Development and psychometric evaluation of the Interactive Test of Interpersonal Behavior (ITIB): A pilot study examining interpersonal deficits in chronic depression

JAN PHILIPP KLEIN,[1] MARIA KENSCHE,[1] NADINE BECKER-HINGST,[1] JÖRG STAHL,[1] CHRISTINA SPÄTH,[1] TILO MENTLER,[2] MATHIAS STOISLOW,[2] MICHAEL HERCZEG,[2] MICHAEL HÜPPE[3] and ULRICH SCHWEIGER[1]

[1]*Psychiatry and Psychotherapy, Lübeck University, Lübeck, Germany*
[2]*Institute for Multimedia and Interactive Systems, Lübeck University, Lübeck, Germany*
[3]*Anesthesiology, Lübeck University, Lübeck, Germany*

Chronic depression is assumed to be caused and maintained by interpersonal deficits. We describe the development and psychometric evaluation of the Interactive Test of Interpersonal Behavior (ITIB) that we developed for self-assessment of these deficits. Participants with chronic depression (CD, N = 15), episodic depression (ED, N = 15) and healthy controls (HC, N = 15) participated in this pilot study. They completed the ITIB and a number of pen and paper questionnaires including the Lübeck Questionnaire of Preoperational Thinking (LQPT) and the inventory of interpersonal problems (IIP). The ITIB was highly acceptable for use in these participants. Internal consistency for the ITIB was adequate for group comparisons (Cronbach's alpha = 0.649). Item-total correlations indicated adequate discriminatory power of five of the six items. The ITIB correlated moderately with the LQPT ($r = 0.524$) and the IIP ($r = -0.568$). The ITIB score differed significantly between the diagnostic groups (ANOVA $F(2,42) = 6.22$, $p = 0.004$). It was the only measure that – albeit at a trend level – was associated with diagnostic group (CD vs. ED) on multinomial logistic regression analysis (B = 0.049 ± 0.029; OR 1.051; $p = 0.088$). We found preliminary evidence that the ITIB is an acceptable and psychometrically adequate measure of interpersonal behavior that distinguishes between patients with CD and patients with ED. If replicated with an improved version of the test, our results could support the hypothesis that having interpersonal problems is a core deficit in patients with CD.

Key words: Chronic depression, persistent depressive disorder, psychometrics, interactive test, interpersonal behavior.

Jan Philipp Klein, *Psychiatry and Psychotherapy*, *Lübeck University*, *Ratzeburger Allee 160*, *Lübeck*, *Germany 23560*. E-mail: philipp.klein@uksh.de

INTRODUCTION

Chronic depression can broadly be defined as a depression lasting two years or longer. In the fifth edition of the Diagnostic and Statistical Manual of the American Psychiatric Association (DSM-5), the different forms of chronic depression (CD) have been summarized in one section under the heading "persistent depressive disorder (dysthymia)" (American Psychiatric Association, 2013). Around a third of patients with a lifetime diagnosis of a depressive disorder suffer from CD (Murphy & Byrne, 2012; Satyanarayana, Enns Cox & Sareen, 2009). CD has been associated with childhood adversity (Klein, Roniger, Schweiger, Späth & Brodbeck, in press; Wiersma, Hovens, van Oppen *et al.*, 2009), increased prevalence of psychiatric comorbidity such as anxiety disorders and personality disorders (Murphy & Byrne, 2012) as well as abnormalities in brain reactivity (Klein, Becker, Hurlemann, Scheibe, Colla & Heuser, 2014b). A better understanding of the factors underlying the development and maintenance of CD might lead to improved treatment and outcomes.

Patients with CD have been found to differ from patients with episodic depression (ED) and healthy controls (HC) in their pattern of interpersonal behavior. They tend to be more hostile and more submissive (Constantino, Manber, DeGeorge *et al.*, 2008). Submission has also been described as having greater difficulty expressing one's needs and feelings (Klein & Belz, 2014). Interpersonal deficits might contribute to the etiology and maintenance of chronic depression. Support for this assumption comes from the stress generation literature (Conway, Hammen & Brennan, 2012; Hammen, 2005) which describes that patients

with depression frequently report suffering from stressful interpersonal life events that are dependent, at least in part, on their own actions.

Interpersonal problems in patients with CD have been hypothesized to be associated with a specific deficit that has variously been termed "theory of mind deficit" (Wilbertz, Brakemeier, Zobel, Harter & Schramm, 2010; Zobel, Werden, Linster *et al.*, 2010), "preoperational thinking" (Kuhnen, Knappke, Otto *et al.*, 2011; McCullough, 2000) and "perceptual disconnection" (McCullough, 2000). These concepts converge on the assumption that patients with CD have difficulties perceiving the consequences of their interpersonal behavior and choosing their interpersonal style according to the needs of a given situation. The perceptual disconnection hypothesis is the central underlying assumption for a treatment model that has specifically been developed to treat CD, namely the Cognitive Behavioral Analysis System of Psychotherapy (CBASP) (McCullough, 2000).

While CBASP has been studied in several successful clinical trials (Brakemeier, Radtke, Engel *et al.*, 2015; Keller, McCullough, Klein *et al.*, 2000; Schramm, Zobel, Dykierek *et al.*, 2011; Wiersma, van Schaik, Hoogendorn *et al.*, 2014) its underlying assumptions have not been studied as extensively. A considerable number (Constantino *et al.*, 2008; Kuhnen *et al.*, 2011; Wolkenstein, Schonenberg, Schirm & Hautzinger, 2011; Zobel *et al.*, 2010) but not all studies (Wilbertz *et al.*, 2010) found evidence for interpersonal deficits in patients with CD. Furthermore, successful treatment with CBASP has been associated with improvements in interpersonal impact (Brakemeier *et al.*, 2015; Constantino, Laws, Arnow, Klein, Rothbaum & Manber,

2012) and functioning (Manber, Arnow, Blasey *et al.*, 2003; Swan, Macvicar, Christmas *et al.*, 2014).

