcome could not be explored. Future research will need to continue this task as the link between process and outcome is process research's ultimate aim.

Final remarks

Although the APQ training and rating process are time consuming, the APQ presents many advantages. Its main contribution is that it provides a language and a rating procedure for describing entire sessions of an *adolescent* in clinically relevant terms that is suitable for quantitative analysis. Other advantages of the APQ are shared with the PQS and CPQ. The analysis of the entire hour has the advantage of allowing the raters to assess the gradual unfolding meaning of events (Jones, Ghannam, Nigg, & Dyer, 1993). Also, like the PQS and CPQ, it can be used in different forms of treatments, including those like psychodynamic psychotherapy that have resisted empirical investigation due to their complexity (Bambery et al., 2007). Another advantage is that the APQ is applicable to both

nomothetic research designs (where groups of sessions are compared) and idiographic research designs (where one case is studied) (Jones, Hall, & Parke, 1991). Its fixed distribution reduces the risk of having halo effect, as ensures multiple discriminations among items (Jones, Krupnick, & Krieg, 1987). Last but not least, unlike other existing measures the APQ is multidimensional, which means that measures a variety of constructs such as therapeutic alliance, therapist's techniques and young person's feelings.

In the study presented, the APQ demonstrated that it could be used to make comparisons between and within treatments. It has also been shown that the APQ can be useful for distinguishing process variables present in the psychotherapy sessions and, by linking those with outcome, it will make it possible to identify which elements are most responsible for the success or failure of the therapies studied. Thus, the APQ has great potential to contribute to current debates in psychotherapy research, and to fill a crucial gap in the study of the psychotherapeutic process with adolescents.

References

- Ablon, J. S., & Jones, E. E. (1998). How expert clinicians' prototypes of an ideal treatment correlate with outcome in psychodynamic and cognitive-behavioral therapy. *Psychotherapy Research, 8*(1), 71-83. doi: 10.1080/10503309812331332207
- Ablon, J. S., Levy, R. A., & Smith-Hansen, L. (2011). The contributions of the Psychotherapy Process Q-set to psychotherapy research. *Research in Psychotherapy, 14*(1), 14-48.
- Bamberg, M., Porcerelli, J. H., & Ablon, J. S. (2007). Measuring psychotherapy process with the Adolescent Psychotherapy Q-set (APQ): Development and applications for training. *Psychotherapy: Theory, Research, Practice, Training, 44*(4), 405-422. doi: 10.1037/0033-3204.44.4.405
- Beutler, L. E., Moleiro, C., Malik, M., Harwood, T. M., Romanelli, R., Gallagher-Thompson, D., & Thompson, L. (2003). A comparison of the dodo, EST, and ATI factors among comorbid stimulant-dependent, depressed patients. *Clinical Psychology and Psychotherapy, 10*(2), 69-85.
- Blagys, M. D., & Hilsenroth, M. J. (2000). Distinctive features of short-term psychodynamic-interpersonal psychotherapy: A review of the comparative psychotherapy process literature. *Clinical psychology: Science and practice, 7*(2), 167-188.
- Blagys, M. D., & Hilsenroth, M. J. (2002). Distinctive activities of cognitive-behavioral therapy: A review of the comparative psychotherapy process literature. *Clinical Psychology Review, 22*(5), 671-706.
- Bychkova, T., Hillman, S., Midgley, N., & Schneider, C. (2011). The psychotherapy process with adolescents: A first pilot study and preliminary comparisons between different therapeutic modalities using the Adolescent Psychotherapy Q-set. *Journal of Child Psychotherapy, 37*(3), 327-348.
- Cregeen, S., Hughes, C., Midgley, N., Rhode, M., & Rustin, M. (In press). Short-term psychoanalytic psychotherapy for adolescents with depression: A treatment manual. In J. Catty (Ed.). London: Karnac.
- Estrada, A. U., & Russell, R. (1999). The development of the Child Psychotherapy Process Scales (CPPS). *Psychotherapy Research, 9*(2), 154-166.
- Field, A. P. (2009). *Discovering statistics using SPSS (and sex and drugs and rock 'n' roll)* (3rd ed.). Los Angeles, London, New Delhi, Singapore, Washington DC: SAGE Publication.
- Fleiss, J. L. (1981). The measurement of interrater agreement. *Statistical methods for rates and proportions* (2nd ed., pp. 212-236). New York: John Wiley & Sons.
- Goldman, G. A., & Gregory, R. (2009). Preliminary relationships between adherence and outcome in dynamic deconstructive psychotherapy. *Psychotherapy: theory, research and practice, 46*(4), 480-485.
- Goldman, R. E., Hilsenroth, M., Owen, J. J., & Gold, J. R. (2013). Psychotherapy integration and alliance: Use of cognitive-behavioral techniques within a short-term psychodynamic treatment model. *Journal of psychotherapy integration, 23*(4), 373-385.
- Goodman, G. (2010). *Transforming the internal world and attachment: Clinical applications* (Vol. 2). Lanham, MD: Jason Aronson.
- Goodman, G. (2015). Interaction structures between a child and two therapists in the psychodynamic treatment of a child with borderline personality disorder. *Journal of Child Psychotherapy, 41*(2), 141-161.
- Goodman, G., & Athey-Lloyd, L. (2011). Interaction structures between a child and two therapists in the psychodynamic treatment of a child with Asperger's disorder. *Journal of Child Psychotherapy, 37*(3), 311-326.
- Goodman, G., Midgley, N., & Schneider, C. (2015). Expert clinicians' prototypes of an ideal child treatment in psychodynamic and cognitive-behavioral therapy: Is mentalization seen as a common

