Construction and validation of the Approve questionnaires – Measures of relating to voices and other people

Mark Hayward [a,*], Björn Schlier [b], Clara Strauss [c], Aikaterini Rammou [c], Tania Lincoln [b]

[a] Research & Development Department, Sussex Partnership NHS Foundation Trust, Nevill Avenue, Hove BN7 3HZ, UK
[b] Universität Hamburg, Institute of Psychology, Department of Clinical Psychology and Psychotherapy, Von-Melle-Park 5, 20146 Hamburg, Germany
[c] School of Psychology, University of Sussex, Brighton BN1 9RH, UK

ABSTRACT

Background: The effectiveness of psychological treatments for auditory hallucinations ('voices') needs to be enhanced. Some forms of novel treatment are working within relational frameworks to support patients to relate assertively to distressing voices. Yet, no measure of assertive relating to voices is available to assess the extent to which this skill is developed during therapy. This study aimed to assess the factor structure and validity of two new questionnaires: a measure of relating to voices and a measure of social relating.

Methods: The relating measures were developed in consultation with members of the international research community and validated in a large sample ($N = 402$) of voice hearing patients within the UK. The measures were subjected to factor analysis and compared to measures of voice hearing, mental health and well-being to evaluate construct, convergent, discriminant, and criterion validity.

Results: Factor analysis confirmed a three-dimensional set of items that measure assertive and non-assertive (passive and aggressive) relating. This resulted in the validation of the 'Approve' questionnaires - two 15-item measures of relating to voices and other people.

Conclusion: The Approve questionnaires can be used to assess a patient's suitability for relationally-based psychological therapies for distressing voices and the extent to which assertive relating skills are developed during the therapy.

1. Introduction

Auditory hallucinations (or 'voices') are a common and distressing experience across mental disorders (Waters and Fernyhough, 2017). Cognitive Behaviour Therapy (CBT) is the recommended psychological treatment for distressing voices (National Collaborating Centre for Mental Health, 2014). However, between-groups outcomes are consistently limited to small-moderate effect sizes (Van der Gaag et al., 2014). These limited effect sizes have prompted attempts to improve the effectiveness of interventions by focusing more specifically on the processes that are maintaining distress (Lincoln and Peters, 2018). Some of these novel developments are characterized by a shift from conceptualizing a voice as a sensory stimulus that the hearer holds beliefs about (the central tenet of CBT), to a voice as a person-like stimulus which the hearer has a relationship with (Hayward et al., 2011). At least three therapies are being developed that attempt to modify the way that patients can relate to the voices they hear: the Voice Dialogue approach conceptualizes a voice as a dissociated 'part' of the self and seeks to facilitate constructive 'live' dialogue between the patient and

the voice (Corstens et al., 2012); within AVATAR therapy, a visual depiction of the voice is created and displayed on a computer screen and the patient is coached to respond assertively to this avatar (whose responses are generated by the therapist in a different room; Craig et al., 2018); and Relating Therapy uses experiential role-plays to practice relating assertively to the typical utterances of the voice (or the social other with whom the patient is in a difficult relationship; Hayward et al., 2017). Despite the development of these relationally-based therapies and their emphasis upon relating assertively to voices, there is no psychometrically robust measure of assertive relating to voices, making it difficult to assess this important proposed change mechanism. To date, the relationally-based therapies have used either single items within broader measures (e.g. the 'Assertiveness' item on the Voice Power Differential scale; Birchwood et al., 2004) or measures that foreground perceptions of the relating of the voice (e.g., 'My voice does not let me have time to myself') and the relating preferences of the patient (e.g., 'I do not wish to spend much time listening to the voice'; items from the Voice & You; Hayward et al., 2008). As each of the therapies aims to teach patients to relate in an assertive manner towards voices, a robust measure is required that can capture changes in the relating of the patient and assess how these changes might mediate changes in voice-related distress. Additionally, a measure of assertive relating

* Corresponding author.
  E-mail address: mih21@sussex.ac.uk (M. Hayward).

within social environments may be useful as: 1) social relating appears to reflect voice relating (Birchwood et al., 2004; Hayward, 2003), and comparable scales for voice and social relating would allow for the assessment of these similarities as a test of convergent validity: and 2) this would facilitate the assessment of the mediating effect of changes in voice relating on social relating. Thus, we aimed to develop and validate self-report questionnaires for assertive/non-assertive relating that assess both relating to voices and relating to others.

2. Methods

2.1. Design

Measures were developed and validated in a large sample of voice hearing patients within the United Kingdom (UK). The measures were evaluated with regard to construct, convergent, discriminant and criterion validity.

2.2. Sample

Participants for this study were patients recruited from inpatient units and community mental health teams in 14 Mental Health Trusts within the National Health Service (NHS) in England. Inclusion criteria were: (1) age of at least 18 years, (2) having heard voices for at least six months, irrespective of diagnosis, and (3) sufficient language skills to complete the questionnaires.

The sample consisted of 402 participants with a mean age of 40.5 years. The majority of participants described themselves as male, of White ethnicity and currently unemployed (see Table 1). The majority of participants self-reported a diagnosis of Schizophrenia Spectrum Disorder (73.6%). On average, participants started hearing voices at the age of 22.4 years.

Table 1
Demographic profile of participants.

