The Development and Psychometric Evaluation of the Group Schema Therapy Rating Scale – Revised

Article *in* Behavioural and Cognitive Psychotherapy · January 2018
DOI: 10.1017/S1352465817000741

8 authors, including:

Simone Verhagen
Maastricht University
11 PUBLICATIONS **19** CITATIONS
SEE PROFILE

Gerhard Zarbock
28 PUBLICATIONS **68** CITATIONS
SEE PROFILE

Joan M Farrell
Indiana University-Purdue University Indianapolis
45 PUBLICATIONS **1,135** CITATIONS
SEE PROFILE

Odette M. Brand-de Wilde
Odette Brand-de Wilde, Klinisch Psycholoog
23 PUBLICATIONS **292** CITATIONS
SEE PROFILE

Behavioural and Cognitive Psychotherapy: page 1 of 18
doi:10.1017/S1352465817000741

The Development and Psychometric Evaluation of the Group Schema Therapy Rating Scale – Revised

Emily Bastick

Murdoch University, Murdoch, Western Australia

Suili Bot

Maastricht University, Maastricht, The Netherlands

Simone J. W. Verhagen

Maastricht University, Maastricht, The Netherlands

Gerhard Zarbock

IVAH – Institut für Verhaltenstherapie-Ausbildung, Hamburg, Germany

Joan Farrell

Indiana University-Purdue University, Indianapolis, Indiana, USA, and Centre for Borderline Personality Disorder Treatment and Research, Indianapolis, Indiana, USA

Odette Brand-de Wilde

University of Amsterdam, Amsterdam, The Netherlands, and De Viersprong, Halsteren, The Netherlands

Arnoud Arntz

University of Amsterdam, Amsterdam, The Netherlands

Christopher William Lee

University of Western Australia, Crawley, Western Australia

Reprint requests to Dr Christopher Lee, Faculty of Health and Medical Sciences, University of Western Australia. E-mail: Chris.Lee@uwa.edu.au

Background: Recent research has supported the efficacy of schema therapy as a treatment for personality disorders. A group format has been developed (group schema therapy; GST), which has been suggested to improve both the clinical and cost-effectiveness of the treatment. **Aims:** Efficacy studies of GST need to assess treatment fidelity. The aims of the present study were to improve, describe and evaluate a fidelity measure for GST, the Group Schema Therapy Rating Scale – Revised (GSTRS-R). **Method:** Following a pilot study on an initial version of the scale (GSTRS), items were revised and guidelines were modified in order to improve the reliability of the scale. Students highly experienced with the scale rated recorded GST therapy sessions using the GSTRS-R in addition to a group cohesion measure, the Harvard Community Health Plan Group Cohesiveness Scale – II (GCS-II). The scores were used to assess internal consistency and inter-rater reliability. Discriminant validity was assessed by comparing the scores on the GSTRS-R with the GCS-II. **Results:** The GSTRS-R displayed substantial internal consistency and inter-rater reliability, and adequate discriminate validity, evidenced by a weak positive correlation with the GCS-II. **Conclusions:** Overall, the GSTRS-R is a reliable tool that may be useful for evaluating therapist fidelity to GST model, and assisting GST training and supervision. Initial validity was supported by a weak association with GCS-II, indicating that although associated with cohesiveness, the instrument also assesses factors specific to GST. Limitations are discussed.

Keywords: group schema therapy, treatment adherence, treatment competence, fidelity, reliability, psychometric

Introduction

Group schema therapy for borderline personality disorder

Research on schema therapy delivered in individual sessions has demonstrated that it is a promising treatment for personality disorders (Bamelis et al., 2014; Giesen-Bloo et al., 2006). Recently, the model has been adapted for use in a group format: group schema therapy (GST; Farrell et al., 2009). GST incorporates psycho-education about schema theory and borderline personality disorder (BPD), skills training for emotional awareness and distress tolerance, and experiential techniques. Initial studies of GST for BPD have shown improvement on both symptom levels, reduced drop-out rate in comparison with transference-focused psychotherapy, treatment as usual, and clarification-oriented psychotherapy, and the possibility of complete remission.

In an initial evaluation of this approach (Farrell et al., 2009), 32 women with BPD who were receiving individual therapy were randomized to receive either an extra 8 months of GST, or no additional treatment. Patients who received GST showed improvements in all BPD symptom domains and at the end of treatment, 94% of the patients no longer met diagnostic criteria for BPD (compared with 16% in the control group). A subsequent study by Dickhaut and Arntz (2014) examined a combination of weekly GST and individual schema therapy for 2 years, with an additional 6 months of individual schema therapy if indicated. The results demonstrated that the combination of GST and individual therapy resulted in significant reductions in symptoms associated with BPD manifestations and dysfunctional schemas. The effect sizes of the improvement were large. Reiss et al. (2014) also took the first steps to test GST in naturalistic clinical settings in a series of pilot studies into intensive in-patient schema therapy treatment. In their studies, they combined both individual and group schema therapy modalities. The results validated that in-patient schema therapy can significantly reduce

symptoms of severe BPD and global severity of psychopathology. A number of more recent studies have also demonstrated significant symptom reduction following GST whether in a short-term GST in-patient programme (Nenadić et al., 2017), in an out-patient programme lasting a year (Fassbinder et al., 2016), or in an out-patient programme of just 20 sessions (Skewes et al., 2015). However, in all three of these studies, there was no allocation to a no-treatment control or other treatment comparison condition.

