

The Berkeley Puppet Interview: A Screening Instrument for Measuring Psychopathology in Young Children

Lisanne L. Stone · Carlijn van Daal · Marloes van der Maten ·
Rutger C. M. E. Engels · Jan M. A. M. Janssens · Roy Otten

Published online: 30 October 2013
© Springer Science+Business Media New York 2013

Abstract

Background While child self-reports of psychopathology are increasingly accepted, little standardized instruments are utilized for these practices. The Berkeley Puppet Interview (BPI) is an age-appropriate instrument for self-reports of problem behavior by young children.

Objective Psychometric properties of the Dutch version of the BPI will be reported, specifically, test–retest reliability, intra-class correlations, congruent and concurrent validity.

Methods In a sample of 300 children ($M_{\text{age}} = 7.04$ years, $SD = 1.15$), the BPI was administered twice, with a 1-year interval. Parents and teachers filled out questionnaires about their children’s problem behavior.

Results Findings from the analyses indicate that the BPI subscales have sufficient test–retest reliability and can be reliably coded. Furthermore, findings suggest adequate congruent validity. More support for concurrent validity is found among externalizing problems in comparison to internalizing problems.

Conclusions With regard to the present study, the BPI seems to have adequate psychometric properties. As such, the BPI enables interviewing young children about their psychopathology-related symptoms in a standardized way. The BPI could be applied in clinical practice as a complement to the diagnostic cycle, allowing children’s self-reports to play an increasingly important role.

Keywords Berkeley Puppet Interview · Psychometric properties · Screening instrument · Self-report

L. L. Stone (✉) · M. van der Maten · R. C. M. E. Engels · J. M. A. M. Janssens · R. Otten
Behavioural Science Institute, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen,
The Netherlands
e-mail: l.stone@pwo.ru.nl

C. van Daal
Pluryn Nijmegen, Nijmegen, The Netherlands

Problem behavior often develops at a young age. A considerable number of children suffer from mental health problems. Prevalence figures show that between three and eighteen percent of children exhibit symptoms of psychopathology (Carter et al. 2010; Costello et al. 2003). Externalizing problems, such as oppositional defiant behavior, antisocial behavior, and attention difficulties, as well as internalizing problems, including separation anxiety, anxiety, and depressive symptoms, are most common in young children (Egger and Angold 2006; Klein et al. 2005; Lavigne et al. 2009). In addition, co-morbidity is quite common, especially with regard to young children (Lavigne et al. 2009; Scheeringa and Zeanah 2008).

It is important to be able to examine psychopathology at a young age, since high degrees of aggressive and oppositional behavior may become permanent and develop into chronic patterns of externalizing and psychopathological behavior at a later age (Reef et al. 2010). Problem behavior is associated with increased risks of poor academic, social and occupational performance, deteriorated physical and mental health, and substance use (Ansary and Luthar 2009; Bayer et al. 2011; Fergusson et al. 2005; Kim et al. 2008; Morcillo et al. 2011; O'Neill et al. 2011). When assessed early in development, interventions may contribute to the reduction of aggressive, oppositional and other externalizing behaviors, before these negative behavioral patterns become integrated into the child's personality (Hill et al. 2004).

Several factors have contributed to the phenomenon that, in both research and clinical practice, the emphasis is on externalizing rather than internalizing problems. Probably, one major reason behind this is that externalizing behavior is easier to observe than internalizing behavior. Externalizing behaviors, such as tantrums and resistance against rules, are outwardly directed, generally troublesome for the environment, and often provocative in terms of negative feelings (Rubin and Mills 1990). On the other hand, internalizing problems are intra-individual in nature, inwardly directed, and more easily shielded from the environment by the child (Luby et al. 2009). These behaviors attract less attention and cause fewer problems for the child's environment. Of course, a child may still experience such internalizing problems and suffer from them. Indeed, research shows that even young children report on internalizing problems (Luby 2010), and that these problems are related to negative developmental outcomes later in life, including recurrent depressive episodes, poor school performance, impaired functioning of peer and family relationships, and an increased risk of suicide (Bhatia and Bhatia 2007; Cicchetti and Toth 1998). The fact that internalizing problems at a young age are predictive of problems at a later age, stresses the need of early intervention (Bayer and Sanson 2003).