Participant characteristic	Category	Count (%)
Age	Mean no. years (sd)	40.5 (13.4)
Gender	Male	245 (60.9)
	Female	151 (37.6)
	Another term	2 (0.5)
	Prefer not to say	1 (0.2)
Employment status	Employed	71 (17.6)
	Unemployed	269 (66.9)
	Student	21 (5.2)
	Retired	21 (5.2)
	Other	16 (4.0)
	Prefer not to say	3 (0.7)
Ethnic group	White British	351 (87.3)
	Asian	16 (4.0)
	Black	8 (2.0)
	Mixed	11 (2.7)
	Other	14 (3.5)
	Prefer not to say	1 (0.2)
Education	No formal qualifications	56 (13.9)
	Completed secondary education	136 (33.8)
	College	91 (22.6)
	University	72 (17.9)
	Other	42 (10.4)
	Prefer not to say	5 (1.2)
Age onset	Mean no. years (sd)	22.4 (11.28)
Diagnosis	None	36 (9.0)
	Schizophrenia Spectrum	296 (73.6)
	Bipolar	16 (4.0)
	BPD/EUPD	30 (7.5)
	PTSD	6 (1.5)
	Depression and/or anxiety	17 (4.2)
	OCD	1 (0.2)

Notes: Missing data for age (n = 2), gender (n = 3), employment (n = 1), ethnicity (n = 1) and age at onset (n = 5).

2.3. Materials

2.3.1. Relating to voices/others questionnaires (the 'Approve' questionnaires)

The items for the self-report Approve questionnaires were created and selected during a 3-stage process. First, the authors created a list of 55 potential items that were drawn from the literature, clinical experience and the experience of developing relationally-based therapies. These items were grouped into three categories: one category for assertive relating and two categories for non-assertive relating (aggressive relating and passive relating). Each item was written so it could apply both to relating to voices and social relating. Second, the list was reviewed by a group of international experts (including researchers, clinicians, and people with lived experience of hearing voices; n = 45) through an online consultation. The experts were asked to indicate whether or not each item measured the category to which it had been allocated and to suggest additional items. All items that were endorsed by at least 70% of the experts were carried forward (37 items), together with 2 items that had lower endorsement (57% and 65%) but were considered important to the balance of the measures. Eight additional items were suggested and also carried forward. This process led to a reduced and refined short-list of 47 items. During the third stage, the wording of the items on the short-list were reviewed by two people with lived and current experience of hearing distressing voices, which led to some word changes and the removal of one item.

The development process resulted in two separate measures: a 46-item measure of relating to voices (Approve-Voices); and a 46-item measure of social relating (Approve-Social). The items were preceded by an introductory text inviting participants to "please select the answer that best reflects your typical response to [voices/other people] on the scale 0 (disagree completely) to 10 (agree completely). Where the item is not relevant to you then please select the not applicable (N/A) option". The following instruction - "when [voices/other people] are being difficult (e.g., treating me badly), I respond by:" was presented before the list of the items (e.g., "Hearing what they are saying but also stating my own views").

2.3.2. Voice hearing

To assess the severity, characteristics, and impact of voice hearing, the Hamilton Program for Schizophrenia Voices Questionnaire (HPSVQ) was used. The HPSVQ is a 9-item measure of the characteristics, content and impact of voices over the past week (Van Lieshout and Goldberg, 2007). Scores on the HPSVQ correlate highly (Kim et al., 2010) with scores on the widely used, clinician-administered Positive Symptoms Rating Scale (PSYRATS) - auditory hallucination (AH; Haddock et al., 1999). For this study we calculated the HPSVQ total score as well as the subscale score for voice characteristics and voice impact.

2.3.3. Well-being

As a measure of well-being, the short Warwick Edinburgh Mental Well-Being Scale (WEMWBS - 7-item self-report questionnaire; Stewart-Brown et al., 2009) was used. The scale has been shown to have strong internal consistency, test-retest reliability and concurrent validity (Stewart-Brown et al., 2011). A sum-score of the 7 items was calculated.

2.3.4. Depression, anxiety, and stress

The 21-item Depression Anxiety Stress Scales (DASS-21) was included as a measure of mental health. The DASS-21 has excellent internal consistency and concurrent validity (Antony et al., 1998) and adequate construct validity (Henry and Crawford, 2005). We used the sum-scores for depression, anxiety, and stress as criteria for mental health.

2.4. Procedure

Clinicians within inpatient units and community mental health teams were asked to identify patients from their existing caseloads who met the inclusion criteria. The clinicians were encouraged to discuss the study with potential participants, using the participant information sheet to guide discussions.

Patients who expressed an interest in participation were invited to meet with a member of the research team. At this meeting, the participant information sheet was reviewed and the patient was encouraged to ask questions about the study. Eligibility was screened and, if appropriate, the consent form was reviewed and signed. Written informed consent was obtained from all patients. The measures were completed in line with an assessment protocol. It was made clear to the participant that they were free to end their participation within the study at any point, without giving a reason and without affecting the care they received.

Upon starting the assessment, participants initially provided demographic data. Next, they answered either the Approve-Voices or Approve-Social questionnaire. The order of presentation of the Approve questionnaires was reversed in half of the recruiting sites. The criterion validation questionnaires followed, starting with the HPSVQ, followed by the WEMWBS and the DASS-21. Next, participants answered the other version of the Approve questionnaires. The two Approve questionnaires were located at either end of the assessment process in an attempt to distinguish between the differing, but related foci of the measures. The assessment concluded with four open-ended questions to provide an opportunity for further feedback on the questionnaires.