These studies differed in the assessments that were used. Two studies (Wilbertz *et al.*, 2010; Wolkenstein *et al.*, 2011) employed a movie depicting a social interaction to explore theory of mind deficits (Dziobek, Fleck, Kalbe *et al.*, 2006), one study (Zobel *et al.*, 2010) assessed theory of mind using cartoon picture story tests. Another study employed a test that was specifically developed to assess preoperational thinking (Kuhnen *et al.*, 2011). Two further studies (Brakemeier *et al.*, 2015; Constantino *et al.*, 2008) assessed the interpersonal impact that patients with CD had on their therapist (Kiesler & Schmidt, 1993). The studies measuring acquisition of interpersonal skills in the course of therapy (Manber *et al.*, 2003; Swan *et al.*, 2014) used the patient performance rating form (PPRF) (McCullough, Lord, Conley & Martin, 2010).

In summary, the current literature suggests that one is more likely to find evidence of interpersonal deficits in CD if one employs measures that assesses the reaction of the patient in a "participant role" ("what do you say or do?") as opposed to an "observer role" ("what do you think person X is likely to say or do?") (Klein *et al.*, 2011; Wilbertz *et al.*, 2010).

To further explore interpersonal deficits in chronic depression we developed and evaluated an Interactive Test of Interpersonal Behavior (ITIB). This test focuses on the behavioral consequences of perceptual disconnection, namely the participant's ability to communicate in a friendly and goal-oriented manner. The test seeks to assess interpersonal behavior in a "participant role." In contrast to previous measures of interpersonal skills and behavior such as the interpersonal message inventory (IMI) (Kiesler & Schmidt, 1993) and the PPRF (McCullough *et al.*, 2010), this test can be self-administered by the patient. In this paper we describe the development of the test and a pilot study where we evaluated its psychometric properties.

We hypothesized that the test would be acceptable to participants, have adequate psychometric properties and find more deficits in interpersonal behavior in patients with chronic depression compared to patients with episodic depression and healthy controls.

METHODS

Sample description

Participants for this study were recruited by Becker-Hingst between February and April 2014. Recruitment of patients took place primarily in the department of psychiatry and psychotherapy and the department of psychosomatics of Lübeck University. A further route of patient recruitment was via local outpatient psychiatric and psychotherapeutic services. Healthy controls were staff members of the above mentioned departments as well as volunteers recruited with announcements. Controls were frequency matched for sex and age to the patient group. Participants did not receive financial compensation.

Inclusion criteria for the study were age between 18 and 65 years and adequate understanding of the German language. Patients with chronic depression were diagnosed according to DSM-5 criteria for persistent depressive disorder (American Psychiatric Association, 2013) using an algorithm we have published elsewhere (Klein & Belz, 2014; Klein, Willenborg & Schweiger, 2014a). Patients with episodic depression were patients meeting DSM-5 criteria for a depressive episode while not meeting criteria for persistent depressive disorder. Healthy controls were participants who did not show evidence of previous psychiatric disorders

on the DSM-5 self-rated cross-cutting symptom measure (adult) (American Psychiatric Association, 2013) and a brief clinical interview.

Informed written consent was obtained from all participants. The study was conducted in accordance with the Declaration of Helsinki. The ethics committee at Lübeck University approved the study. All participants received the questionnaires listed below prior to the actual assessment. The assessment visit started with the application of the Interactive Test of Interpersonal Behavior. A trained study assistant was available for questions while the patients completed the test. This was followed by a brief diagnostic interview to rule out psychiatric morbidity in the controls and ascertain group membership (episodic versus chronic depression) in patients. This sequence was established to ascertain that the study assistant would be ignorant to the diagnostic status while administering the ITIB to avoid unconsciously biasing the participants. The assessment also included the Wortschatztest (WST), which was used as a measure of verbal IQ (Schmidt & Metzler, 1992). In total, the assessment visit took between 30 and 40 minutes.

Instruments

Interactive Test of Interpersonal Behavior (ITIB). The Interactive Test of Interpersonal Behavior (ITIB) was developed by Kensche, with support from all the authors. It is based on situational analysis, a central technique in the Cognitive Behavioral Analysis System of Psychotherapy (CBASP) (McCullough, 2000). Through situational analysis, patients learn interpersonal behavior that is in service of a realistic and attainable desired outcome. Specifically, they learn to openly express their needs and feelings in a friendly way wherever that is possible. The goal of the ITIB is to assess how often and how quickly participants are able to do that.

Interpersonal scenarios from different areas of social life were constructed for the ITIB. Each scenario starts with the presentation of an interpersonal dilemma (i.e., discussing an overdue project with an angry boss). Participants can chose between three answers: one that is rather hostile and not goal-oriented (i.e., talking back to the boss "do not use that language with me"), one that is rather friendly and goal-oriented (i.e., "I can understand your anger, I have good reasons to be late and would like to find a solution with you") and one that is still friendly but not goal-oriented (i.e., "I understand your anger but the delay it is not my fault"). Participants are instructed to honestly choose the option that best describes how they would react in the given situation.

If they chose the friendly but not goal-oriented answer, they will be taken to the next step of the same scenario (i.e., boss answers "I really needed you to complete this assignment on time"). Again, participants can chose between a hostile/not goal-oriented, a friendly/not goal-oriented and a friendly/goal-oriented response. The scenarios can therefore be understood as simulated dialogues. This contributes to the ecological validity of the test as subjects could be said to be "interacting" with the imagined person that the test system presents.

Once participants chose either the hostile/not goal-oriented or the friendly/goal-oriented response, the scenario is over and the participant is taken to the next scenario. The total number of possible steps for each scenario is six, in the final step the participant has to choose between a hostile/not goal-oriented or a friendly/goal-oriented response. The scenarios were discussed with experts in the field and the final set included six scenarios (see Table 1).