As demonstrated above, the GST format appears promising. It also has several advantages over the individual schema therapy format. First, group therapy is more cost-effective than individual therapy and mental health services (both private and publicly funded) would benefit from interventions that are less costly. Private patients have less out-of-pocket expenses as a group session would cost less per person than an individual session, and public-funded services can treat more patients or treat existing patients for longer if the cost per patient is reduced. Secondly, GST therapy allows for social/psychological factors that are not present in individual therapy. More specifically, Yalom (1995) empirically identified eleven factors that open the pathway to therapeutic change, including universality, group cohesiveness, and corrective recapitulation of the primary family group. In this light, Farrell and Shaw (2012) noted that limited re-parenting is enhanced in GST as added to the transitional parental figure (the therapist) is a transitional family (the group), a combination which seems to amplify the effects that each has on their own. Common to reports from patients who attend group therapy of varying modalities, Farrell and colleagues (2009) noted that their participants reported that the group was 'the first time [they] felt a sense of belonging and acceptance' and that they were 'not alone' and 'not crazy' (p. 10).

An important research question is the extent to which these group therapy factors are enhanced by the specific aspects of the schema therapy model, or the extent to which the schema model accounts for change beyond these common factors. In order to assess the relative contribution of the specific aspects of schema therapy, it is important to reliably assess what constitutes schema therapy and whether therapists are adhering to the model.

Aims and hypotheses

The aim of the present study was to construct and assess the reliability and validity of a rating scale of fidelity for group schema therapy. It was hypothesized that the scale would display substantial internal consistency as demonstrated by Cronbach's alpha coefficients of .61 or above (Landis and Koch, 1977) and substantial inter-rater reliability when used by trained raters. Discriminant validity would be assessed by comparing the scale with a group cohesion tool, the Harvard Community Health Plan Group Cohesiveness Scale-II (GCS-II; Soldz et al., 1987). As group cohesion is only one component of GST, and only partially accounts for the effects of specific factors in therapy such as the use of experiential techniques, it is expected that the competence ratings of the GST fidelity scale would have a low correlation with the GCS-II. However, the size of the correlation would be in the low range (between 0.21 and 0.60).

Method

Participants

For the present study, GST session recordings were taken from an ongoing international multi-centre randomised controlled trial (RCT; Wetzelaer et al., 2014) investigating the clinical

effectiveness of GST for BPD. Participants involved in the trial were aged between 18 and 65 years and had a primary diagnosis of BPD. For further details of the inclusion and exclusion criteria, see Wetzelaer and colleagues (2014). All participants and therapists, directly or indirectly involved in the study, had given informed consent to participate, and for the therapy sessions to be recorded and used for the purposes of this study.

Each group was run by two schema therapists, at least one of which was a senior (or advanced) schema therapist. These therapists received 6 days of training in GST. During the broader trial, intensive supervision sessions were held once or twice a year. In addition, weekly peer supervision was provided locally and central supervision by the developers of the GST model was provided through teleconferencing and viewing of video recordings weekly during the first year, biweekly for the 6 months thereafter, and monthly in the last 6 months. The final sample for the current study featured a total of four therapists and 14 patients across two groups (named after their locations, Peel and Rockingham).

Raters

All three raters involved in the pilot study (Bot, 2013) of the GSTRS were involved in the revision process. Two of the raters (S.V. and S.B.) were Dutch masters level psychology students, who had sufficient English language capacities, and the third rater (E.B.) was an Australian psychology doctoral student. The raters were not certified schema therapists; however, all raters received clinical training during their education, and were required to study the book *Group Schema Therapy for Borderline Personality Disorder* (Farrell and Shaw, 2012). In addition, the raters were required to watch the video produced by the IVAH institute of Hamburg (Zarbock et al., 2011), which has been used to train therapists in GST.

Procedure

Development of the GSTRS-R. An initial version of a fidelity measure for GST, the GSTRS, was originally developed by Zarbock et al. (2012) based on the techniques and approaches described in the GST protocol created by Farrell and Shaw (2012). The GSTRS consisted of two subscales: *General* and *Specific*. The 27 items on the *General* subscale reflected a range of behaviours present throughout each session, such as 'therapist has a positive presence' and were divided into six sections: *General Therapist Behaviour*, *Limited Reparenting*, *Group Therapeutic Skills*, *Group Climate*, *Structure* and *Co-Therapist Team*. In contrast, the 27 items on the *Specific* subscale reflected the use of particular techniques or interventions, present at a particular point in time, such as a 'physical grounding exercise' and were divided into four sections: *Mode Awareness and Change Work*, *Cognitive Interventions*, *Experiential Interventions* and *Behavioural Pattern Breaking Interventions*. Each item was initially rated for treatment integrity (or adherence), which refers to the degree to which techniques and methods as described in the treatment protocol are implemented. If implemented, the item was then rated for competence, which refers to the skill or quality with which the intervention was delivered.

In a pilot test of the ratings, the inter-rater reliability of adherence was poor on both *General* and *Specific* subscales with non-significant average measures ICCs of –.11 and –.03, respectively (Bot et al., 2013). In order to improve reliability, four aspects of the pilot scale were changed. These were changes in item content, improvements in when the therapists

were rated, the addition of descriptors to help with assessing adherence to the specific subscale items, and the addition of descriptors to help with assessing competency ratings.