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human subjects/patients were approved by an NHS Research Ethics Committee (reference 18/LO/0046) and the NHS Health Research Authority.

2.5. Data analysis

All analyses were calculated using R 3.4.2. utilizing the package lavaan (Rosseel, 2012). We followed a two-step approach consisting of (1) scale construction and item selection using the Approve- Voices data and (2) cross-validation of the final version of the Approve questionnaires using the Approve-Social data.

2.5.1. Scale construction and item selection

In order to arrive at a final version for the Approve questionnaires, we first tested whether the answers to the 46-item list fitted a three-dimensional model of assertive vs. aggressive vs. submissive relating using exploratory factor analysis (EFA) with promax rotation and extraction of three factors. Fit to the three-dimensional model was determined based on sufficient item loading on the preconceived factor (loading > 0.4, Bandalos and Gerstner, 2016) and sufficient difference between primary loading and cross-loadings on the remaining factors ($\Delta_{loadings} < 0.2$). Next, the 46 Approve-Voices items were subjected to Confirmatory Factor Analyses (CFA). Model-fit was double-checked by comparing the three factor-model (assertive vs. aggressive vs. passive relating) with more parsimonious one-factor and two-factor (assertive vs. non-assertive relating) models. Finally, repeated CFA of the Approve-Voices items guided an iterative process of item elimination. Since the estimates of theory-guided primary loadings in CFA correspond to EFA values (minus the additional information on cross-loadings from EFA, see: Bandalos and Gerstner, 2016) we used loading estimates from CFA and changes in fit-indices to guide item elimination (for a pre-existing example of this approach, see: Leung et al., 2012). In the first iteration, we excluded items with lower standardized loading on their respective factor and re-ran the CFA on the reduced item list. In the following iterations, further items were excluded based on item

loading and content (i.e., including items that provided unique content and excluding items with content overlapping with other items). We ended the iteration process when sufficient model-fit had been achieved.

To determine model fit we used the three indices CFI (good fit: CFI > 0.95; Hu and Bentler, 1999, sufficient fit CFI > 0.90; McDonald and Ho, 2002), RMSEA (good fit: RMSEA < 0.06; Hu and Bentler, 1999, sufficient fit: RMSEA < 0.08; McDonald and Ho, 2002), and SRMR (good fit: SRMR < 0.08; Hu and Bentler, 1999).

2.5.2. Cross-validation of the final version of the scales

The final item selection was then cross-validated with the Approve-Social items. Again, model-fit was determined based on CFI, RMSEA, and SRMR. All factor analyses were calculated with maximum likelihood estimation with robust Huber-White standard errors and a scaled test statistic asymptotically equal to the Yuan-Bentler test statistic using both complete and incomplete data sets.

2.5.3. Validity assessment

For the assessment of convergent, discriminant, and criterion validity, we calculated Bravais-Pearson correlation between the Approve scores and the respective criteria. Approve scores were calculated based on the items answered with 0–10; N/A-answers were set to missing and not used in the calculation of a participant's score.

3. Results

3.1. Item characteristics

For all 46 candidate items of the Approve-Voices and Approve-Social questionnaires, the range of responses was 0 to 10. The amount of N/A-answers for Approve-Voices ranged between 5 and 19, and there were between 7 and 11 missing values per item with 3 participants responding to all items with "N/A" and 7 participants responding to none of the items. Regarding the Approve-Social items, the amount of N/A answers ranged from 0 to 11 whereas missing values ranged from 1 to 4, with only one participant providing a complete set of missing values. Item mean scores ranged from 2.30 to 6.33 (*SD* range: 3.20 to 3.84) for Approve-Voices and from 1.83 to 6.27 (*SD* range: 2.83 to 3.82) for Approve-Social.

3.2. Scale construction and item elimination

The EFA loadings for the full Approve-Voices 46 item list are shown in Table 2. As can be seen, extraction of three factors yielded an overall pattern consistent with the theory-based allocation of the items: All but six items showed substantial loading on their primary factor (40 items with loading > 0.4). Of the 40 items with substantial loading, only three showed substantial cross-loading. Thus, the overall pattern of assertive vs. aggressive vs. submissive relating emerged from the data of the raw item list.

Initial CFA of the full Approve-Voices item list yielded better model fit for a three-dimensional model ($CFI = 0.741, RMSEA = 0.068, SRMR = 0.114$, loadings: 0.185–0.807) than a one-dimensional ($CFI = 0.339, RMSEA = 0.109, SRMR = 0.178$, loadings: −0.721-0.447) or two-dimensional model ($CFI = 0.543, RMSEA = 0.091, SRMR = 0.158$, loadings: 0.099–0.796). Based on the cross-loadings found in EFA, the limited absolute fit was likely due to item quality, so we started an iterative elimination process.

The three-dimensional CFA yielded 27 items with a standardized loading above 0.60, with 8 items for the assertive relating factor, 6 items for the aggressive relating factor, and 13 items for the passive relating factor. In order to include a roughly equal amount of items for all factors, only the 8 items with the highest loadings on the passive relating factor were retained for further analysis, whereas all 8 and 6 items

for the assertive and aggressive relating factors with loadings over 0.60 were included.