- process factor? *Psychotherapy Research*. doi:
<http://dx.doi.org/10.1080/10503307.2015.1049672>
- Goodyer, I. M., Tsancheva, S., Byford, S., Dubicka, B., Hill, J., Kelvin, R., . . . Fonagy, P. (2011). Improving Mood with Psychoanalytic and Cognitive Therapies (IMPACT): A pragmatic effectiveness superiority trial to investigate whether specialised psychological treatment reduces the risk for relapse in adolescents with moderate to severe unipolar depression. Study protocol for a randomised controlled trial. *Trials*, *12*, 175-186. doi: 10.1186/1745-6215-12-175
- Hilsenroth, M. J. (2007). A programmatic study of short-term psychodynamic psychotherapy: Assessment, process, outcome, and training. *Psychotherapy Research*, *17*(1), 31-45.
- Hilsenroth, M. J., Blagys, M. D., Ackerman, S. J., Bonge, D. R., & Blais, M. A. (2005). Measuring psychodynamic, interpersonal and cognitive-behavioral techniques: Development of the Comparative Psychotherapy Process Scale. *Psychotherapy: Theory, Research, Practice, Training*, *42*(3), 340-356. doi: 10.1037/0033-3204.42.3.340
- Jacobson, C. M., & Mufson, L. (2010). Treating adolescent depression using interpersonal psychotherapy. In J. R. Weisz & A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (2nd ed., pp. 140-155). New York: Guilford Press.
- Jones, E. E. (1985). *Manual for the Psychotherapy Process Q-set*. Unpublished manuscript. University of California, Berkeley.
- Jones, E. E., Cumming, J. D., & Horowitz, M. J. (1988). Another look at the nonspecific hypothesis of therapeutic effectiveness. *Journal of Consulting and Clinical Psychology*, *56*(1), 48-55.
- Jones, E. E., Ghannam, J., Nigg, J. T., & Dyer, J. F. (1993). A paradigm for single-case research: The time series study of a long-term psychotherapy for depression. *Journal of Consulting and Clinical Psychology*, *61*(3), 381-394.
- Jones, E. E., Hall, S., & Parke, L. (1991). The process of change: The Berkeley psychotherapy research group. In L. E. Beutler & M. Crago (Eds.), *Psychotherapy research: An international review of programmatic studies* (pp. 98-106).
- Jones, E. E., Krupnick, J. H., & Krieg, P. K. (1987). Some gender effects in a brief psychotherapy. *Psychotherapy*, *24*, 336-352.
- Jones, E. E., Parke, L., & Pulos, S. M. (1992). How therapy is conducted in the private consulting room: A multidimensional description of brief psychodynamic treatments. *Psychotherapy Research*, *2*(1), 16-30. doi: 10.1080/10503309212331333568
- Jones, E. E., & Pulos, S. M. (1993). Comparing the process in psychodynamic and cognitive-behavioral therapies. *Journal of Consulting and Clinical Psychology*, *61*(2), 306-316.
- Kazdin, A. E. (2000). *Psychotherapy for children and adolescents: Directions for research and practice*. New York: Oxford University Press.
- Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, *3*, 1-27. doi: 10.1146/annurev.clinpsy.3.022806.091432
- Kernberg, P. F., Chazan, S. E., & Normandin, L. (1998). The children's play therapy instrument (CPTI): description, development, and reliability studies. *The Journal of psychotherapy practice and research*, *7*(3), 196-207.
- Kline, P. (1994). *An easy guide to factor analysis*. London, New York: Routledge.
- Kronmüller, K., Stefini, A., Geiser-Elze, A., Horn, H., Hartmann, M., & Winkelmann, K. (2010). The Heidelberg study of psychodynamic psychotherapy for children and adolescents. In J. Tsiantis & J. Trowell (Eds.), *Assessing change in psychoanalytic psychotherapy of children and adolescents: Today's challenge* (pp. 115-138). London: Karnac for the European Federation for Psychoanalytic Psychotherapy.
- Lahey, M. A., Downey, R. G., & Saal, F. E. (1983). Intra-class correlations: There's more there than meets the eye. *Psychological Bulletin*, *93*(3), 586-595.
- McLeod, B. D., & Weisz, J. R. (2005). The therapy process observational coding system-alliance scale: Measure characteristics and prediction of outcome in usual clinical