Yet, while 50 % of children expressing externalizing behaviors receive help, this is true for only 20 % of children suffering from internalizing problems (Merikangas et al. 2011). Some researchers suppose that internalizing problems are generally better recognized by children themselves than by other informants (Achenbach et al. 1987). In one respect, it is possible that an informant's background distorts his/her perception of a child's behavior, particularly when the behavior is more ambiguous, as is the case with internalizing problems (Kroes et al. 2003). For example, personality characteristics such as hostility and inadequate interpersonal sensitivity, are associated with reporting on internalizing problems. In another respect, it is likely that children behave differently in several environments (e.g., at home versus at school), which ensures that information derived from different informants is related to the specific context by definition. Hence, the problem with obtaining information from different informants is that these perceptions are context specific and biased by personal backgrounds (Los Reyes and Kazdin 2005a, b). Alongside conventional screening instruments that are used during the problem analysis phase in

clinical practice, including the CBCL/TRF and SDQ (Achenbach and Ruffle 2000; Goodman et al. 2000), it seems worthwhile to pay attention to the possibility of adopting instruments that refer to the child as an informant. This is in accordance with the so-called ‘multi-informant approach’, in which it is recommended to take into account *context* (i.e., at home and elsewhere), and *perspective* (i.e., self and other), when selecting informants (Kraemer et al. 2003). By using self-report instruments, the risk of under-reporting of internalizing problems may be reduced and a more comprehensive picture of the existing problems will arise (Kraemer et al. 2003).

Screening instruments use self-reports of young children to a minor extent. Young children are not always considered reliable informants of their own behavior (Mutsaers 2009; Scheeringa and Haslett 2010). Children’s vocabulary and cognitive development may affect their understanding of questions and interfere with the duration of administration (Arseneault et al. 2005). Furthermore, it is doubted whether children are capable of self-perception, as this concept is related to cognitive development. Moreover, young children are very sensitive to suggestion, which makes interviewing children a challenge and requires specific interviewing skills. Still, already in the 80’s, Harter (1982) showed that children from the age of eight can meaningfully differentiate between various competence scales (cognitive, social, and physical competence, and general self-esteem). Measelle et al. (1998) stated that children’s self-perceptions can indeed be reliably measured by using an age-appropriate instrument. In clinical practice it is also known that children from 6 years can be interviewed as a part of the diagnostic cycle (Van Leeuwen 2002), thereby adding unique information to the diagnostic process. In the last few years, children’s self-reports are valued increasingly (Arseneault et al. 2005; Ialongo et al. 2001; Luby et al. 2007). Specific self-report questionnaires are available for children from 8 years onwards, such as the Child Depression Inventory (CDI; Kovacs 2001), Screen for Child Anxiety Related Emotional Disorders (SCARED; Birmaher et al. 1999), and Perceived Competence Scale for Children (PCSC; Harter 1982). However, in practice, there is no screening instrument available in the Netherlands, that uses children younger than 8 years old as informants for the assessment of their psychopathology. The Berkeley Puppet Interview (BPI; Measelle et al. 1998; Morris et al. 2002) is an interactive interviewing technique, developed in the USA and designed to elicit perceptions of 4, 5 to 8-year-olds in an age-appropriate way. During the BPI, children are interviewed by two hand puppets in order to simulate a conversation between three peers. Each time, these two hand puppets make opposing statements. For example, one puppet indicates: ‘I am a sad child’, whereas the other puppet states: ‘I am not a sad child’. Then, they ask the child together: ‘How about you?’. Influencing the child in the direction of the question that is asked by the interviewer is thus largely avoided (Fig. 1).

In previous studies the BPI has proven to be reliable and valid (Ablow et al. 1999; Arseneault et al. 2005; Luby et al. 2007; Measelle et al. 1998; Morris et al. 2002; Ringoot et al. 2013). However, only one of these studies used longitudinal data and the sample of this study was rather small with less than 100 participants (Measelle et al. 1998). In addition, recent former studies investigated specific problem clusters of the BPI, such as conduct problems or depression (Arseneault et al. 2005; Luby et al. 2007), with one exception (Ringoot et al. 2013). Our aim is to investigate the BPI as a whole. Further, more research into the BPI’s psychometric properties may facilitate its use in clinical practice. As such, the BPI may be suitable for embedding into the diagnostic cycle. Clinicians naturally conduct interviews with children, and the BPI allows doing so in a standardized manner, without disregarding particular case-dependent questions. In addition, it is an age-appropriate instrument of which the administration will take less



Fig. 1 Pictures of the Berkeley Puppet Interview

time than a diagnostic interview. Recently, the BPI was used as a research instrument as part of two large-scale studies in the Netherlands: the Kind in Zicht study (Stone et al. 2013a), and the Generation R study (Jaddoe et al. 2012). Kind in Zicht is a longitudinal research project on incipient emotional and behavioral problems in young children (Stone et al. 2013a). Generation R involves research into early influences on growth and development within a longitudinal multi-ethnic birth cohort (Jaddoe et al. 2012). For the BPI to be used in these studies, a Dutch version was developed in collaboration with the developers of the instrument.

In the present article, we introduce the Dutch version of the BPI as a useful instrument complementary to the diagnostics in the field of psychopathology, and we examine the test–retest reliability, and the congruent and concurrent validity of the BPI in the Kind in Zicht study. We expected the Dutch version of the BPI—like the American version—to be a reliable and valid instrument for self-reports of psychopathology in young children.