A second three-dimensional CFA based on the 22 items with the highest loadings yielded increased but still non-sufficient model fit ($CFI = 0.879$, $RMSEA = 0.069$, $SRMR = 0.084$, loadings: 0.563–0.866). Further item deletion was guided by item-loading and uniqueness of item content. For the assertive relating factor, the two items with the lowest loading were excluded ("…Defending my own view in a sensitive manner" and "…Defending myself in a calm and confident manner"). Regarding the aggressive relating factor, the item with the second-lowest loading ("…Insulting them") was excluded due to high content overlap with other items (e.g., "…Swearing at them" and "…Yelling at them", see Table 2). Finally, the two items with the lowest loading on the passive relating factor were excluded ("…Feeling helpless" and "…Listening to them and telling myself they are right") and the item with the fourth-lowest loading ("…Giving in, even if I don't agree with what they are saying") was excluded due to content-overlap.

A CFA of the resulting 16 Approve-Voices items yielded sufficient fit according to one indicator ($CFI = 0.935$) and good fit according to the other two indicators (RMSEA = 0.059, SRMR = 0.070). In order to provide an equal amount of items for each subscale and to ensure the final measure minimized participant burden, one item was eliminated from the final list based on its complexity and narrow scope in terms of content ("…Letting them know that I have heard all this before and am not prepared to listen to it anymore at the moment"). The final 15-item Approve-Voices questionnaire with 5 items in each subscale yielded unchanged high standardized loadings (Table 3) and sufficient model fit ($CFI = 0.940$, $RMSEA = 0.059$, $SRMR = 0.069$).

3.3. Cross-validation

A Cross-validation CFA using the corresponding item-selection from the Approve-Social questionnaire yielded sufficient to good fit ($CFI = 0.930$, $RMSEA = 0.060$, $SRMR = 0.073$) and consistently high

Table 2
Item loadings for the EFA of the initial 46 item Approve-Voices.

Item text: When voices are being difficult (e.g., treating me badly), I respond by…	F1 (Submissive)	F2 (Assertive)	F3 (Aggressive)
Assertive relating items			
Defending my own view in a sensitive manner.	0.153	**0.795**	−0.240
Presenting and defending my own view.		**0.753**	
Hearing what they are saying but also stating my own views.		**0.686**	
Defending my own view in a sensitive manner.	−0.121	**0.674**	−0.213
Repeating my own opinion if they don't listen the first time.		**0.644**	0.232
Negotiating with them in a way equal partners would.	0.406	**0.625**	−0.182
Standing up for myself.	−0.302	**0.608**	0.115
Letting them know that I have heard all this before and am not prepared to listen to it anymore at the moment.		**0.608**	0.207
Respectfully disagreeing with what they are saying.		**0.593**	−0.123
Calmly stating that I don't agree with them.	−0.132	**0.580**	−0.125
Letting them know that I wish to be left in peace right now.		**0.576**	0.244
Trying to see their point of view without necessarily agreeing with it.	*0.382*	**0.421**	*−0.275*
Sticking with what I feel is true even if they say otherwise.	−0.267	**0.413**	
Just not giving in, even if they continue to disagree with me.	−0.221	0.384	
Aggressive relating items			
Yelling at them.			**0.852**
Shouting and screaming.		−0.110	**0.814**
Swearing at them.			**0.769**
Shouting out loud at them.			**0.750**
Insulting them.	−0.178	0.152	**0.648**
Threatening them.		0.111	**0.567**
Telling them to shut up.		0.194	**0.550**
Arguing with them.		0.174	**0.505**
Shouting at them in my head.	0.176	0.126	**0.470**
Throwing things.	0.157	−0.141	**0.450**
Fighting back.		0.424	**0.429**
Feeling like I want to destroy them.	*0.239*		*0.389*
Blocking my ears.	*0.202*		*0.372*
Feeling violent towards others.	*0.241*		*0.360*
Taking my anger out on someone else.	*0.252*	−0.119	*0.360*
Submissive relating items			
Giving in.	**0.721**		
Telling myself they are right even though I don't believe they are right.	**0.712**		
Finding myself at their mercy.	**0.703**		
Listening to them and telling myself they are right.	**0.701**		
Doing what they want.	**0.701**	−0.126	
Trying to stick up for myself but eventually giving in if they persist.	**0.695**	0.204	
Giving in, even if I don't agree with what they are saying.	**0.675**		
Trying to satisfy them so they leave me alone.	**0.672**	0.184	
Allowing them to get on top of me.	**0.641**	−0.12	
Feeling afraid to say anything.	**0.637**	−0.166	
Feeling helpless.	**0.612**	−0.14	
Begging them to stop.	**0.601**	0.151	0.223
Wanting to run away.	**0.578**	−0.124	0.138
Telling them they are right.	**0.502**		
Silently enduring it.	**0.459**		−0.319
Trying to hide from them in some way.	**0.425**		
Doing something to try to calm them down.	0.277	*0.325*	

Note. Only factor loadings > 0.1 are depicted. Items are presented in descending order of primary loading. Primary loadings > 0.4 are printed in bold; Cross-loadings with $|\Delta_{primary\ loading}| < 0.2$ are printed in italics. Items that are retained after the item CFA-guided item reduction are printed in bold.

standardized factor loadings (Table 3). Results for both CFA in terms of fit indices and loading remained stable when conventional maximum-likelihood estimation was used (Approve-Voices: CFI = 0.931, RMSEA = 0.071, SRMR = 0.068; Approve-Social: CFI = 0.926, RMSEA = 0.071, SRMR = 0.068). Internal consistencies for all three subscales were good in both the Approve-Voices (assertive relating: $\alpha = 0.85$, aggressive relating: $\alpha = 0.86$, passive relating: $\alpha = 0.83$) and the Approve-Social (assertive relating: $\alpha = 0.79$, aggressive relating: $\alpha = 0.88$, passive relating: $\alpha = 0.81$).