To improve item content, the three raters highlighted items on the *General* subscale of the GSTRS that appeared unclear to them, lacked precision or were too similar to other items. In collaboration with the raters, a group of international researchers and practitioners met regularly via audio conferences. The researchers and practitioners were all certified schema therapists involved in the ongoing multi-site RCT of GST for BPD. This revision process was repeated several times until the three raters were confident that they were interpreting the items and their corresponding guidelines in a similar manner. Table 1 shows changes made to specific items on the *General* subscale of the GSTRS. Following the changes in item content, the structure of the subsections was altered. The original six *General* subsections were reduced to five, eliminating the subsection of *Group Therapeutic Skills.* There was also the addition of a subsection to the *Specific* subscale: *Anticipatory Socialization to the Group Modality* and *Schema Therapy Education*.

The second aspect of the GSTRS to be changed was the instructions to rate Therapist A in the first 30 minutes of the recording, and Therapist B in the second 30 minutes of the recording. This rule was abolished and instead, in all tapes both therapists were rated simultaneously throughout the entire recording.

Thirdly, in order to improve rater consistency, guidelines were created for rating the subscales of the GSTRS-R. These guidelines included specific examples rather than relying on raters' interpretations based on information contained in Farrell and Shaw (2012). Figure 1 shows an example of a guideline developed for one of the items in the *Specific* subscale. The final guidelines for the GSTRS-R *General* subscale included descriptors of what would warrant a score of non-adherence or adherence. Figure 2 shows an example of a guideline from the *General* subscale.

The final change to the pilot version of the scale was to change the competency rating system. In the pilot version, competency ratings occurred over a 6-point scale. To improve scoring discrimination this 6-point scale was changed to a 7-point scale corresponding to descriptors such as: 0 = *very poor*, 1 = *poor*, 2 = *unsatisfactory*, 3 = *adequate*, 4 = *good*, 5 = *very good* and 6 = *excellent*. In addition, examples were provided of good or poor competence to improve reliability. The most recent version of the GSTRS-R can be accessed at: https://doi.org/10.4225/23/585a265e14ab8 (Zarbock et al., 2014).

Tape selection. The final analysis of the GSTRS-R utilized 10 randomly selected video-recorded therapy sessions from two Australian mental healthcare sites that were not used during any of the previous review processes. In order to select the GST recordings, recordings from each site were numbered in chronological order. A random number generator was used to determine a recording number for each stage of therapy (one recording for months 1–4, 5–9, 10–14, 15–19 and 20–24). Inclusion criteria for the tapes were as follows:

(1) Therapy sessions were complete (no more than 20 minutes missing, nor the beginning or end missing).
(2) There were three or more participants, plus two (of the original) therapists present in the group.
(3) The video recording was of sufficient quality that all group members were clearly audible.

Table 1. Revision of GSTRS *General* subscale items

	Original GSTRS item	Problem with the item		Revised GSTRS-R item
04	Therapist acts as a role model for healthy adult behaviour (e.g. by accepting own imperfections, apologizing for mistakes, acknowledging and taking care of one's own needs)	It may be possible that the therapists were not making any mistakes and it was unclear how to recognize that a therapist has accepted his/her own imperfections. Too similar to item 6	04	Therapist addresses and resolves alliance ruptures using schema therapy technique and terminology (e.g. apologizing for not seeing the VC, pointing out one's demanding parent, explaining the idea behind an intervention, apologizing for having been too quick, impatient or for not providing enough details before an intervention, or for overlooking an important patient response)
05	Therapist combines both rational and emotional behaviours (e.g. experiential activation, cognitive reflection)	Required more clarification and an example	05	Therapist addresses both cognitive and emotional processes of the patient in an integrated manner. Integration means that both processes are included for the same issue or are present in the same intervention
06	Therapist self-discloses in an appropriate manner (e.g. to reduce participants' shame, to show that nobody is perfect, to model being aware of and open about one's own schema driven reactions and therefore more in control and able to cope)	It was not clear what self-disclosure specifically referred to. More guidelines were needed. Raters had differing views on the definition of self-disclosure	06	Therapist self-discloses in an appropriate manner that serves the therapy process (e.g. to reduce participants' shame, to show that nobody is perfect, to model how being aware of and open about one's own schema-driven reactions leads to more control and more effective coping)
08	Therapist attends to the modes of participants (e.g. validating/protecting VC, limiting IC, allows venting AC's anger, reinforcing HA and HC, disempowering DemP/PP, addressing DP).	More detail and clarity was required	08	Therapist attends to the need that is present for a patient based upon the modes he or she is in (e.g. validating/protecting VC, limiting IC, allowing AC's to vent anger, reinforcing HA and HC, disempowering DemP/PP, addressing DP)