Method

Sample and Procedure

In this study, 300 children were interviewed during the first measurement (T1). One child was excluded due to missing data and another child because she was over 8 years old. One year later (T2), 288 of these children (96 %) were re-interviewed, of whom one was excluded because of her advanced age. This resulted in a sample of 298 children at T1, and 287 children at T2. Of these participating children, 50 % was male and the mean age was 6.95 years ($SD = 1.13$; range 5–8 years). The majority of the children was of Dutch origin (97.4 %) and grew up in a two-parent family (92.2 %). Teachers (T1 $n = 282$, T2 $n = 245$) and parents (T1 $n = 289$, T2 $n = 269$) completed questionnaires about the children at both time points. In addition, teachers ($n = 287$) and parents ($n = 287$) completed a questionnaire about the children 1 year before the interviews took place, and this measurement point is referred to as T0. At T0, the teachers' mean age was 36.57 years ($SD = 10.43$), and 93.9 % of them was female. The parents who filled out the questionnaires were on average 38.29 years old ($SD = 3.88$), and 92.9 % of them was female. Over half of the parents were highly educated (54.8 %), 37.3 % had an intermediate education level, and 6.6 % lower education. Slightly over 1 % received some other type of education.

For the present study, longitudinal data (2011(T1)–2012(T2)) from the Kind in Zicht project were used (Stone et al. 2013a), which was approved by the committee on ethics. Within this project, information was collected about the individual children, using multiple informants. Informed consent from the children's parents was obtained. Each year, the BPI was administered to the children by five certified master students or researchers. They all completed a training course in which the interviewing techniques of the BPI were extensively practiced. Subsequently, they each conducted eight practice interviews, and were then evaluated. The interviews were administered at primary schools in January and February of 2011 and 2012. Children were interviewed in an empty classroom to ensure confidentiality. Interviews were videotaped and after completion, the children received a pair of stickers to thank them for their participation.

Measures

BPI

The Berkeley Puppet Interview (BPI; Measelle et al. 1998) is an interactive and age appropriate interviewing technique, designed to elicit self-perceptions in 4.5–8 year-olds. During the interview, children were asked questions by two identical hand puppets: Iggy and Ziggy. Prior to the interview, the puppets introduced themselves and explained in a playful way how the interview is carried out. Using three practice items, the interviewer assessed whether the procedure was clear to the child, and continued with the actual interview or repeated the practice items until the procedure was clear. An example of such a practice item is: Puppet 1: 'I like chocolate', Puppet 2: 'I do not like chocolate. How about you?'. Throughout the interview, the puppets exchanged opposing statements and then asked the child: 'How about you?'. The puppet with which the child agreed repeated the response, thereby confirming the child's answer.

After administration of the interviews, the children's answers were coded by trained observers on a 7-point scale (see Fig. 2). Answers that reflected the absence of psychopathology were coded as either 5, 6, or 7, depending on possible amplifications or

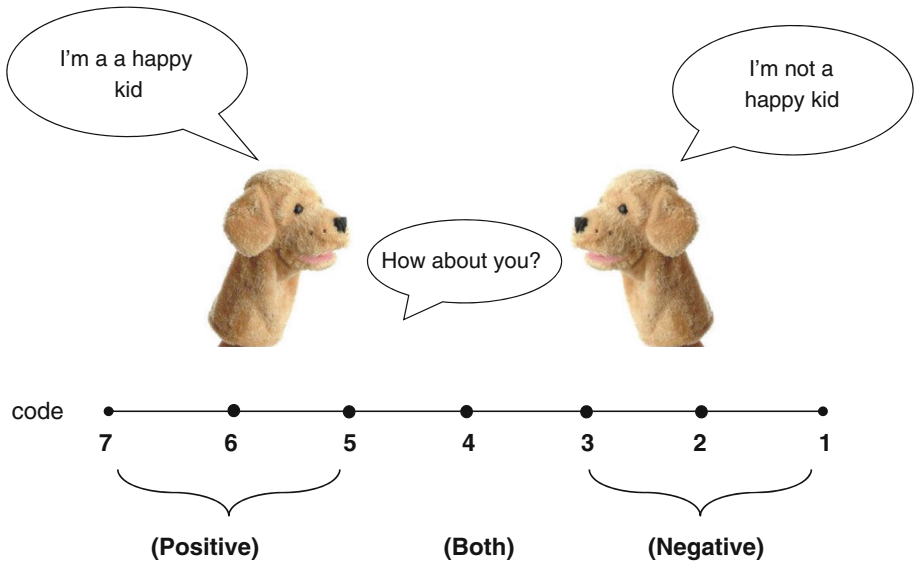


Fig. 2 Coding scale of the Berkeley Puppet Interview

attenuations in the child's response. Code 7 comprised the strongest absence of psychopathology (e.g., 'I am never a sad child'), whereas code 6 meant a neutral absence (e.g., 'I am not a sad child'), and code 5 represented a hesitant response (e.g., 'Usually, I am not a sad child'). On the other side of the spectrum, code 1, 2, or 3 reflected the presence of psychopathology. Code 1 stood for a strong presence (e.g., 'I am always a sad child'), while code 2 represented a neutral response (e.g., 'I am a sad child'), and code 3 was equivalent to a hesitant response (e.g., 'Usually, I am a sad child'). When a child was unable to choose between the two statements, this response was coded as 4. In order to test the reliability of the coding, 15 % of the interviews were double-coded.