3.4. Convergent and discriminant validity: associations between relating styles

Significant and substantial correlation was found between the scales of the Approve-Voices and Approve-Social measures assertive relating ($r = 0.43$, $t(386) = 9.47$, $p < 0.001$), aggressive relating ($r = 0.47$, $t(387) = 10.55$, $p < 0.001$), and passive relating scales ($r = 0.45$, $t(388) = 9.80$, $p < 0.001$), respectively, attesting to convergent validity of the factors.

Furthermore, correlations between different relating styles to the same targets (voices or other people) were lower than the aforementioned correlations between the same relating styles across different targets, indicating discriminant validity: Assertive relating and passive relating were found to be negatively correlated when relating to voices ($r = -0.24$, $t(386) = -4.90$, $p < 0.001$) and when relating to others ($r = -0.39$, $t(399) = -8.40$, $p < 0.001$). There was a positive association between assertive and aggressive relating to voices ($r = 0.28$, $t(387) = 5.78$, $p < 0.001$) and to others ($r = 0.23$, $t(399) = 4.63$, $p < 0.001$). Finally, the association between aggressive and passive relating to voices was moderately positive ($r = 0.35$, $t(387) = 7.34$, $p < 0.001$), but only small for relating to others ($r = 0.13$, $t(399) = 2.62$, $p = 0.009$).

3.5. Criterion validation: association between relating to voices/others and mental health

There was no significant correlation between assertive relating to voices and the HPSVQ total score ($r = 0.08$, $t(381) = 1.66$, $p = 0.098$) or either of its subscales (voice characteristics: $r = 0.09$, $t(382) = $ 1.77, $p = 0.078$; voice impact: $r = 0.06$, $t(385) = 1.16$, $p = 0.245$). Assertive relating to others correlated significantly with lower HPSVQ total ($r = -0.12$, $t(392) = -2.38$, $p = 0.018$) and subscale scores (voice characteristics: $r = -0.11$, $t(393) = -2.20$, $p = 0.030$; voice impact: $r = -0.12$, $t(396) = -2.35$, $p = 0.019$). Aggressive and passive relating to voices were significantly associated with increased HPSVQ total scores (aggressive relating: ($r = 0.26$, $t(382) = 5.23$, $p < 0.001$; passive relating: $r = 0.32$, $t(383) = 6.59$, $p < 0.001$), and each if its subscales - voice characteristics (aggressive relating: ($r = 0.21$, $t(383) = 4.28$, $p < 0.001$; passive relating: $r = 0.30$, $t(384) = 6.27$, $p < 0.001$) and voice impact scale (aggressive relating: ($r = 0.26$, $t(386) = 5.22$, $p < 0.001$; passive relating: $r = 0.28$, $t(387) = 5.81$, $p < 0.001$). A similar pattern was found for passive relating to others ($0.18 \leq r \leq 0.22$, all $p < 0.001$), but not for aggressive relating to others ($0.05 \leq r \leq 0.08$, all $p > 0.050$).

Regarding mental health, there was a consistent pattern for Approve-Voices and Approve-Social subscales (see Table 4 for an overview of the correlation coefficients). The assertive relating scales showed moderate correlations with increased well-being and small negative associations with the severity of depression. The aggressive relating scales showed small negative associations with well-being, and small to moderate associations with depression, anxiety, and stress. Finally, the passive relating scales showed moderate associations with decreased well-being, as well as increased severity of depression, anxiety and stress.

3.6. Exploratory subgroup analysis of the association between assertive relating and voice impact

Given the unexpected lack of an association between assertive relating to voices on Approve-Voices and voice impact on HPSVQ, we explored the possible influence of the frequency and content of voices. Sub-group analyses were conducted based on two assumptions:

1. in order to report voice impact meaningfully during the last week (i.e., HPSVQ voice impact score > 0), voices needed to be present (i.e. HPSVQ item 1 "how frequently do you hear a voice or voices?" > 0) during the last week

2. in order for relating "when voices are treating me badly" to be associated with voice impact, participants had to indicate that during the last week, voice content had been negative to some degree (i.e., HPSVQ item 2 "how bad are the things the voices say to you?" > 0 ("no voices saying bad things").

As we had no preconceived ideas about the minimal frequency of voices or amount of negative voice content required for assertive relating to be associated with reduced voice impact, multiple subsamples were created based on incrementally increasing thresholds for voice hearing frequency and negative voice content. For each subsample, the association between assertive relating and voice impact (i.e., the sum of the voice impact items excluding HPSVQ item 2) was recalculated. The results for the associations are shown in Fig. 1 and illustrate a significant negative association between assertive relating and

Table 3
Item loadings for the final CFA on the 15-item Approve.