To increase ease of administration and usability of the interactive test we decided to implement it as an application for a tablet computer. Usability was defined as the extent to which the test system can be used by the pre-specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use (International Organization for Standardization (ISO), 2010; International Organization for Standardization (ISO), 1998). Development of this application was performed by Mentler, Stoislow and Hüppe according to ISO 9241-210. This standard refers to human-centered design of interactive systems and is part of the multi-part standard from the International Organization for Standardization (ISO) covering ergonomics of human-system interactions. It describes an iterative and participatory design process based on explicit understanding of users, tasks and environments (the context of use). All

Table 1. *Item analysis and factor analysis analysis for the Interactive Test of Interpersonal Behaviour (ITIB)*

| | | | | | Rotated factor loadings | |
| | | | | | 1: Explaining one's mistakes | 2: Getting one's needs met |
Item		Mean (SD)	r_{tot}	Cronbach's alpha if item was deleted		
1	Dinner with partner	8.58 (± 4.73)	0.317	0.622	0.318	0.390
2	Arrangement with friend	9.82 (± 3.91)	0.381	0.594	**0.930**	−0.043
3	Conversation with colleague	9.33 (± 3.64)	0.501	0.554	0.292	**0.744**
4	Conversation with boss	9.76 (± 4.06)	0.524	0.538	**0.891**	0.197
5	At the cinema with a friend	9.87 (± 4.01)	0.212	0.652	−0.064	**0.638**
6	Arrangement with partner	9.47 (± 4.36)	0.330	0.613	0.040	**0.744**
	Eigenvalues				2.251	1.307
	% of variance				37.513	21.785
	Cronbach's alpha				0.834	0.561

Notes: r_{tot} = corrected item-total correlation, SD = standard deviation. Items loading on one of the factors above the predefined threshold are printed in bold.

the authors contributed to the development to insure multidisciplinary input.

The application starts with a screen introducing the test. The six scenarios are presented in random order. Also the location of the three possible responses on the screen (bottom left, bottom middle, bottom right) is randomized to ascertain that participants carefully consider each option rather than repeatedly opting for the same response based on the screen-location of their prior response. All responses are recorded in a log-file that can be accessed by the researchers or clinicians applying the test. No feedback is given to the test participants.

The test score was developed to reflect both frequency and immediacy of friendly/goal-oriented responses in one score. To this end, participants can receive between zero and 12 points in each scenario. Twelve points are awarded if the patient immediately choses the friendly/goal-oriented response, 11 points are awarded if the patient choses this response at the second step and so forth. Zero points are awarded if the patient immediately choses the hostile/not goal-oriented response, one if he chooses this response at the second step and so forth. The total score is the sum of the scores of the individual scenarios and can range from 0 to 72. Higher scores reflect the ability to frequently choose a friendly/goal-oriented response over an hostile/not goal-oriented one at the earliest possible moment in the interaction.

Lübeck Questionnaire for Recording Preoparational Thinking (LQPT). The LQPT is a standardized self-assessment instrument for recording the specific cognitive psychopathology of chronically depressed patients (Kuhnen *et al.*, 2011). It contains 22 items where participants are confronted with difficult situations and are required to choose between two response options: one reflecting a high and the other a low level of preoperational thinking. The different characteristics of preoperational thinking (snapshot perspective, prelogical thinking, egocentrism, lack of perceived functionality, lack of empathy) are covered by the test. A low total score indicates a high level of preoperational thinking. The LQPT has been shown to be a reliable (Cronbach's alpha 0.901) and valid instrument (Kuhnen *et al.*, 2011).

Self-rated Inventory of Depressive Symptomatology (IDS-SR). The IDS-SR assesses the severity and frequency of depressive symptoms. It contains 32 items that are scored from 0 to 3, with higher scores reflecting greater psychopathology. Total IDS scores range from 0 to 96 (Rush, Giles, Schlesser, Fulton, Weissenburger & Burns, 1986; Rush, Gullion, Basco, Jarrett & Trivedi, 1996), with scores of 13 or lower indicative of no depression, scores from 14 to 25 indicating mild depression, 26 to 38 indicating moderate depression, 39 to 48 reflecting severe depression, and total scores greater than 48 indicating very severe depression.

DSM-5 self-rated cross-cutting symptom measure (adult). This measure is a self-rated measure that assesses mental health domains that are important across psychiatric diagnoses (American Psychiatric Association, 2013). It

may be used to track changes in the individual's symptoms over time. It contains 23 questions that assess 13 psychiatric domains including depression, anger, mania, anxiety, somatic symptoms, suicidal ideation, psychosis, sleep problems, memory, repetitive thoughts and behaviors, dissociation, personality functioning, and substance use. Each item inquires about how much (or how often) the individual has been bothered by the specific symptom during the past 2 weeks. As this measure had not been officially translated at the onset of our study we translated this measure for our purposes. Test-retest reliabilities of the cross-cutting symptom items have been shown to be good to excellent (Narrow, Clarke, Kuramoto *et al.*, 2013).

Childhood Trauma Questionnaire (CTQ). The CTQ consists of 28 self-report items that assess childhood maltreatment before the age of 18. Patients rate the accuracy of statements about childhood experiences are on a five-point Likert scale (1, never true; 5, very often true). The CTQ consists of five subscales: emotional abuse, physical abuse, sexual abuse, emotional neglect and physical neglect. The scores for these subscales range from 5 (low level of childhood maltreatment) to 25 (high level of childhood maltreatment) and provide a quantitative index of the severity of abuse.

Inventory of Interpersonal Problems (IIP). The short version (64 items) of the IIP was used to assess the severity of interpersonal problems (Brahler, Horowitz, Kordy, Schumacher & Strauss, 1999; Horowitz, Rosenberg, Baer, Ureno & Villasenor, 1988). The IIP consists of eight subscales reflecting different octants in the interpersonal space. Each construct is measured by eight items, each rated on a five-point Likert scale (0 = not at all; 4 = very much). For the analyses presented here we only used the sum score of all the eight subscales that describes the extent of interpersonal difficulties.

Competence and Control Beliefs Questionnaire (FKK). The FKK was used to assess expectations regarding ones competence and contingencies of ones behavior across different classes of behavior (Krampen, 1991). The questionnaire is composed of 32 items that are based on Levenson's "Internal, Powerful Others and Chances" scale (IPC-scale) (Levenson, 1973). It consists of four subscales that can be summarized into two scales: one reflecting internal control beliefs (FKK-SKI) and one reflecting external control beliefs (FKK-PC). Higher scores on each subscale reflect high internality and high externality respectively.