Table 1. Continued

	Original GSTRS item	Problem with the item		Revised GSTRS-R item
09	Therapist establishes and maintains boundaries of group interactions (i.e. time and task management, reminding of ground rules)	Item was too similar to items 10 and 18 and required clarification. It was unclear whether maintaining boundaries needed to be explicitly stated or could be implicit within the session	09	Limit setting: therapist limits dysfunctional and disruptive behaviour (e.g. violation of ground rules, verbal attack of another member) immediately, firmly and directly. Empathic confrontation is a different intervention (Note: maintaining boundaries was relocated to item 20)
10	Therapist deals with dysfunctional behaviour (e.g. violation of ground rules, verbal attack of another member) and confronts participants in an empathic way (e.g. by validating underlying needs)	Items 9, 10 and 18 were too similar and needed clarification	10	Empathic confrontation: therapist confrontation is done in firm but friendly manner; the patient's underlying need is addressed. (a) Name the problem behaviour, (b) strengthen connection, (c) connect to history or underlying feelings, (d) point out the result of the action, (e) discuss more effective options to get the patient's needs met
11	Therapist responds to the overall group needs and atmosphere of the group (e.g. addresses low energy level, responds to level of vulnerability)	Items 11, 12 and 16 are combined to form a new item	11	Therapist uses schema and mode language to label, identify, comment on, or regulate participant's experiences and behaviour (in session or in a reported event outside of group). If appropriate, the underlying needs of the individual and the group as a whole are also labelled and addressed, including individual modes (e.g. 'Tough Tammy, Mean Mommy')
12	Therapist openly labels currently activated modes and identifies the needs of individuals and the group as a whole			

Table 1. Continued

	Original GSTRS item	Problem with the item		Revised GSTRS-R item
14	Therapist is an active leader, who allows enough room (silence) for participants' involvement, but not so much that anxiety builds up	More detail added	13	Therapist is an active leader, who allows enough room (silence) for participants' involvement, but not so much that anxiety builds up. Therapist uses direction and limit setting actively to keep the group in the 'working window' of activation, preventing over activation (high tension, turmoil, verbal attacks) as well as under-activation (e.g. detachment, lethargy)
16	Therapist uses schema and mode language to identify, comment on, or regulate participant's experiences and behaviour	Items 11, 12 and 16 are combined to form a new item	11	See item 11 above.
18	Therapist addresses and manages conflicts that occur (between members, between co-therapists and between members and co-therapists)	Items 9, 10 and 18 were too similar and needed clarification	16	Therapist addresses and manages interpersonal tensions, irritations, quarrels and/or open conflicts that occur (between members, between co-therapists and between members and co-therapists) according to the stage of the group. Healthy conflict can occur without breaking ground rules
20	Therapist creates an atmosphere that encourages and engages the playful child mode of each member and the group as a whole	Needed clarification. It was unclear whether the playful child mode could be facilitated with non-verbal behaviour and attitudes (e.g. implicit behaviours)	18	Therapist creates an atmosphere that encourages and engages the playful child mode of each member and the group as a whole. This could be done explicitly by using an exercise or task, implicitly by smiling, laughing, non-verbal teasing, or by para-lingual tone
21	Therapist fosters group cohesion and acceptance (e.g. by pointing out similarities and supporting the acceptance of differences)	More detail added	19	Therapist fosters group cohesion and acceptance (e.g. by pointing out similarities among group members while also supporting the acceptance of differences; limiting any negative evaluations of other members, encouraging 'I feel' language instead of judgements of the other)

Table 1. Continued

	Original GSTRS item	Problem with the item		Revised GSTRS-R item
22	Therapist negotiates session agenda and topics, or issues that need to be addressed with the group collaboratively	It was unclear about whether the agenda/plan needed to be set explicitly	20	Therapist establishes and maintains the working frame of a group by time and task management and reminders of ground rules. High level of competence is defined by: the balance of structure and flexibility, and the therapist setting the stage for the task or topic and guiding the group actively toward a goal while also adjusting to the group needs
23	Therapist balances structure and flexibility (i.e. going in with a goal but being ready to change the plan as the group necessitates)			
26	Therapist explains the (Schema Therapy) rationale behind techniques and approaches to provide transparency	More detail was required	23	Therapist explains the (Schema Therapy) rationale behind techniques and approaches to provide transparency. The point at which this is done may vary – sometimes before and sometimes coming after (e.g. following experiential exercises). Therapist chooses the most suitable point in time for this explanation, so that the emotional process is facilitated (and not disturbed or closed down etc.)
27	Therapist gives homework and either refers to it, works with it or collects it to provide comments during the group session	It was unclear what was considered homework and whether it was required every session	24	Therapist gives some assignment or task (could be a question to consider further). These assignments must be followed up on in some way in the next session – either used in the session or collected for therapist review and returned with comments

VC, vulnerable child; IC, impulsive child; AC, angry child; HA, healthy adult; HC, happy child; DemP, demanding parent; PP, punitive parent; DP, detached protector.

15) Imagery change work(Rescripting): (pp. 169, 115, 170-74, 178-95)

The purpose of this exercise is to change the endings of childhood memories.'Patients access their vulnerable child mode and go in imagery to a childhood scene where a core need was not met. They allow the therapist, and later the group also, to enter the image to meet the need. This usually takes the form of protecting the child, comforting him/her, telling the punitive parent mode to stop, that they were wrong to treat him/her like that, and then taking the child in imagery to a safe place. The group is brought into imagery work in a variety of ways and at different depths of involvement. '

Figure 1. Example of a guideline from the *Specific* subscale.