Factor	Voices version		Social version	
Item text: When voices/other people are being difficult (e.g., treating me badly), I respond by…	Loading	SE	Loading	SE
Assertive responding factor				
…Repeating my own opinion if they don't listen the first time.	0.798	0.036	0.621	0.052
…Presenting and defending my own view.	0.792	0.033	0.759	0.034
…Hearing what they are saying but also stating my own views.	0.705	0.042	0.678	0.055
…Standing up for myself.	0.692	0.046	0.797	0.041
…Letting them know that I wish to be left in peace right now.	0.641	0.044	0.419	0.060
Aggressive responding factor				
…Yelling at them.	0.866	0.022	0.871	0.023
…Shouting and screaming.	0.811	0.029	0.753	0.036
…Shouting out loud at them.	0.804	0.032	0.849	0.028
…Swearing at them.	0.714	0.036	0.757	0.035
…Telling them to shut up.	0.559	0.043	0.626	0.043
Passive/submissive responding factor				
…Allowing them to get on top of me.	0.743	0.039	0.678	0.051
…Finding myself at their mercy.	0.735	0.041	0.731	0.043
…Giving in.	0.733	0.039	0.665	0.039
…Doing what they want.	0.691	0.045	0.720	0.043
…Telling myself they are right even though I don't believe they are right.	0.647	0.043	0.578	0.054

Note. Standardized loadings. All loadings significant with $p < 0.001$. SE = Standard Error.

Table 4
Association of Approve subscales and general mental health parameters.

	Well-being	Depression	Anxiety	Stress
Assertive to voices	0.27***	−0.19***	−0.09+	−0.11*
Aggressive to voices	−0.18***	0.28***	0.33***	0.29***
Passive/submissive to voices	−0.47***	0.49***	0.41***	0.41***
Assertive to other people	0.28***	−0.15**	−0.14**	−0.06
Aggressive to other people	−0.16**	0.22***	0.26***	0.29***
Passive/submissive to other people	−0.31***	0.37***	0.35***	0.33***

Note. Well-Being denotes the WEMWBS-sumscore. Depression, Anxiety, and Stress are scores of the respective DASS subscale. *** = p < 0.001; ** = p < 0.01; * = p < 0.05; + = p < 0.1.

voice impact that becomes evident when voices are present on a daily basis ($r = -0.13$, $t(303) = -2.21$, $p = 0.028$) and further increases when negative voice content is present ($r = -0.19$, $t(287) = -3.30$, $p = 0.001$). The associations between assertive relating and mental health followed a similar pattern and were also increased in the sub-sample characterized by voice hearing once or twice a day and at least some amount of negative content (well-being: $r = 0.33$, $t(285) = 5.95$, $p < 0.001$; depression: $r = -0.27$, $t(283) = -4.81$, $p < 0.001$; anxiety: $r = -0.18$, $t(282) = -3.01$, $p = 0.003$; stress: $r = -0.20$, $t(279) = -3.35$, $p < 0.001$).

4. Discussion

Novel psychological treatments for distressing voices are exploring the use of a variety of therapeutic techniques to enable a patient to respond more assertively within these difficult relationships (Corstens et al., 2012; Hayward et al., 2017; Craig et al., 2018). However, there is currently no psychometrically validated questionnaire available to assess the extent to which assertiveness skills are being developed during therapy. This study aimed to assess the psychometric properties of measures of assertive and non-assertive relating to voices and other people (named Approve-Voices and Approve-Social, respectively) that were developed in consultation with experts from the international research community.

The Approve questionnaires were completed by a large trans-diagnostic sample of voice-hearing patients who were using mental health services within the NHS in England. Data were factor analyzed and generated two 15-item measures across three sub-scales – one sub-scale assessing assertive relating, and two sub-scales assessing non-assertive (aggressive and passive) relating. The assessment of relating styles to voices and other people suggested that the measures and their sub-scales had some degree of convergence (associations ranging from $r = 0.43$–0.47). These findings were consistent with previous studies reporting that relationships with voices and other people share some similarities, but also have some differences (Hayward, 2003; Mawson et al., 2011). Discriminant validity was evidenced across both measures by the expected positive associations between the non-assertive (passive and aggressive) forms of relating, and the expected negative association between passive and assertive relating. However, an unexpected positive association was found between assertive relating and aggressive relating ($r = 0.28$), suggesting that these sub-scales may not be entirely independent.

Criterion validity was evident across assertive and non-assertive forms of relating for both measures as passive and aggressive relating

were associated with poorer mental health and assertive relating was associated with increased mental health. A similar pattern of associations was also expected for associations with the severity of voice hearing but was evident for only non-assertive relating. The expected association between assertive relating and less severity of voice hearing was evident only in patients where voices were more frequent and communicating negative content. These findings suggest that an assertive style of relating may be helpful for mental health generally but may only be impacting upon voice hearing distress when this experience is current and voice content is negative.