System Usability Scale (SUS). The SUS is a reliable and valid measure of perceived usability of a computer system (Bangor, Kortum & Miller, 2009; Brooke, 2013). The questionnaire consists of 10 items with five response options varying from "strongly agree" to "strongly disagree." Five items are positively respectively negatively worded. Score contribution is scale position minus 1 respectively 5 minus scale position. The overall value is calculated by multiplying the sum of the single scores by 2.5. This leads to a range from 0 to 100 which must not be interpreted as percentage of

usability. Rather, SUS scores should be interpreted based on percentiles and adjective ratings. For example, a SUS score of 71.4 respectively 85.5 correlates to "good" respectively "excellent" ratings by individuals (Bangor *et al.*, 2009). The SUS has been found to be a reliable (Cronbach's alpha 0.911) and valid instrument (Bangor, Kortum & Miller, 2008).

Data analyses

Statistical analyses were conducted using SPSS 22.0 (SPSS, Chicago, IL). All statistical tests were evaluated as two-sided tests with significance levels set at $p \leq 0.05$. No missing values were substituted.

Reliability

Cronbach's alpha and Guttman's lambda 2 were calculated as measures of reliability. As Cronbach's alpha can give falsely low estimates of reliability, reporting Guttman's lambda 2 in addition to Cronbach's alpha has been suggested as a measure to improve reliability estimation practice (Sijtsma, 2009). By convention, Cronbach's alpha of above 0.70 was regarded as acceptable to detect individual differences (Tavakol & Dennick, 2011). Mean-inter-item correlation was calculated as an index of homogeneity of items. Mean-inter-item correlations of between 0.20 and 0.40 were regarded as acceptable (Bühner, 2011).

Item analysis

For item analysis of the individual ITIB items several analyses were performed. Means were expected to be around 6, else the responses to this item would be predominantly at one end of the scale. Standard deviation was calculated as an index of the extent of interindividual differences in response to each item. Corrected item-total correlations were calculated as an index of discrimination, where individuals who possess the same attribute respond to all the items in in a similar way. The corrected item-total correlations were expected to be about 0.3 (Coaley, 2009; Nunnally & Bernstein, 1994). Also it was determined for each ITIB item, what Cronbach's alpha for the entire scale would be if the item was deleted. Cronbach's alpha if an item was deleted should be close to overall Cronbach's alpha to signify that the item does not negatively impacting overall reliability (Coaley, 2009).

Factor analysis

A principal component analysis (PCA) was performed on the ITIB with orthogonal rotation (varimax). Factors with eigenvalue greater than 1.0 were extracted. The Kaiser-Meyer-Olkin-coefficient (KMO) verified the sampling adequacy for the analysis, KMO = 0.565 (Field, 2009) and all KMO values for individual items are ≥ 0.518 which is above the acceptable limit 0.5 (Field, 2009). Bartlett's test of sphericity chi^2 (15) = 58.805, $p < 0.001$ indicated that correlations between items were sufficiently large for PCA. An item was considered to be part of a factor if its factor loading on that factor was above 0.55 as these loadings are considered to be good or better (Tabachnik & Fidell, 2013).

Validity

To determine concurrent and divergent validity, Pearson correlations between the total scores of the ITIB, the LQPT, the IIP, the FKK, the IDS, the CTQ and the WST were computed. Correlations between two measures of 0.55 and higher were considered to be an excellent indicator of a measuring the same construct. Correlations between 0.45 and 0.54 were considered to be a good indicator of measuring the same construct. Correlations of under 0.20 are indicative of measuring a different construct (Coaley, 2009; McDowell, 2006).

We also determined discriminant validity by comparison of the three groups (chronic depression, episodic depression, healthy controls) using one-way independent ANOVAs. Based on previous research (Kuhnen

et al., 2011; Constantino *et al.*, 2008), we expected that the total scores for the ITIB and the LQPT would be highest for healthy controls, medium for patients with episodic depression and lowest for patients with chronic depression. As a sensitivity analysis, we also calculated an ANCOVA with the ITIB and the LQPT as dependent variable to control for age.

Finally we computed a multinomial regression analysis to analyze associations with diagnostic status. The ITIB and the LQPT were entered as independent variables with diagnostic status as the dependent variable. Predictors were included stepwise, the analysis was adjusted for age.

RESULTS

Sample

A total of 45 participants were enrolled. For demographic and clinical details of the participants please refer to Table 2. Briefly, the majority of participants in all groups was female but patients with episodic depression were younger and less often married than those with chronic depression and healthy controls. With regards to comorbidity, the three groups differed on measures of anxiety, somatization, sleep disturbance, obsessive-compulsiveness, personality functioning and substance use. On post hoc testing there were no significant differences between patients with episodic and chronic depression.

Usability

Mean scores (SD) for SUS were 91.73 (9.38). Therefore, the application has been rated "good" respectively "excellent" by most of the participants and seems to be usable in the specified context of use. Participants were also observed regarding human-computer interaction aspects. While some users were afraid of not being able to get along with the tablet PC prior to usage, they had no or only minor problems while using it. Most of the problems concerned layout of the start screen, text lengths and finger input, for example, with long fingernails. All in all, the subjective impressions and direct quotes of patients reflected the SUS score. Some participants also found the instruction in the start screen confusing.

Reliability

Cronbach's alpha, Guttman's lambda 2 and mean-inter-item consistency for the ITIB were 0.649, 0.664 and 0.24 (SD 0.057) respectively. Reliability estimates were outside the pre-defined limits of acceptability for Cronbach's alpha and within the predefined limit for mean-inter-item consistency. A scale which has an index of reliability outside these limits is still suitable for group comparisons (Lienert & Raatz, 1998).

Item analysis

The results of the item and factor analysis are presented in Table 1. Regarding the item analysis, our results indicate that the responses given by participants in our study are at the higher end of the scale.