The BPI includes 8 subscales (i.e., the symptom scales), that constitute the basis for two overall scales: *internalizing problems* and *externalizing problems*. The internalizing problems scale comprises three subscales: *depression* (7 items; e.g., 'I am a sad child/I am not a sad child'), *anxiety* (7 items; e.g., 'I do have many bad dreams/I do not have many bad dreams'), and *separation anxiety* (6 items; e.g., 'When I am at school, I miss my mum or dad/When I am at school, I do not miss my mum or dad'). We used the internalizing problems scale, as well as the separate symptom scales. The externalizing problems scale also comprised three subscales: *oppositional defiant behavior* (6 items; e.g., 'Sometimes I curse, or I use bad language/I do not curse, or use bad language'), *behavioral problems* (9 items; e.g., 'Sometimes I act cruel towards animals/I do not act cruel towards animals'), and *aggression and hostility towards peers* [from here referred to as aggression] (6 items; e.g., 'I often fight with other children/I do not fight with other children'). In addition, two subscales focus on relationships with peers: *acceptance and rejection by peers* [from here referred to as acceptance/rejection] (5 items; e.g., 'Other children ask me to play along/Other children do not ask me to play along'), and *being bullied* (4 items; e.g., 'Children hit me, or beat me up/Children do not hit me, or beat me up'). The negative and positive statements were presented in a random order. No Cronbach's alpha's will be reported regarding the BPI, since the interview is considered an index scale instead of a Likert scale,

making it unsuitable for calculating this reliability coefficient (Stone et al. 2013b). The interrater reliability is reported in the results section.

SDQ

The Dutch parent and teacher version of the Strengths and Difficulties Questionnaire (SDQ) was used to assess internalizing and externalizing problems (van Widenfelt et al. 2003). The subscales measuring emotional problems (e.g., often unhappy, down-hearted or tearful) and behavioral problems (e.g., often lying or cheating) each consist of five items. Parents or teachers judged children on a 3-point scale, from 0 (not true) to 2 (very true). The scoring manual is available online (www.sdqinfo.com). In the Kind in Zicht study, the psychometric properties of the SDQ were adequate, as described elsewhere (Stone et al. 2013b).

CBCL/TRF

The Dutch versions of the Child Behavior Check List (CBCL) and Teacher Report Form (TRF) were also used (at T0) to measure internalizing and externalizing behavior, as reported by parents and teachers (Achenbach and Rescorla 2000; Achenbach and Rescorla 2001; Verhulst et al. 1997). The C-TRF and C-CBCL are intended for children aged 1.5–5 years and contain 100 items; the TRF and CBCL are intended for 5–18 year-olds and contain 118 items. The C-TRF and TRF were filled out by teachers, whereas the C-CBCL and CBCL were filled out by parents. Items were scored on a 3-point Likert scale, where 0 represents ‘not true’, and 2 stands for ‘very true or often true’. Three scales (i.e., somatic symptoms, anxious-depressed, and withdrawn) were combined in order to constitute the internalizing scale. Combining two scales (i.e., violation of rules and aggressive behavior) resulted in the externalizing scale. The psychometric properties of this instrument in the Kind in Zicht study were again adequate (Stone et al. 2013b).

Strategy for Analysis

First, descriptive statistics that provide insight into the level of psychopathology for the whole sample will be shown, disaggregated for gender and age group (4–5 and 6–7 years). Besides, an independent *t* test was conducted to test whether the mean scores of boys and girls, and younger and older children, respectively, differ statistically. Originally, the BPI is scored in such a way that higher scores reflect lower levels of psychopathology. In our opinion, this is somewhat confusing. For the sake of clarity regarding the interpretation, the scores were therefore coded the other way around (i.e., 1 becomes 7, and vice versa), such that higher means reflected higher levels of problem behavior. These reversed scores were used for calculating means and standard deviations.

Subsequently, the reliability of the BPI codes was examined using intra-class correlations and test–retest correlations. The intra-class correlation coefficient [ICC] was calculated to determine the reliability between two coders per BPI subscale. The higher the ICC, the more reliable the coding, where a score of 1 represents absolute agreement. ICC values of $>.60$ are considered good and values $>.75$ are considered excellent (Cicchetti et al. 2011). Pearson correlations were used for calculating test–retest correlations. These test–retest correlations were calculated for the entire group, and for gender and age separately.