This study has limitations in several respects. Firstly, diagnosis was self-reported by participants and was not clinician verified. The more robust collection of diagnostic information in a future study would facilitate the exploration of any differences in relating across groups of patients with different diagnoses. This would be particularly pertinent at a time when the trans-diagnostic assessment and treatment of distressing voices is attracting attention (Hazell et al., 2018). Secondly, there was no attempt to measure the participants' stage of recovery – a further variable that may influence the use of assertive and non-assertive relating. Thirdly, further research is required to explore the conceptual distinctiveness of assertive and aggressive forms of relating which were unexpectedly found to be positively associated. Fourthly, as self-report questionnaires, the Approve measures cannot facilitate an objective evaluation of relating styles. Future research could include the validation of self-reported relating styles in a behavioural role-play task. Fifthly, the Approve questionnaires assess only relating within relationships that are perceived to be 'difficult'. Hearers can develop positive relationships with voices (e.g., Jackson et al., 2011), but the measurement of responding within these relationships would require a separate assessment. Sixthly, the Approve questionnaires make no attempt to explore and determine the cause/origin of voice hearing experiences. Whilst this is consistent with the focus of relationally based therapies on responding to voices in the here-and-now, therapy informed by a longitudinal formulation will require exploration beyond the information generated by the Approve measures. Finally, the cross-validation of the final item-selection based on independent data (the Approve-Social) from the same sample provides some evidence for the validity of the model fit. However, it falls short of the gold standard for scale construction (validating model fit on a new, independent sample; Matsunaga, 2010), so future research is needed to provide further evidence for the three-dimensional model of relating to voices/others. Nevertheless, this study has a significant strength in relation to the size of the sample allowing robust CFA to be conducted. Seminal measures in the field have previously been developed and psychometrically evaluated on small samples (e.g., BAVQ-r; Chadwick et al., 2000) and/or involved the secondary analysis of combined datasets (e.g., PSYRATS-AH; Woodward et al., 2014).

The Approve questionnaires[1] can be used prior to and at the conclusion of relationally-based psychological therapy for distressing voices. Prior to the offering of therapy, an assessment of relating styles could help to identify the patients who might be most likely to benefit from developing assertive ways of relating – possibly patients who score highly on non-assertive relating (Hayward et al., 2016; Strauss et al., 2018) – and this suggestion can be tested through future research. At the conclusion of therapy, assessment can inform the identification of the mechanisms (e.g., assertive relating) that may have influenced beneficial outcomes. Future research is needed to examine if improvements in assertive relating and reductions in non-assertive relating mediate improved outcomes. Such an assessment would be crucial at a time when relationally-based therapies are seeking to understand the mechanisms and processes by which they are generating encouraging outcomes (Alderson-Day and Jones, 2018; Hayward, 2018). Moreover, the use of the Approve-Social questionnaire would facilitate an assessment
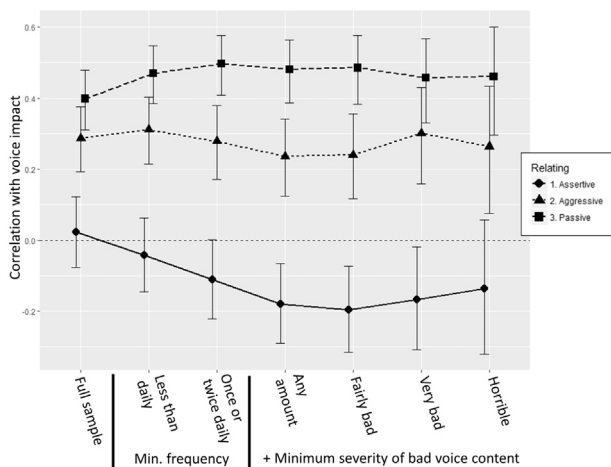

Fig. 1. Correlation between voice impact and relating to voices factors in the full sample and subsamples based on minimum amount of voice hearing (HPSVQ item 1) and minimum amount of bad content of voices (HPSVQ item 2). Correlation coefficient with 95% confidence interval.

[1] The Approve questionnaires can be downloaded from https://www.sussexpartnership.nhs.uk/sussex-voices-clinic.

of the extent to which any changes in relating to voices may be mirroring changes within difficult social relationships.

Author contribution

All of the authors were involved in the design of the study. MH and AR co-ordinated the consultation and the collection of data from participants. BS and TL conducted the statistical analysis. All of the authors were involved in the drafting and refinement of the manuscript.

Declaration of competing interest

None.

References

Alderson-Day, B., Jones, N., 2018. Understanding AVATAR therapy: who, or what, is changing? The Lancet Psychiat 5, 2–3.

Antony, M.M., Bieling, P.J., Cox, B.J., Enns, M.W., Swinson, R.P., 1998. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. Psychol Assessment 10, 176–181.

Bandalos, D.L., Gerstner, J.J., 2016. Using factor analysis in test construction. In: Schweizer, K., DiStefano, C. (Eds.), Psychological Assessment–Science and Practice. Principles and Methods of Test Construction: Standards and Recent Advances 3. Hogrefe Publishing, pp. 26–51.

Birchwood, M., Gilbert, P., Gilbert, J., Trower, P., Meaden, A., Hay, J., Murray, E., Miles, J.N.V., 2004. Interpersonal and role-related schema influence the relationship with the dominant "voice" sin schizophrenia: a comparison of three models. Psychol. Med. 34, 1571–1580.

Chadwick, P., Lees, S., Birchwood, M., 2000. The revised beliefs about voices questionnaire (BAVQ-R). Brit J Psychiat 177, 229–232.

Corstens, D., Longden, E., May, R., 2012. Talking with voices: exploring what is expressed by the voices people hear. Psychosis 4, 95–104.

Craig, T.K.J., Rus-Calafell, M., Ward, T., Leff, J.P., Huckvale, M., Howarth, E., Emsley, R., Garety, P.A., 2018. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiat 5, 31–40.

Haddock, G., McCarron, J., Tarrier, N., Faragher, E.B., 1999. Scales to measure dimensions of hallucinations and delusions: the psychotic symptom rating scales (PSYRATS). Psychol. Med. 9, 879–889.

Hayward, M., 2003. Interpersonal relating and voice hearing: to what extent does relating to the voice reflect social relating? Psychol and Psychother-T 76, 369–383.