Corrected item-total correlations between 0.317 and 0.524 indicate that all ITIB-items except for item 5 ("at the cinema with a friend": 0.212) discriminate well between different levels of impairment in interpersonal behavior. Conversely, the results for the calculation of Cronbach's alpha if an item is deleted do not

Table 2. *Demographics and clinical characteristics*

	Chronic depression N = 15	Episodic depression N = 15	Healthy controls N = 15	P
Female (%)	11 (73.3%)	11 (73.3%)	11 (73.3%)	1.00
Age (yrs), mean (SD)	49.8 (± 6.12)	35.6 (± 14.49)	42.0 (± 12.46)	0.002
Education (%)				0.109
No degree	1 (6.7%)	0	0	
Lower secondary	3 (20.0%)	4 (26.7%)	0	
Middle secondary	7 (46.7%)	4 (26.7%)	11 (73.3%)	
Higher secondary	2 (13.3%)	6 (40.0%)	4 (26.7%)	
University	2 (13.3)	1 (6.7%)	0	
Marital status (%)				0.361
Not married	3 (20%)	8 (53.3%)	4 (26.7%)	
Married	8 (53.3%)	4 (26.7%)	7 (46.7%)	
Divorced	4 (26.7%)	3 (20.0%)	4 (26.7%)	
Employment status (%)				0.027
Fulltime	4 (26.7%)	6 (40.0%)	6 (40.0%)	
Part-time	3 (20.0%)	0	8 (53.3%)	
Disability pension	4 (26.7%)	4 (26.7%)	0	
Other	4 (26.7%)	5 (30.0%)	1 (6.7%)	
Depression severity, mean (SD)				
IDS-SR	41.3 (± 16.82)	47.1 (± 14.25)	10.8 (± 5.23)	0.008
DSM-5 symptom measure (adult) sum score, mean (SD)				
Depression	5.53 (± 2.26)	6.00 (± 1.77)	1.07 (± 1.16)	<0.001
Anxiety	5.60 (± 3.85)	6.87 (± 3.12)	0.60 (± 0.74)	<0.001
Somatic symptoms	2.67 (± 2.90)	3.00 (± 2.75)	0.80 (± 1.57)	0.041
Sleep problems	3.20 (± 0.86)	2.73 (± 1.22)	0.67 (± 0.72)	<0.001
Repetetive thougths and behaviours	3.00 (± 2.14)	3.53 (± 2.39)	0.33 (± 0.49)	<0.001
Personality functioning	3.67 (± 3.12)	4.60 (± 2.61)	0.27 (± 0.59)	<0.001
Substance use	3.20 (± 3.52)	1.87 (± 1.77)	1.53 (± 1.60)	0.158

Notes: DSM-5 = 5th edition of the Diagnostic and Statistical Manual of the American Psychiatric Association, IDS-SR = Self-Rated Inventory of Depressive Symptomatology, SD = standard deviation, yrs = years.

indicate that one of the items in the scale negatively affects reliability of the total scale.

Factor analysis

The principal component analysis resulted in the extraction of two components with eigenvalues greater than Kaiser's criterion of 1. In combination, these two factors explained 59.30% of the variance. The scree plot showed inflections that justified retaining two components. Given the convergence of the scree plot and Kaiser's criterion on two components, these were retained in the final analysis. The factors were labeled "explaining one's mistakes" and "getting one's needs met." One item (dinner with partner) did not meet this criterion. This item loaded on factor two above a less conservative threshold of 0.32 (Tabachnik & Fidell, 2013).

Validity

Intercorrelations between measures are presented in Table 3. The results presented there indicate that the ITIB has good concurrent validity with the preoperational thinking construct measured by the LQPT and excellent concurrent validity with the interpersonal problems captured by the IIP. Among the subscales of the IIP, the ITIB has correlations indicating good concurrent validity with the following subscales: being too vindictive (BC –0.456), too introverted (FG –0.485), too exploitable (JK –0.513) but also overly nurturing (LM –0.548). The ITIB correlates less strongly with the degree of depression measured by the IDS-SR than the LQPT. The ITIB does not correlate significantly with measures of early trauma and verbal intelligence.

Results for discriminant validity can be found in Tables 4 and 5. In summary, scores for all the scales including the ITIB but except

Table 3. *Correlations of Lübeck Questionnaire for recording Preoperational Thinking (LQPT) and the Interactive Test of Interpersonal Behaviour (ITIB) sum score with scores of other measures that measure a similar construct (IIP) and other constructs (FKK, CTQ, IDS-SR, WST)*

	LQPT	ITIB	IIP	FKK-SKI	FKK-PC	IDS-SR	CTQ	WST
LQPT		*0.524***	**−0.753***	**0.679***	**−0.544***	**0.608***	0.344*	0.168
ITIB	*0.524***		**−0.568***	0.399**	−0.326*	0.391**	−0.094	−0.065

Notes: Correlations between two measures of 0.55 and higher were considered to be an excellent indicator of a measuring the same construct and are printed in bold. Correlations between 0.45 and 0.54 were considered to be a good indicator of measuring the same construct and are printed in italic. ** significant at p ≤ 0.05, ** significant at p ≤ 0.01. CTQ: Childhood Trauma Questionnaire, FKK: Competence and Control Beliefs Questionnaire, FKK-SKI: FKK-subscale measuring self-efficiency and internality, FKK-PC: FKK-subscale measuring externality, IDS-SR = Self-Rated Inventory of Depressive Symptomatology, IIP: Inventory of Interpersonal Problems, WST: measure of verbal IQ.

Table 4. *Comparison of patients with regard to diagnosis. Means and standard deviations*