In terms of validity, congruent validity was examined first by mutually correlating the BPI subscales. Additionally, concurrent validity was defined by correlating the BPI outcomes with the outcomes of the other questionnaires; again using Pearson correlations. When comparing the BPI with the SDQ and CBCL, the BPI subscales were ranged under two headings; the internalizing problems scale and the externalizing problems scale. These were compared with the emotional and behavioral problems scale of the parent and teacher versions of the SDQ. The CBCL also used an internalizing and externalizing problems scale, that was completed by both parents (CBCL) and teachers (TRF). Because of the ages of a restricted group of children, alternative versions were deployed; the C-CBCL and the C-TRF. In order to clearly show the possible similarities and differences between the BPI and CBCL, the standardized T-scores of the CBCL and C-CBCL, and those of the TRF and the C-TRF were combined.

Results

Descriptive Statistics

The descriptive statistics of the BPI subscales appear in Table 1. The mean scores on the subscales were low. *T* tests for paired observations showed that the mean scores of depression, separation anxiety, anxiety, behavioral problems, and being bullied, declined from T1 to T2. In addition, it was tested whether mean differences regarding age and gender at T1 and T2 were present. The *t* test for gender at T1 showed that there were statistically significant mean differences for separation anxiety ($t(286) = -2.25$, $p < 0.05$), aggression ($t(289) = 3.56$, $p < 0.01$), and acceptance/rejection ($t(284) = 2.04$, $p < 0.05$), but not for depression, anxiety, behavioral problems, oppositional defiant behavior, and being bullied. The mean scores of boys on the aggression and acceptance/rejection subscales were higher than those of girls, while girls scored higher on separation anxiety than boys. At T2, the *t* test for gender was statistically significant for the subscales separation anxiety ($t(279) = -3.37$, $p < 0.05$), oppositional defiant behavior ($t(280) = 3.02$, $p < 0.05$), behavioral problems ($t(280) = 2.07$, $p < 0.05$), aggression ($t(279) = 3.96$, $p < 0.01$), acceptance/rejection ($t(280) = 2.49$, $p < 0.05$), and being bullied ($t(279) = 2.39$, $p < 0.05$), but not for depression and anxiety. Mean scores of boys at T2 were higher than those of girls on the subscales oppositional defiant behavior, behavioral problems, aggression, acceptance/rejection, and being bullied, whereas girls reported higher scores on separation anxiety than boys. In conclusion, boys generally reported more externalizing problems than girls at both time points.

As regards the *t* test for age, mean scores for depression ($t(282) = 2.46$, $p < 0.05$) and acceptance/rejection ($t(276) = 2.22$, $p < 0.05$) were found to be higher for younger children as opposed to older children at T1. At T2, younger children also reported more symptoms of depression ($t(273) = 2.76$, $p < 0.01$), as well as aggression ($t(272) = 2.12$, $p < 0.05$), and they indicated to be bullied more than older children ($t(272) = 3.77$, $p < 0.01$).

Intra-class Correlations

The following ICC's were obtained for the separate subscales, for T1 and T2 respectively: depression (.74, .86), anxiety (.70, .80), separation anxiety (.70, .83), oppositional defiant behavior (.66, .71), behavioral problems (.81, .66), aggression (.78, .77), acceptance/

Table 1 Descriptive statistics of the BPI subscales at T1 and T2

	T1 N = 291–297		T2 N = 286–287		<i>t</i> value
	<i>M</i> (<i>SD</i>)	Range	<i>M</i> (<i>SD</i>)	Range	
Depression	2.61 (.62)	1.71–5.00	2.51 (.59)	1.67–6.00	2.31*
Anxiety	3.10 (.83)	1.71–6.57	2.89 (.84)	1.86–5.86	3.63**
Separation anxiety	3.33 (.91)	1.83–6.67	3.07 (.98)	1.83–6.00	4.15**
Oppositional/defiant	2.67 (.61)	1.33–4.80	2.58 (.59)	1.67–4.50	1.89
Behavioral problems	2.56 (.64)	1.22–5.11	2.43 (.51)	1.44–4.56	3.23**
Aggression	2.32 (.57)	1.33–5.33	2.32 (.51)	1.67–5.33	–1.00
Acceptance/rejection	2.67 (.84)	1.60–6.00	2.56 (.84)	1.60–6.00	1.84
Being bullied	3.01 (1.06)	1.50–6.25	2.72 (1.04)	1.75–6.00	3.69**

* $p < 0.05$; ** $p < 0.01$

Table 2 Longitudinal associations of the BPI subscales by gender and age-group

Scales	Total <i>r</i> (n)	Boys <i>r</i> (n)	Girls <i>r</i> (n)	Younger <i>r</i> (n)	Older <i>r</i> (n)
Depression	.29** (283)	.32** (137)	.30** (141)	.23** (134)	.34** (137)
Separation anxiety	.39** (282)	.35** (135)	.39** (131)	.28** (133)	.47** (137)
Anxiety	.35** (278)	.38** (134)	.33** (140)	.34** (130)	.34** (136)
Oppositional/defiant	.34** (281)	.26** (136)	.40** (141)	.31** (133)	.37** (136)
Behavioral problems	.38** (279)	.33** (134)	.43** (141)	.35** (132)	.42** (135)
Aggression	.31** (282)	.30** (136)	.25** (141)	.34** (134)	.25** (136)
Acceptance/rejection	.37** (279)	.33** (135)	.38** (140)	.21* (133)	.54** (134)
Being bullied	.26** (280)	.13 (135)	.37** (142)	.14 (134)	.40** (137)

* $p < 0.05$; ** $p < 0.01$

rejection (.82, .82), and being bullied (.74, .88). These correlations indicated that the BPI subscales can be reliably coded by multiple coders.