- practice. *Journal of Consulting and Clinical Psychology*, 73(2), 323-333.
- Mufson, L., Dorta, K., Wickramaratne, P., Nomura, Y., Olfson, M., & Weissman, M. (2004). A randomized effectiveness trial of interpersonal psychotherapy for depressed adolescents. *Archives of general psychiatry*, 61, 577-584.
- Mundfrom, D. J., Shaw, D. G., & Tian, L. K. (2005). Minimum sample size recommendations for conducting factor analysis. *International Journal of Testing*, 5(2), 159-168.
- Owen, J., Hilsenroth, M. J., & Rodolfa, E. (2013). Interaction among alliance, psychodynamic-interpersonal and cognitive-behavioural techniques in the prediction of post-session change. *Clinical Psychology and Psychotherapy*, 20, 513-522.
- Price, P. B., & Jones, E. E. (1998). Examining the alliance using the Psychotherapy Process Q-Set. *Psychotherapy: Theory, Research, Practice, Training*, 35(3), 392-404.
- Schmolck, P. (2002). PQMethod (version 2.33). Retrieved 23 December, 2012, from <http://schmolck.userweb.mwn.de/qmethod/index.htm>
- Schneider, C., & Jones, E. E. (2004). *Child Psychotherapy Q-Set coding manual*. Unpublished manuscript. University of California, Berkeley.
- Schneider, C., Pruetzel-Thomas, A., & Midgley, N. (2009). Discovering new ways of seeing and speaking about psychotherapy process: The Child Psychotherapy Q-Set. . In N. Midgley, J. Anderson, E. Grainger, T. Nesic-Vuckovic & C. Urwin (Eds.), *Child psychotherapy and research: New approaches, emerging findings* (pp. 72-84). New York; London: Routledge.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428.
- Smith-Hansen, L., Levy, R., Seybert, C., Erhardt, I., & Ablon, S. J. (2012). The contributions of the Psychotherapy Process Q-set to psychotherapy research. *Psychodynamic psychotherapy research: Evidence-based practice and practice-based evidence* (pp. 381-400). New York, Doedrecht, Heidelberg, London: Springer.
- Stainton Rogers, R. (1995). Q methodology. In J. A. Smith, R. Harré & L. Van Langenhove (Eds.), *Rethinking methods in psychology* (pp. 178-192). London: SAGE.
- Stein, M., Pesale, F., Slavin, J., & Hilsenroth, M. (2010). A training outline for conducting psychotherapy process ratings: An example using therapist technique. *Counselling and psychotherapy research*, 10(1), 50-59.
- Sub-group, I. S. C. (2010). *Cognitive-behaviour therapy for depression in young people: Manual for therapists*. Unpublished manual.
- Tolan, P. H., & Titus, J. (2011). Adolescents. In J. C. Norcross, G. R. VandenBos & D. K. Freedheim (Eds.), *History of psychotherapy: Continuity and change* (2nd ed., pp. 449-457). Washington, DC: American Psychological Association.
- Velicer, W., & Fava, J. (1998). Affects of variable and subject sampling on factor pattern recovery. *Psychological methods*, 3(2), 231-251.
- Verduyn, C., Rogers, J., & Wood, A. (2009). Depression in adolescence. *Depression: cognitive behaviour therapy with children and young people* (pp. 1-12). London: Routledge.
- Watts, S., & Stenner, P. (2012). *Doing Q methodological research: Theory, method and interpretation*. Los Angeles, London, New Delhi, Singapore, Washington DC: SAGE Publications.
- Weersing, V. R., Weisz, J. R., & Donenberg, G. R. (2002). Development of the Therapy Procedures Checklist: A therapist-report measure of technique use in child and adolescent treatment. *Journal of clinical child psychology*, 31(2), 168-180.