	Chronic depression N = 15	Episodic depression N = 15	Healthy controls N = 15	ANOVA F(2,42)		Post Hoc Test	
				F	p	CD vs. HC	CD vs. ED
LQPT	14.33 (± 5.42)	15.93 (± 5.86)	19.40 (± 2.69)	4.25	0.021	0.02	n.s.
ITIB	47.67 (± 16.61)	58.00 (± 14.74)	64.73 (± 6.43)	6.22	0.004	0.003	n.s.
ITIB "explaining one's mistakes"	15.47 (± 8.77)	19.87 (± 7.83)	23.40 (± 0.91)	5.18	0.010	0.008	n.s.
ITIB "getting one's needs met"	23.93 (± 10.67)	30.40 (± 7.94)	31.67 (± 5.12)	3.81	0.030	0.041	n.s.
CTQ total	66.60 (± 24.33)	58.73 (± 23.50)	39.60 (± 11.98)	6.74	0.003	0.003	n.s.
Emotional abuse	15.67 (± 5.74)	12.87 (± 6.21)	7.73 (± 4.33)	8.07	0.001	0.001	n.s.
Physical abuse	9.33 (± 5.65)	6.53 (± 2.90)	5.87 (± 1.51)	3.57	0.037	0.047	n.s.
Sexual abuse	8.27 (± 4.54)	8.73 (± 6.67)	5.47 (± 1.81)	2.05	n.s.	n.s.	n.s.
Emotional neglect	15.73 (± 6.29)	13.47 (± 6.31)	8.93 (± 4.43)	5.45	0.008	0.007	n.s.
Physical neglect	9.47 (± 3.39)	8.53 (± 4.00)	6.13 (± 1.60)	4.42	0.018	0.019	n.s.
IIP	15.02 (± 3.94)	14.18 (± 5.35)	8.52 (± 3.63)	9.83	<0.001	0.001	n.s.
FKK-SKI	53.00 (± 7.89)	53.93 (± 9.44)	67.13 (± 8.25)	12.82	<0.001	<0.001	n.s.
FKK-PC	55.00 (± 18.71)	55.40 (± 12.38)	44.13 (± 10.18)	3.03	0.059	n.s.	n.s.

Notes: CTQ: Childhood Trauma Questionnaire, FKK: Competence and Control Beliefs Questionnaire, FKK-SKI: FKK-subscale measuring self-efficiency and internality, FKK-PC: FKK-subscale measuring externality, IDS-SR = Self-Rated Inventory of Depressive Symptomatology, IIP: Inventory of Interpersonal Problems, ITIB: Interactive Test of Interpersonal Behaviour, LQPT: Lübeck Questionnaire for recording Preoperational Thinking (LQPT).

Table 5. *Associations with diagnosis (multinomial logistic regression analysis, adjusted for age)*

	B	S.E.	95% CI for Odds Ratio			p
			Lower	OR	Upper	
Predictors ITIB and LQPT included with stepwise method. R² = 0.269 (Nagelkerke). Model chi²(2) = 12.286, *p* = 0.002.						
Outcome: chronic depression vs. episodic depression.						
ITIB	0.049	0.029	0.993	1.051	1.112	0.088
Outcome: chronic depression vs. healthy controls.						
ITIB	0.118	0.043	1.035	1.125	1.224	0.006
Predictors FKK-SKI and FKK-PC included with stepwise method. R² = 0.425 (Nagelkerke). Model chi²(2) = 21.36, *p* < 0.001.						
Outcome: chronic depression vs. episodic depression.						
FKK-SKI	0.013	0.043	0.931	1.013	1.103	n.s.
Outcome: chronic depression vs. healthy controls.						
FKK-SKI	0.217	0.071	1.083	1.243	1.427	0.002

Notes: FKK: Competence and Control Beliefs Questionnaire, FKK-SKI: FKK-subscale measuring self-efficiency and internality, FKK-PC: FKK-subscale measuring externality, ITIB: Interactive Test of Interpersonal Behaviour, LQPT: Lübeck Questionnaire for recording Preoperational Thinking (LQPT).

for the FKK-PC subscale significantly differed between the diagnostic groups ($p < 0.037$). On post-hoc testing, there were significant differences between patients with chronic depression and healthy controls on all these scales. Differences between the chronic and the episodic depression group were not significant though. On sensitivity analysis, the result for the ITIB (ANCOVA $F(2,41) = 5.29$, $p = 0.009$) and the LQPT (ANCOVA $F(2,41) = 5.51$, $p = 0.008$) was stable after controlling for age.

As the sample size was small we also calculated between groups effect sizes (Cohen's d for chronic vs. episodic depression) to further explore the non-significant differences between these two diagnostic groups (Cohen, 1992). These were below 0.2 for the IIP (0.18), FKK-SKI (0.12) and FKK-PC (0.02), between 0.2 and 0.5 (small to medium) for the CTQ 0.32) and the LQPT .28) and between.5 and.8 (medium to large) only for the ITIB .66). A further exploration of the ITIB results shows that the range of total scores is overlapping (CD: 2–69, ED:

32–72, HC: 54–71) but the median of the CD group (51) is outside the range of the HC group.

On our final analysis using multinomial regression (Table 5), the ITIB emerged as the best predictor of diagnostic status. As the accuracy of classification was not improved with further addition of the LQPT, its score was not included in the final model. There was a significant association between the ITIB and diagnostic status for the comparison between healthy controls and chronic depressives ($p = 0.006$) and a trend for the comparison between episodic and chronic depressives ($p = 0.088$).

DISCUSSION

In this paper we describe the development and psychometric properties of the Interactive Test of Interpersonal Behavior (ITIB). The test was developed to test the hypothesis that patients with chronic depression have deficits in interpersonal behavior. We

found that the test is highly acceptable. The reliability estimates are not adequate yet for using the test to detect individual differences. In its current form, the ITIB should therefore probably not be used as a pre-post measure. The instrument is reliable enough however to make group comparisons. The estimates of concurrent validity were within the pre-determined limits.

Using the newly developed ITIB we found that compared to patients with episodic depression and healthy controls, patients with chronic depression had more deficits in interpersonal behavior. On a trend level we also found that the ITIB is associated with group membership (CD vs. ED). This could indicate that interpersonal deficits are indeed a defining characteristic of CD but not the Competence and Control Beliefs Questionnaire (FKK). The findings regarding locus of control support this notion. In keeping with earlier research (Benassi, Sweeney & Dufour, 1988), we did find a difference between patients with depression and healthy subjects on a measure of locus of control (the FKK). Interestingly, patients with CD and ED did not differ on this scale (see Tables 4 and 5). This might imply that patients with CD do not just suffer from more severe general depressive psychopathology but from a specific pathology associated with chronicity. This specific psychopathology may be particularly pronounced with regards to getting one's needs met in interpersonal situations, one of the two factors that emerged from factor analysis of the ITIB. Here, patients with ED did not differ from HC while patients with CD differed markedly.