Test–Retest Reliability

In Table 2, the results with regard to test–retest reliability, with a time interval of 1 year, are presented. These showed that, overall, the psychopathology self-reports as provided by the children were rather stable. Boys appeared to report somewhat less stable than girls, in terms of oppositional defiant behavior, behavioral problems, and being bullied. Moreover, the correlations regarding depression, separation anxiety, acceptance/rejection, and being bullied were less pronounced in young children than in older children. The test–retest reliability of these scales thus increased with age.

Congruent Validity

As is apparent from Table 3, the BPI subscales correlated significantly at T1 and T2. The correlations were weak to moderate, and the pattern of correlations was as expected; the reports of certain types of problem behaviors were associated with the reports of other types

Table 3 Correlations among the BPI subscales at T1 and T2

	1	2	3	4	5	6	7	8
1. Depression	–	.19**	.41**	.18**	.20**	.13*	.15*	.31**
2. Separation anxiety	.38**	–	.32**	–.05	.01	.03	–.03	.17**
3. Anxiety	.40**	.53**	–	.08	.19**	.14*	.03	.25**
4. Oppositional/defiant	.20**	.06	.11	–	.43**	.38**	.26**	.28*
5. Behavioral problems	.24**	.05	.03	.53**	–	.46**	.13*	.29**
6. Aggression	.16**	.09	.19**	.37**	.44**	–	.07	.26**
7. Acceptance/rejection	.36**	.11	.19**	.27**	.24**	.24**	–	.26**
8. Being bullied	.34**	.23**	.28**	.24**	.21**	.26**	.39**	–

Numbers above the diagonal T1 (N = 291–297), numbers below the diagonal T2 (N = 286–287)
 * $p < 0.05$; ** $p < 0.01$

Table 4 Correlations among the BPI subscales and the SDQ T1 and T2 scale scores

		SDQ			
		Parent		Teacher	
		Emotional problems	Behavioral problems	Emotional problems	Behavioral problems
		T1			
BPI	Internalizing	.14*	.05	.09	.10
	Externalizing	.11	.29**	–.03	.25**
		T2			
	Internalizing	.08	–.03	.17**	.06
	Externalizing	.06	.23**	.08	.27**

* $p < 0.05$; ** $p < 0.01$

of problem behaviors (e.g., anxiety was correlated with depression). The internalizing subscales depression, separation anxiety, and anxiety, correlated weakly with the externalizing subscales oppositional defiant behavior, behavioral problems, and aggression. The correlations between the internalizing subscales themselves were stronger, especially between anxiety and depression, and anxiety and separation anxiety. Furthermore, oppositional defiant behavior, behavioral problems, and aggression correlated relatively strongly with one another. Acceptance/rejection correlated predominantly with depression and oppositional defiant behavior, and to a lesser extent with behavioral problems, aggression, and anxiety. The subscale being bullied was correlated with all other subscales. In summary, various problem behaviors were meaningfully intercorrelated within this young age group.

Concurrent Validity

The externalizing subscales of the BPI and the SDQ were correlated at T1 and T2, concerning both parents and teachers (see Table 4). The more externalizing problems the children reported, the more behavioral problems parents and teachers reported likewise. It is noteworthy that the internalizing subscales of the BPI and the SDQ correlated to a lesser extent than the externalizing subscales. In order to explain this difference, the individual

internalizing BPI subscales (i.e., anxiety, depression, and separation anxiety), were correlated to the SDQ emotional problems scale score. Depression, separation anxiety, and anxiety were uncorrelated with emotional problems as reported by teachers at T1: $r(277) = .09$, n.s.; $r(277) = .05$, n.s.; $r(277) = .09$, n.s., respectively. Similarly, separation anxiety ($r(237) = .11$, n.s.) and anxiety ($r(237) = .10$, n.s.) did not correlate with emotional problems as reported by teachers at T2, but depression did: $r(238) = .21$, $p < .01$.

As for the parents as informants, it was noticed that whereas at T1 the internalizing BPI scale was correlated with emotional problems, it was no longer at T2. Next, idem, the separate BPI subscales were correlated to the SDQ emotional problems scale score. At both time points, no correlation was found between separation anxiety and emotional problems (T1: $r(280) = .04$, n.s.; T2: $r(255) = .02$, n.s.) and between anxiety and emotional problems (T1: $r(276) = .08$, n.s.; T2: $r(255) = .02$, n.s.). Depression was found to be associated with emotional problems at both T1 and T2 (T1: $r(280) = .12$, $p < .05$; T2: $r(256) = .17$, $p < .01$). From these results, we can conclude that children's self-reports of depression corresponded to some extent to the emotional problems reports by teachers and parents; the more emotional problems teachers and parents reported, the more depression children reported. However, children's self-reports of anxiety and separation anxiety did not correspond to teachers' and parents' reports of emotional problems.