Item	Exemplary descriptions	
	0	6
	GENERAL THERAPIST BEHAVIOR	
1.	**Insecure/anxious**: Therapist avoids eye contact, does not speak up (hard to understand, mumbles), insecure/ intimidated body language (e.g., crossed arms, slumped/ducked posture, wringing his/her hands) or constantly takes notes or reads off his/her notes or lacks distinctly positive mimic and gestures (physical expression). **Cold/negative/reserved**: Therapist is stern, cool and rigid. Openness and friendliness seem posed, **artificial** and forced. Therapist is appearing to be **overconfident** about his/her own ability in an arrogant manner or is trying too hard to impress.	**Confident**: Therapist seeks eye contact, has a clear and easily audible voice and a confident body language, does not depend on his/her notes (hands are free), is expressive (responds physically via mimic and gestures) and emphasizes his/her own statements through body language. **Warm/positive/open**: Therapist seems kind (understanding, approving, forgiving) and encourages participants through smiling and nodding. Therapist seems **authentic**.

Figure 2. Example of *General* subscale guidelines for item 1.

If one of the randomly selected sessions did not meet the sufficient criteria, the following tape in the series was selected. This tape then underwent the same review process.

Each video file was provided to the three raters (E.B., S.V. and S.B.) who watched and rated the recordings independently. The raters concurrently rated the recordings using the GSTRS-R and the GCS-II for the purpose of determining the discriminant validity of the GSTRS-R.

Companion measure

The GCS-II (Soldz et al., 1987). The GCS-II is a group cohesion tool, and was selected to explore the discriminant validity of the GSTRS-R. Prior research has indicated that group cohesion

is positively correlated with favourable treatment outcome within group psychotherapy (see Kivlighan and Lilly, 1997; Ogrodniczuk and Piper, 2003; Soldz et al., 1990).

The GCS-II is an observer-rated group cohesion measure that consists of nine ratings on the following dimensions: focus, interest/involvement, trust, facilitative behaviour, bonding, global cohesiveness, affective intensity, conflict, and global quality. Each dimension is rated on a 1 (very slight) to 9 (very strong) scale. Although currently there does not appear to be any published evaluation of the internal consistency of inter-rater reliability of the GCS-II, it was selected due to its successful use in a range of observer-rated group cohesion studies. Raters were provided with a full copy of the scale, the GCS-II manual, and a set of blank score sheets.

Statistical analyses

Adherence and competence ratings for the *General* and *Specific* subscales of the GSTRS-R were analysed separately. The ratings of the two therapists for each tape were averaged to obtain one score per tape, per rater.

All data were analysed using SPSS (version 23). Ratings were added to the data set and were initially scrutinized for outliers. This involved manually examining all three raters' scores for each item across the 10 recordings. No scores were removed following this step.

Items 9 and 10 from the competence ratings of the GSTRS-R *General* subscale were removed prior to analyses due to a particularly low frequency of endorsement of these items [3 (1%) and 7 (2.3%), respectively]. The remaining missing items (5%) were excluded from analysis.

Internal consistency. The internal consistency of the adherence ratings was initially calculated using the Kuder–Richardson Formula 20 (the equivalent to Cronbach's alpha when computed for binary items). Items with zero variance were removed. This accounted for 20 of the 28 items in the *General* subscale, as all raters had deemed these behaviours were present, and 15 out of 28 items for the *Specific* subscale, as all raters had deemed these behaviours were not present. Due to the significant reduction in data and the few remaining number of items, it appeared meaningless to report internal consistency on the remaining items.

The internal consistency of the competence rating for the *General* subscale was computed using Cronbach's alpha coefficient (Cronbach, 1951) on the mean raters' item scores. The competence ratings for the *Specific* subscale were not analysed for internal consistency due to the low frequency with which the items were endorsed (mean number of items per tape: 1.8).

Inter-rater reliability of the GSTRS-R. Two-way mixed intraclass correlation coefficients with a 95% confidence interval (ICC; Shrout and Fleiss, 1979) were used to measure the absolute agreement between the three raters across the 10 recordings on the adherence ratings from each subscale (*General* and *Specific*), and the competence ratings for the *General* subscale. Inter-rater reliability was not calculated for the competence ratings of the *Specific* subscale due to the low frequency of endorsement of items.

Validity of the GSTRS-R. Prior to any validity analysis, the internal consistency and inter-rater reliability of the GCS-II were calculated using Cronbach's alpha and the ICC, respectively. A Pearson's bivariate correlation was used to analyse the relationship between the item scores for the competence ratings on the General subscale of the GSTRS-II and the GCS-II.

The scores from all three tests (Cronbach's alpha, ICC and Pearson's r) were interpreted as: <0, poor agreement/no linear relationship; 0–.20, slight agreement/very weak relationship;

Table 2. Item-total statistics and standard deviations (*SD*) for the competence ratings for the GSTRS-R *General* subscale

Item number	Corrected item: total correlation	*SD*
1	.39	.57
2	.08	.63
3	.78	.72
4	.35	.59
5	.43	.63
6	.27	.70
7	.34	.57
8	.56	.70
11	.26	.76
12	.71	.65
13	.36	.81
14	.84	.66
15	.13	.71
16	–.12	.77
17	.76	.78
18	.71	.64
19	.89	.73
20	.76	.72
21	.70	.79
22	.66	1.11
23	.67	.61
24	.46	.77
25	.09	.74
26	.33	.63
27	.38	.66
28	.35	1.01

.21–.40, fair agreement/weak relationship; .41–.60, moderate agreement/moderate relationship; .61–.80, substantial agreement/strong relationship; and .81–1, almost perfect to perfect agreement/relationship (Landis and Koch, 1977).