Hayward, M., 2018. Continuing the conversation about AVATAR therapy. Lancet Psychiat 5, 196.

Hayward, M., Denney, J., Vaughan, S., Fowler, D., 2008. The voice and you: development and psychometric evaluation of a measure of relationships with voices. Clin Psychol Psychot 15, 45–52.

Hayward, M., Berry, K., Ashton, A., 2011. Applying interpersonal theories to the understanding of and therapy for auditory hallucinations: a review of the literature and directions for further research. Clin. Psychol. Rev. 31, 1313–1323.

Hayward, M., Slater, L., Berry, K., Perona-Garcelán, S., 2016. Establishing the "fit" between the patient and the therapy: the role of patient gender in selecting psychological therapy for distressing voices. Front. Psychol. 7, 424.

Hayward, M., Jones, A.-M., Bogen-Johnston, L., Thomas, N., Strauss, C., 2017. Relating therapy for distressing auditory hallucinations: a pilot randomized controlled trial. Schizophr. Res. 183, 137–142.

Hazell, C.M., Greenwood, K., Fielding-Smith, S., Rammou, A., Bogen-Johnston, L., Berry, C., Jones, A.-M., Hayward, M., 2018. Understanding the barriers to accessing symptom-specific cognitive behavior therapy (CBT) for distressing voices: reflecting on and extending the lessons learnt from the CBT for psychosis literature. Front. Psychol. 9, 727.

Henry, J.D., Crawford, J.R., 2005. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Brit J Clin Psychol 44, 227–239.

Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 6, 1–55.

Jackson, L., Hayward, M., Cooke, A., 2011. Developing positive relationships with voices: a preliminary grounded theory. Int J Soc Psychiatr 57, 487–495.

Kim, S.H., Jung, H.Y., Hwang, S.S., Chang, J.S., Kim, Y., Ahn, Y.M., Kim, Y.S., 2010. The usefulness of a self-report questionnaire measuring auditory verbal hallucinations. Prog Neuro-Psychoph 34, 968–973.

Leung, D.Y., Wong, E.M., Chan, S.S., Lam, T., 2012. Psychometric properties of the big five inventory in a Chinese sample of smokers receiving cessation treatment: a validation study. J Nurs Educ Pract 3. https://doi.org/10.5430/jnep.v3n6p1.

Lincoln, T.M., Peters, E., 2018. A systematic review and discussion of symptom specific cognitive behavioural approaches to delusions and hallucinations. Schizoph Res https://doi.org/10.1016/j.schres.2017.12.014 Available from.

Matsunaga, M., 2010. How to factor-analyze your data right: do's, don'ts, and how-to's. Int J Psychol Res 3, 97. https://doi.org/10.21500/20112084.854.

Mawson, A., Berry, K., Murray, C.D., Hayward, M., 2011. Voice hearing within the context of hearer's social worlds: an interpretative phenomenological analysis. Psychology Psychother-T 84, 256–272.

McDonald, R.P., Ho, M.H.R., 2002. Principles and practice in reporting structural equation analyses. Psychol. Methods 7, 64–82.

National Collaborating Centre for Mental Health, 2014. Psychosis and Schizophrenia in Adults: Treatment and Management. NICE Clinical Guideline n° 178.

Rosseel, Y., 2012. Lavaan: an R package for structural equation modeling. J. Stat. Softw. 48, 1–36.

Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., Weich, S., 2009. Internal construct validity of the Warwick-Edinburgh mental well-being scale (WEMWBS): a rasch analysis using data from the Scottish health education population survey. Health Quality Life Out 7, 15.

Stewart-Brown, S., Platt, S., Tennant, A., Maheswaran, H., Parkinson, J., Weich, S., Al, E., 2011. The Warwick-Edinburgh mental well-being scale: a valid and reliable tool for measuring mental well-being in diverse populations and projects. J Epidemiol & Commun H 11.

Strauss, C., Hugdahl, K., Waters, F., Hayward, M., Bless, J., Falkenberg, L.E., Krakvik, B., AsBjornsen, A.E., Johnsen, E., Sinkevicuite, I., Kroken, R.A., Loberg, E.-M., Thomas, N., 2018. The beliefs about voices questionnaire–revised: a factorstructure from 450 participants. Psychiatry Res. 259, 95–103.

Van der Gaag, M., Valmaggia, L.R., Smit, F., 2014. The effects of individually tailored formulation-based cognitive behavioural therapy in auditory hallucinations and delusions: a meta-analysis. Schizoph Res 156, 30–37.

Van Lieshout, R.J., Goldberg, J.O., 2007. Quantifying self-reports of auditory verbal hallucinations in persons with psychosis. Can J Beh Sci 39, 73–77.

Waters, F., Fernyhough, C., 2017. Hallucinations: a systematic review of points of similarity and difference across diagnostic classes. Schizophrenia Bull 43, 32–43.

Woodward, T.S., Jung, K., Hwang, H., Yin, J., Taylor, L., Menon, M., Peters, E., Kuipers, E., Waters, F., Lecomte, T., Sommers, I.E., Dallman, K., van Lutterved, R., Hubl, D., Kindler, J., Homan, P., Badcock, J.C., Chhabra, S., Cella, M., Keedy, S., Allen, P., Mechelli, A., Preti, A., Siddi, S., Erickson, D., 2014. Symptom dimensions of the psychotic symptom rating scales in psychosis: a multisite study. Schizophrenia Bull 40, 265–274 SUPPL. 4.