All of these interpretations have to be made with great caution though. Due to the very small sample size and the pilot character of this study, these results are very preliminary and might be underpowered. For instance, on a descriptive level we could also replicate earlier findings with regards to more severe early trauma in patients with CD compared to patients with ED (Wiersma *et al.*, 2009; van Randenborgh, Huffmeier, Victor, Klocke, Borlinghaus & Pawelzik, 2012). This difference failed to reach statistical significance though due to our small sample size.

The acceptability ratings, the reliability estimates and the item analysis imply that we will need to make minor adjustments to the ITIB before we can use it in a larger setting. We found that item 5 ("at the cinema with a friend") had a lower corrected item-total correlation than the other items and may therefore discriminate less well between individuals with a similar pattern of interpersonal behavior. On closer inspection of this item we found that the answers participants could choose from did not fully reflect the underlying construct (differentiating hostile/not goal oriented vs. friendly/not goal-oriented vs. friendly/goal-oriented responses). This item may therefore have to be rewritten. Some minor improvements may also be necessary with regards to the instructions in the start screen.

Finally it has to be acknowledged that the different tests previously used to assess interpersonal psychopathology of chronic depression measure different constructs. Specifically we found that compared to the correlation with the Inventory of Interpersonal Problems (IIP) there was a slightly lower correlation of the ITIB with our previously developed assessment of preoperational thinking, the LQPT. That is probably due to the fact that the LQPT also measures preoperational thinking in non-interpersonal situations (i.e., losing one's passport). A comparison of the ITIB with third-party assessments of the patients behavior would be a valuable contribution to the literature as it would

substantiate that the test adequately reflects actual interpersonal behavior. This could be achieved by comparing ITIB ratings with ratings on the Interpersonal Message Inventory (IMI) that can be collected from patients (Constantino *et al.*, 2008; Brakemeier *et al.*, 2015) or significant others (Grosse, Altenstein, Ansell, Schneider & Caspar, 2012).

In summary we found preliminary evidence that the ITIB is accepted by participants and has mostly adequate psychometric properties. Also we found very preliminary evidence that the test differentiates between patients with chronic depression and episodic depression better than other assessments and therefore might actually measure psychopathology that is unique to chronic depression. We have now slightly modified the ITIB and test it in a larger sample of participants. Further research is needed to ascertain whether the interpersonal deficits found with this test are indeed unique to chronic depression or if they can also be found in other chronic psychiatric disorders.

REFERENCES

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th edn). Arlington, VA: American Psychiatric Association.

Bangor, A., Kortum, P. & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4, 114–123.

Bangor, A., Kortum, P. T. & Miller, J. T. (2008). An empirical evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24, 574–594.

Benassi, V. A., Sweeney, P. D. & Dufour, C. L. (1988). Is there a relation between locus of control orientation and depression? *Journal of Abnormal Psychology*, 97, 357–367.

Brahler, E., Horowitz, L. M., Kordy, H., Schumacher, J. & Strauss, B. (1999). Zur Validierung des Inventars zur Erfassung Interpersonaler Probleme (IIP)-Ergebnisse einer Repräsentativbefragung in Ost-und Westdeutschland. *Psychotherapie Psychosomatik Medizinische Psychologie*, 49, 422–431.

Brakemeier, E. L., Radtke, M., Engel, V., Zimmermann, J., Tuschen-Caffier, B. & Hautzinger, M. *et al.* (2015). Overcoming treatment resistance in chronic depression: A pilot study on outcome and feasibility of the cognitive behavioral analysis system of psychotherapy as an inpatient treatment program. *Psychotherapy and Psychosomatics*, 84, 51–56.

Brooke, J. (2013). SUS. A retrospective. *Journal of Usability Studies*, 8, 29–40.

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion.* München: Pearson Studium.

Coaley, K. (2009). *An introduction to psychological assessment and psychometrics.* London: Sage.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Sciences*, 1, 98–101.

Constantino, M. J., Laws, H. B., Arnow, B. A., Klein, D. N., Rothbaum, B. O. & Manber, R. (2012). The relation between changes in patients' interpersonal impact messages and outcome in treatment for chronic depression. *Journal of Consulting and Clinical Psychology*, 80, 354–364.

Constantino, M. J., Manber, R., DeGeorge, J., McBride, C., Ravitz, P., Zuroff, D. C. *et al.* (2008). Interpersonal styles of chronically

depressed outpatients: Profiles and therapeutic change. *Psychotherapy: Theory. Research, Practice, Training*, 45, 491–506.

Conway, C. C., Hammen, C. & Brennan, P. A. (2012). Expanding stress generation theory: Test of a transdiagnostic model. *Journal of Abnormal Psychology*, 121, 754–766.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J. & Brand, M. et al. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders*, 36, 623–636.

Field, A. (2009). *Discovering Statistics using SPSS* (3rd edn). London: Sage.

Grosse, H. M., Altenstein, D., Ansell, E., Schneider, C. & Caspar, F. (2012). Impact messages of depressed outpatients as perceived by their significant others: Profiles, therapeutic change, and relationship to outcome. *Journal of Clinical Psychology*, 68, 319–333.

Hammen, C. (2005). Stress and depression. *Annual Review of Clinical Psychology*, 1, 293–319.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G. & Villasenor, V. S. (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, 56, 885–892.

International Organization for Standardization (ISO). (2010). *ISO 9241: Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems.*

International Organization for Standardization (ISO). (1998). *ISO 9241-11: Ergonomic requirements for office work with visual display terminals - Part 11: Guidance on usability.*

Keller, M. B., McCullough, J. P., Klein, D. N., Arnow, B., Dunner, D. L., Gelenberg, A. J. et al. (2000). A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *New England Journal of Medicine*, 342, 1462–1470.

Kiesler, D. J. & Schmidt, J. A. (1993). *The Impact Message Inventory: Form IIA Octant Scale Version*. Redwood City, CA: Mind Garden.

Klein, J. P., Willenborg, B. & Schweiger, U. (2014a). Diagnostik affektiver Störungen im DSM-5: Neuerungen und Kontroversen. *Fortschritte der Neurologie und Psychiatrie*, 82, 655–664.

Klein, J. P., Becker, B., Hurlemann, R., Scheibe, C., Colla, M. & Heuser, I. (2014b). Effect of specific psychotherapy for chronic depression on neural responses to emotional faces. *Journal of Affective Disorders*, 166, 93–97.