Results

Reliability

Internal consistency. Internal consistency of the competence ratings for the *General* subscale was excellent, with a Cronbach's alpha value of .90. Corrected item-total correlations ranged from –.12 to .89, as indicated in Table 2.

Internal consistency for the GCS-II was substantial with a Cronbach's alpha value of .76. Corrected item-total correlations ranged from .01 to .81, as indicated in Table 3. Removing the

Table 3. Item-total statistics and standard deviations (*SD*) for the GCS-II

Item number	Corrected item: total correlation	Cronbach's alpha if item deleted	*SD*
1	.36	.75	.90
2	.49	.74	.54
3	.66	.71	.78
4	.36	.75	.91
5	.53	.72	1.10
6	.81	.69	.76
7	.58	.72	.67
8	.51	.73	1.07
9	.01	.81	1.04

worst performing item (item 9) increased the internal consistency to a Cronbach's alpha value of .81.

Inter-rater reliability. Scores from the three raters across the 10 recordings showed that the inter-rater reliability of the adherence ratings for the *General* subscale of the GSTRS-R was almost perfect, with an average measures ICC of .91 [95% CI: .89 to .93; F (279,558) = 11.69, $p < .001$].

The inter-rater reliability for the adherence ratings for the *Specific* subscale was substantial, with an average measures ICC of .75 [95% CI: .70 to .80; F (279,558) = 4.10, $p < .001$].

Therapist competence overall was excellent. The agreement between raters was very high; 91% of the items were rated as 4, 5 or 6. Due to a lack of variance across the three raters' scores for the competence ratings of the *General* subscale, the ICC could not be calculated. If, however, the inter-rater reliability is computed according to the definition of Finn (1970, p.73), the total possible range from 0 to 6 is taken into account, resulting in an inter-rater reliability of .98.

Finally, the inter-rater reliability for the GCS-II was almost perfect, with an average measures ICC of .86 [95% CI: .80 to .91; F (89,178) = 7.84, $p < .001$]. Reviewing the scale without item 9 resulted in an average measures ICC of .82 [95% CI: .74 to .88; F (79,158 = 5.81, $p < .001$].

Discriminant validity

The competence ratings for the GSTRS-R General subscale item scores were weakly and significantly correlated to the GCS-II item scores (r = .45, p < .001).

Discussion

Largely, the stated hypotheses were confirmed. The results identified that the newly revised GSTRS-R is adequately reliable and valid for assessing therapist adherence to the GST model.

In terms of reliability, the Cronbach's alpha coefficient for the competence ratings of the *General* subscale was excellent (.90), indicating that the core subscale within the GSTRS-R has considerable internal consistency as predicted, and the items are operating as a cohesive construct measuring the same underlying elements of the intervention.

Although inter-rater reliability analyses could not be performed for the competence ratings for the *Specific* subscale, two out of the remaining three inter-rater reliability hypotheses were supported as outlined. There was close to perfect agreement between raters for both the adherence and competence ratings for the *General* subscale (.91 and .98, respectively). Inter-rater reliability for adherence ratings for the *Specific* subscale was good but lower than predicted with an ICC of .75. The validity of the *Specific* subscale is questionable.

Again, the internal consistency of the GCS-II was substantial with a Cronbach's alpha value of .76, and would improve moderately if item 9 were removed (.81). Furthermore, the inter-rater reliability of the GSC-II was excellent, with an ICC of .86. These results add to the void in the literature by providing evidence for the reliability of the GCS-II. Moreover, the results indicated that it was acceptable to use the GCS-II for the purpose of discriminant analysis in the current study.

In line with the stated hypothesis, there was a significant, positive weak correlation between the item scores of the competence ratings for the GSTRS-R *General* subscale and the GCS-II (.45). This result provides support for the notion that the GSTRS-R incorporates an element of group cohesion whilst still measuring an independent construct (i.e. fidelity to the schema model).

Limitations

Despite the generally positive findings, there are several limitations of the present study and further investigation of the psychometric properties of the GSTRS-R is recommended. First, further replications of this study should involve a larger sample of recordings from a broader range of sites and countries in order to combat any cultural effects and idiosyncratic behaviour of particular therapists, as the tapes rated only involved four therapists. More heterogeneity in the behaviours of the therapists would help to reduce the lack of variance that was observed in some items.

In addition, the therapy sessions rated in this investigation were taken from a research trial evaluating the new delivery mode of GST. Therefore, it is likely that the therapists had received more training and supervision in GST than therapists outside of the trial would. This probably skewed the range of competence to the upper end with rating primarily falling in the 4 to 6 range, with very few ratings falling below 4. The scale should be re-evaluated using recordings of therapists who are only moderately trained in GST in order to produce data that are more generalizable to the community health care setting and produce a wider range of variance in techniques and competencies.