The BPI subscales measured at T1 have also been compared with the CBCL/TRF scale scores at T0. Children's self-reported internalizing problems did not correlate with parent's and teachers' reported problems ($r(278) = -.00$, n.s.; $r(286) = -.02$, n.s., respectively). However, the correlations between children's self-reports and the reports of their parents ($r(279) = .20$, $p < .01$) and teachers ($r(287) = .14$, $p < .05$) on externalizing problems were significant. Children's reports regarding internalizing problems were not correlated with the reports of parents and teachers about the children's behaviors in the previous year, while children's reports regarding externalizing problems were.

Discussion

At present, no standardized instrument is available in the Netherlands for measuring self-perceptions of problem behavior in young children (Mutsaers 2009). This is problematic, since it is known that there may be great differences in reports of parents and teachers about children's behaviors (Los Reyes and Kazdin 2005a, b). As a consequence, certain problem behaviors may not be recognized. Therefore, it is important that attention is paid to self-reports of problem behavior by young children. In this article, the Dutch version of the Berkeley Puppet Interview (BPI) was presented, which is a standardized and age-appropriate instrument for interviewing young children about their self-perceptions of problem behaviors. In addition, several psychometric properties of the BPI were presented.

We expected that the results regarding reliability and validity would be consistent with earlier research into the BPI. The results suggest that the BPI scales can be sufficiently reliably coded, that the subscales are correlated after 1 year, and that the subscales are meaningfully intercorrelated, which indicates congruent validity. The analyses concerning the intra-class correlation coefficients and test-retest reliability imply that the BPI is a consistent, reliable interviewing method. Though, it should be noted that the intra-class correlation for oppositional defiant behavior were somewhat lower. The interpretation of the results of this subscale should be interpreted with some caution. Still, even after a 1-year interval, during which, of course, not only reliability was assessed, but also development, there appeared to be clear patterns in the behaviors children report. The test-

retest coefficients are not as high as typically found in studies that focus on adults, but are similar to other studies investigating the BPI's psychometric properties (Measelle et al. 1998). Furthermore, theoretically speaking, it was to be expected that the BPI subscales were meaningfully interrelated. This indicates that the BPI seems to measure the constructs that are intended to be measured. However, for determining congruent validity, it is also necessary that the BPI will be compared to external measures, such as standardized tests that assess school performance. Although children are sometimes still not considered reliable informants of their own problems (Mutsaers 2009; Scheeringa and Haslett 2010), the results of this study seem to indicate the opposite. This is in line with other studies that have been conducted into the BPI (Arseneault et al. 2005; Luby et al. 2007; Measelle et al. 1998), and with recommendations to clinicians, that children from the age of 6 years can be interviewed as part of the diagnostic cycle (Van Leeuwen 2002). The comparison of the BPI with the SDQ and CBCL/TRF, shows that differences between reports of multiple informants are indeed great. It is important to note that comparing scores on the BPI on the one hand, and the SDQ and CBCL/TRF on the other hand is difficult, given the nature of the instruments; an interviewing technique versus a questionnaire. In spite of this difference in method, the correlations between comparable concepts measured using the BPI and SDQ or CBCL/TRF, remain weak.

This phenomenon, 'informant disagreement', is a well-known issue when comparing reports from multiple informants (Los Reyes and Kazdin 2005a, b). As expected, the agreement was greater in terms of externalizing behavior, than with respect to internalizing behavior, although the agreement on externalizing behavior was also very low. These results underscore that reports of problem behavior by parents and teachers cannot simply be regarded as corresponding to children's perceptions (Achenbach et al. 1987; Los Reyes and Kazdin 2005a, b), particularly when it comes to reporting internalizing problems, where agreement between children and parents and teachers was very limited (Achenbach et al. 1987). These results also imply that child reports provide important information additional to the process of information gathering in the problem analysis phase. In this respect, the BPI could be a useful instrument. Based on the current state of research into the BPI, however, clinicians are recommended to also keep in mind the limitations of the BPI, when using this instrument. It is not recommended to use the BPI as a single instrument, but it seems suitable for gaining more insight into certain symptoms and for confirming or rejecting hypotheses regarding a child's symptoms. In addition to the BPI, another promising instrument is available for children aged 6–11 years old: the Dominic Interactive (DI; Valla 2000; Kuijpers et al. 2013). The DI is a structured digital questionnaire that assesses the most common internalizing and externalizing problems in children. It takes into account the child's developmental level, by means of supporting the questions by visual and auditory stimuli. The item is both displayed through an image of the problem situation, and made audible by being read out loud by the program.