Klein, J. P. & Belz, M. (2014). *Psychotherapie chronischer depression. Praxisleitfaden CBASP*. Göttingen: Hogrefe.

Klein, J. P., Roniger, A., Schweiger, U., Späth, C. & Brodbeck, J. (in press). The association of childhood trauma and personality disorder with adult chronic depression. A cross-sectional study in depressed outpatients. *The Journal of Clinical Psychiatry*.

Klein, J. P., Steinlechner, S., Sipos, V. & Schweiger, U. (2011). Psychotherapie chronischer depression nach dem Cognitive Behavioral Analysis System of Psychotherapy (CBASP)-Konzept. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 61, 526–535.

Krampen, G. (1991). *Fragebogen zu Kompetenz -und Kontrollüberzeugungen*. Götingen: Hogrefe – Verlag für Psychologie.

Kuhnen, T., Knappke, F., Otto, T., Friedrich, S., Klein, J. P., Kahl, K. G. et al. (2011). Chronic depression: Development and evaluation of the Lübeck Questionnaire for Recording Preoperational Thinking (LQPT). *BMC Psychiatry*, 11, 199.

Levenson, H. (1973). Multidimensional locus of control in psychiatric patients. *Journal of Consulting and Clinical Psychology*, 41, 397–404.

Lienert, G. & Raatz, U. (1998). *Testaufbau und testanalyse* (6th edn). Weinheim: Psychologie Verlags Union.

Manber, R., Arnow, B., Blasey, C., Vivian, D., McCullough, J. P. & Blalock, J. A. et al. (2003). Patient's therapeutic skill acquisition and response to psychotherapy, alone or in combination with medication. *Psychological Medicine*, 33, 693–702.

McCullough, J. P. (2000). *Treatment for chronic depression: Cognitive behavioral analysis system of psychotherapy (CBASP)*. New York: The Guilford Press.

McCullough, J. P., Lord, B. D., Conley, K. A. & Martin, A. M. (2010). A method for conducting intensive psychological studies with early-onset chronically depressed patients. *American Journal of Psychotherapy*, 64, 317–337.

McDowell, I. (2006). *Measuring health: A guide to rating scales and questionnaires*. New York: Oxford University Press.

Murphy, J. A. & Byrne, G. J. (2012). Prevalence and correlates of the proposed DSM-5 diagnosis of Chronic Depressive Disorder. *Journal of Affective Disorders*, 139, 172–180.

Narrow, W. E., Clarke, D. E., Kuramoto, S. J., Kraemer, H. C., Kupfer, D. J., Greiner, L. et al. (2013). DSM-5 field trials in the United States and Canada. Part III: Development and reliability testing of a cross-cutting symptom assessment for DSM-5. *The American Journal of Psychiatry*, 170, 71–82.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (3rd edn). New York: McGraw-Hill.

Rush, A. J., Giles, D. E., Schlesser, M. A., Fulton, C. L., Weissenburger, J. & Burns, C. (1986). The Inventory for Depressive Symptomatology (IDS): Preliminary findings. *Psychiatry Research*, 18, 65–87.

Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B. & Trivedi, M. H. (1996). The Inventory of Depressive Symptomatology (IDS): Psychometric properties. *Psychological Medicine*, 26, 477–486.

Satyanarayana, S., Enns, M. W., Cox, B. J. & Sareen, J. (2009). Prevalence and correlates of chronic depression in the Canadian community health survey: Mental health and well-being. *Canadian Journal of Psychiatry*, 54, 389–398.

Schmidt, K.-H. & Metzler, P. (1992). *Wortschatztest*. Weinheim: Beltz Test.

Schramm, E., Zobel, I., Dykierek, P., Kech, S., Brakemeier, E. L., Kulz, A. et al. (2011). Cognitive behavioral analysis system of psychotherapy versus interpersonal psychotherapy for early-onset chronic depression: A randomized pilot study. *Journal of Affective Disorders*, 129, 109–116.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's Alpha. *Psychometrika*, 74, 107–120.

Swan, J. S., Macvicar, R., Christmas, D., Durham, R., Rauchhaus, P. & McCullough, J. P., Jr. et al. (2014). Cognitive Behavioural Analysis System of Psychotherapy (CBASP) for chronic depression: Clinical characteristics and six month clinical outcomes in an open case series. *Journal of Affective Disorders*, 152–154, 268–276.

Tabachnik, B. G. & Fidell, L. S. (2013). *Using multivariate statistics*. Boston, MA: Pearson Education.

Tavakol, M. & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55.

van Randenborgh, A., Huffmeier, J., Victor, D., Klocke, K., Borlinghaus, J. & Pawelzik, M. (2012). Contrasting chronic with episodic depression: An analysis of distorted socio-emotional information processing in chronic depression. *Journal of Affective Disorders*, 141, 177–184.

Wiersma, J. E., Hovens, J. G., van Oppen, P., Giltay, E. J., van Schaik, D. J., Beekman, A. T. et al. (2009). The importance of childhood trauma and childhood life events for chronicity of depression in adults. *The Journal of Clinical Psychiatry*, 70, 983–989.

Wiersma, J. E., van Schaik, D. J., Hoogendorn, A. W., Dekker, J. J., Van, H. L., Schoevers, R. A. et al. (2014). The effectiveness of the Cognitive Behavioral Analysis System of Psychotherapy for Chronic Depression: A randomized controlled trial. *Psychotherapy and Psychosomatics*, 83, 263–269.

Wilbertz, G., Brakemeier, E. L., Zobel, I., Harter, M. & Schramm, E. (2010). Exploring preoperational features in chronic depression. *Journal of Affective Disorders*, 124, 262–269.

Wolkenstein, L., Schonenberg, M., Schirm, E. & Hautzinger, M. (2011). I can see what you feel, but I can't deal with it: Impaired theory of mind in depression. *Journal of Affective Disorders*, 132, 104–111.

Zobel, I., Werden, D., Linster, H., Dykierek, P., Drieling, T., Berger, M. et al. (2010). Theory of mind deficits in chronically depressed patients. *Depression and Anxiety*, 27, 821–828.