Furthermore, when averaging the ratings of the two therapists prior to statistical analyses, the authors made the assumption that if one therapist had poor adherence, and one therapist had very high adherence, then the rating of the pair would be in between that of two non-adherent and two highly adherent therapists. Alternatively, it might be the case that a very good therapist compensates for the effects of a poorly functioning therapist, or that a poorly functioning therapist damages the effects of the good therapist. The GST model assumes good collaboration between therapists, requiring that both therapists adhere to the GST protocol. The

possibility that differences in adherence within the pair might influence the outcome deserves further study.

It must also be noted that there was a high degree of concordance between the three raters. This may reflect the fact that the three raters worked very closely on the development of the scale in addition to having a greater understanding of the items in the scale than would someone new to the scale or relatively new to GST. As the scale was developed with the intention of being rated by psychology/mental health students who are not professionally trained in GST, it is imperative that the scale be tested on a new selection of raters that are limited in their knowledge and experiences of conducting GST.

Importantly, the reliability of the competency ratings for the items in the *Specific* subscale remains unknown. The *Specific* subscale items were seldom endorsed, reflecting that each item, or schema therapy technique, is just one of a large set of possible techniques that can be utilized within the schema therapy model. The considerably low rate of endorsement was surprising in the present study, in that typically only a couple of the techniques described by items in the *Specific* subscale were utilized within any one session of GST. Therefore, in order to determine the internal consistency and the inter-rater reliability of the competency ratings of the *Specific* subscale, future studies should include a sample size four times larger than that of the present study.

Finally, there are other ways that the validity of the GSTRS-R could be assessed, including identifying the factor structure of the scale, which can reduce the number of items within the scale for ease of use. Another option is to explore its predictive validity such as comparing competence ratings of the GSTRS-R *General* subscale to BPD patients' change scores on the Borderline Personality Disorder Severity Index (BPDSI) following GST treatment. As widely noted in literature (e.g. Guydish et al., 2014; Hogue et al., 2008), better treatment fidelity probably predicts better treatment outcome. Therefore, the validity of the GSTRS-R will be further supported if the ratings are positively correlated with change scores.

Application of the GSTRS-R

There are a number of potential applications for the GSTRS-R, for clinical practice, training and research. The GSTRS-R may be used within the clinical environment to ensure that clients are receiving a high quality of therapy and could provide clinicians with information on how well they are adhering to the GST model, and highlight areas in need of improvement for future therapy sessions. Additionally, as different approaches of providing GST to patients with BPD are trialled internationally, a scale for assessing the fidelity of GST would be invaluable for reducing the variance between different therapists' adherence and competence to the GST model. Furthermore, the GSTRS-R would be a useful tool within training programmes to facilitate the supervision and progression of new GST therapists, and to assess their progress throughout their training and career. In this sense, the scale could be used as a 'checklist' to ensure that all components are being adequately adhered to.

Conclusions

Overall, the results from this study have shown that the revised version of the *General* subscale of the GSTRS is a practical, reliable and valid tool for evaluating fidelity to GST for BPD. Reflective of the success of the scales revision process, the reliability of the scale was good,

and the weak positive correlation with the GCS-II (Soldz et al., 1987) supports the discriminant validity of the scale. As a whole, the findings of this study indicate that the scale could potentially be invaluable to GST research, supervision and training practices, and provide clinicians with an important method for evaluating and improving their GST therapeutic technique.

Acknowledgements

The authors would like to thank those who contributed to the development and refinement of the GSTRS-R, including Eelco Muste and Neele Reiss. A special thank you is also extended to all patients, research assistants and therapists for their involvement in the parent study, the ongoing international multi-centre RCT (Wetzelaer et al., 2014).

Conflicts of interest: Emily Bastick, Suili Bot, Simone Verhagen, Gerhard Zarbock, Joan Farrell, Odette Brand, Arnoud Arntz and Christopher Lee have no conflicts of interest with respect to this publication.

Ethical statement: The authors abided by the Ethical Principles of Psychologists and Code of Conduct as set out by the APA. Ethical approval was approved by the Murdoch University Human Research Ethics Committee.

Financial support: This study was supported by grants from Australian Rotary Health, the Netherlands Organization for Health Research and Development (ZonMW; 80-82310-97-12142), the Netherlands Foundation for Mental Health (2008 6350), Maastricht University, The Netherlands, the Else Kröner-Fresenius-Stiftung, and IVAH – Institut für Verhaltenstherapie-Ausbildung, Hamburg, Germany. Funding bodies played no role in the design of the study, in the collection, analysis and interpretation of data, in the writing of the manuscript or in the decision to submit the manuscript for publication.

References

Bamelis, L. L. M., Evers, S. M. M. A., Spinhoven, P. and Arntz, A. (2014). Results of a multicenter randomized controlled trial of the clinical effectiveness of schema therapy for personality disorders. *American Journal of Psychiatry*, *171*, 305–322. https://doi.org/10.1176/appi.ajp.2013.12040518

Bot, S. (2013). *A pilot study: what is the inter-rater reliability of the Group Schema Therapy Rating Scale and is there a cultural difference between countries?* Bachelor thesis, MARBLE: Empirical Research. Maastricht University.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334. https://doi.org/10.1007/BF02310555

Dickhaut, V. and Arntz, A. (2014). Combined group and individual schema therapy for borderline personality disorder: a pilot study. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*, 242–251. https://doi-org/10.1016/j.jbtep.2013.11.004

Fassbinder, E., Schuetze, M., Kranich, A., Sipos, V., Hohagen, F., Shaw, I. et al. (2016). Feasibility of group schema therapy for outpatients with severe borderline personality disorder in Germany: a pilot study with three year follow-up. *Frontiers in Psychology*, *7* (1851). doi: 10.3389/fpsyg.2016.01851

Farrell, J. M. and Shaw, I. A. (2012). *Group Schema Therapy for Borderline Personality Disorder: a Step-by-Step Treatment Manual with Patient Workbook*. Chichester, West Sussex, UK: John Wiley and Sons Ltd.

Farrell, J. M., Shaw, I. A. and Webber, M. A. (2009). A schema-focused approach to group psychotherapy for outpatients with borderline personality disorder: a randomized controlled trial. *Journal of Behavior Therapy and Experimental Psychiatry*, *40*, 317–328. https://doi.org/10.1016/j.jbtep.2009.01.002

Giesen-Bloo, J., van Dyck, R., Spinhoven, P., van Tilburg, W., Dirksen, C., van Asselt, T., Kremers, I., Nadort, M. and Arntz, A. (2006). Outpatient psychotherapy for borderline personality disorder: randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Archives of General Psychiatry*, *63*, 649–658. https://doi.org/10.1001/archpsyc.63.6.649

Guydish, J., Campbell, B. K., Manuel, J. K., Delucchi, K., Le, T., Peavy, M. and McCarty, D. (2014). Does treatment fidelity predict client outcomes in 12-step facilitation for stimulant abuse? *Drug and Alcohol Dependence*, *134*, 330–336. https://doi.org/10.1016/j.drugalcdep.2013.10.020

Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A. and Liddle, H. A. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology*, *76*, 544–555. https://doi.org/10.1037/0022-006X.76.4.544

Kivlighan, D. M. and Lilly, R. L. (1997). Developmental changes in group climate as they relate to therapeutic gain. *Group Dynamics: Theory, Research, and Practice*, *1*, 208–208. https://doi.org/10.1037/1089-2699.1.3.208

Landis, R. J. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. https://doi.org/10.2307/2529310

Nenadić, I., Lamberth, S. and Reiss, N. (2017). Group schema therapy for personality disorders: a pilot study for implementation in acute psychiatric in-patient settings. *Psychiatry Research*, *253*, 9–12. https://doi-org./10.1016/j.psychres.2017.01.093

Ogrodniczuk, J. S. and Piper, W. E. (2003). The effect of group climate on outcome in two forms of short-term group therapy. *Group Dynamics: Theory, Research, and Practice*, *7*, 64–76. https://doi.org/10.1037/1089-2699.7.1.64

Reiss, N., Lieb, K., Arntz, A., Shaw, I. and Farrell, J. (2014). Responding to the treatment challenge of patients with severe BPD: results of three pilot studies of inpatient schema therapy. *Behavioural and Cognitive Psychotherapy*, *42*, 355–367. doi: 10.1017/S1352465813000027

Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Skewes, S. A., Samson, R. A., Simpson, S. G. and van Vreeswijk, M. (2015). Short-term group schema therapy for mixed personality disorders: a pilot study. *Frontiers in Psychology*, *5* (1592). doi: 10.3389/fpsyg.2014.01592

Soldz, S., Bernstein, E., Rothberg, P., Budman, S. H., Demby, A., Feldstein, M. et al. (1987). *Harvard Community Health Plan Group Cohesiveness Scale – Version II (GCS-II): Rater Manual*. Harvard Community Health Plan, Mental Health Research Program, Boston, MA.

Soldz, S., Budman, S. H., Demby, A. and Feldstein, M. (1990). Patient activity and outcome in group psychotherapy: New findings. *International Journal of Group Psychotherapy*, *40*, 53–62. https://doi.org/10.1080/00207284.1990.11490583

Wetzelaer, P., Farrell, J., Evers, S., Jacob, G. A., Lee, C. W., Brand, O., van Breukelen, G., Fassbinder, E., Fretwell, H., Harper, R. P., Lavender, A., Lockwood, G., Malogiannis, I. A., Schweiger, U., Startup, H., Stevenson, T., Zarbock, G. and Arntz, A. (2014). Design of an international multicentre RCT on group schema therapy for borderline personality disorder. *BMC Psychiatry*, *14*, 319. https://doi.org/10.1186/s12888-014-0319-3

Yalom, I. D. (1995). *The Theory and Practice of Group Psychotherapy* (4th edition). New York: Basic Books.

Zarbock, G., Rahn, V., Farrell, J. M. and Shaw, I. A. (2011). Group schema therapy: an innovative approach to treating patients with personality disorder [DVD]. Hamburg: IVAH.

Zarbock, G., Farrell, J. M., Schikowski, A., Heimann, A., Shaw, I., Reiss, N., Verhagen, S. and Bot, S. (2012). Group Schema Therapy Rating Scale (unpublished work). Hamburg: IVAH, Germany.

Zarbock, G., Farrell, J. M., Schikowski, A., Heimann, A., Shaw, I., Reiss, N., Verhagen, S., Bot, S., Lee, C. W., Arntz, A. and Bastick, E. (2014). Group Schema Therapy Rating Scale–Revised (GSTRS-R). Available at: https://doi.org/10.4225/23/585a265e14ab8.