Limitations and Future Directions

The present study showed that the BPI has adequate psychometric properties, although we believe that more research into the internal structure of the BPI is necessary and highly recommended for further research. A recent study did confirm the internal structure of the BPI and reported Cronbach's alpha's for the subscales (Ringoot et al. 2013). Yet, a thorough test of the internal structure of the BPI is hampered by the bimodal frequency distribution, and in our opinion as such, not suitable for the execution of conventional reliability analyses, such as calculating Cronbach's alpha and testing the factor structure.

The BPI thus appears to be a sound and useful instrument which could be used in child and youth care. Still, it is important that, in the future, the experiences using the BPI in clinical practice, and its functioning in a clinical setting will be explored. After all, little is known about using the BPI in clinical practice. Thus far, the results that have emerged from studies into the BPI are promising (Arseneault et al. 2005; Measelle et al. 1998; Ringoot et al. 2013), and suggest that the BPI can constitute a valuable supplement to youth care practices. When research from a clinical setting on use of the BPI is available, it may possibly be embedded in evidence-based practice (Mash and Hunsley 2005). In conclusion, by means of this article we hope to have provided greater BPI publicity, to allow for optimal utilization of this instrument within youth care.

Acknowledgments This research was granted by the Dutch Organization for Health Research and Care Innovation (ZonMW: 80-82435-98-8026).

References

- Ablow, J. C., Measelle, J. R., Kraemer, H. C., Harrington, R., Luby, J., Smider, N., et al. (1999). The MacArthur three-city outcome study: Evaluating multi-informant measures of young children's symptomatology. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38(12), 1580–1590. doi:10.1097/00004583-19991200-00020.
- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlation for situational specificity. *Psychological Bulletin*, 101, 213–232.
- Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for ASEBA school-age forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Achenbach, T. M., & Ruffle, T. M. (2000). The child behavior checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in Review*, 21, 265–279.
- Ansary, N. S., & Luthar, S. S. (2009). Distress and academic achievement among adolescent of affluence: A study of externalizing and internalizing problem behaviors and school performance. *Development and Psychopathology*, 21, 319–341.
- Arseneault, L., Kim-Cohen, J., Taylor, A., Caspi, A., & Moffit, T. E. (2005). Psychometric evaluation of 5- and 7-year old children's self-reports of conduct problems. *Journal of Abnormal Child Psychology*, 33, 537–550.
- Bayer, J. K., Rapee, R. M., Hiscock, H., Ukourmunne, O. C., Mihalopoulos, C., & Wake, M. (2011). Translational research to prevent internalizing problems early in childhood. *Depression and Anxiety*, 28, 50–57.
- Bayer, J. K., & Sanson, A. V. (2003). Preventing the development of emotional mental health problems from early childhood: recent advances in the field. *International Journal of Mental Health Promotion*, 5, 4–16.
- Bhatia, S. K., & Bhatia, S. C. (2007). Childhood and adolescent depression. *American Academy of Family Physicians*, 75, 73–80.
- Birmaher, B., Brent, D. A., Chiappetta, L., Bridge, J., Monga, S., & Baugher, M. (1999). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): A replication study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 1230–1236.
- Carter, A. S., Wagmiller, R. J., Gray, S. A. O., McCarthy, K. J., Horwitz, S. M., & Briggs-Gowan, M. J. (2010). Prevalence of DSM-IV disorder in a representative, healthy birth cohort at school entry: Sociodemographic risks and sociale adaptation. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47, 686–698.
- Cicchetti, D. V., Koenig, K., Klin, A., Volkmar, F. R., Paul, R., & Sparrow, S. (2011). From Bayes through marginal utility to effect sizes: A guide to understanding the clinical and statistical significance of the results of autism research findings. *Journal of Autism and Developmental Disorders*, 41, 168–174.
- Cicchetti, D., & Toth, S. L. (1998). The development of depression in children and adolescents. *American Psychologist*, 53, 221–241.

- Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorder in childhood and adolescence. *Archives of General Psychiatry*, *60*, 837–844.
- De Los Reyes, A., & Kazdin, A. E. (2005a). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, *132*, 483–509.
- De Los Reyes, A., & Kazdin, A. E. (2005b). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, *131*, 483–509.
- Egger, H. L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*, *47*, 313–337.
- Fergusson, D. M., Horwood, L. J., & Ridder, E. M. (2005). Show me the child at seven: The consequences of conduct problems in childhood for psychosocial functioning in adulthood. *Journal of Child Psychology and Psychiatry*, *46*, 837–849.
- Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the strengths and difficulties questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, *177*, 534–539.
- Harter, S. (1982). The perceived competence scale for children. *Child Development*, *53*, 87–97.
- Hill, L. G., Lochman, J. E., Coie, J. D., & Greenberg, M. T. (2004). Effectiveness of early screening for externalizing problems: Issues of screening accuracy and utility. *Journal of Consulting and Clinical Psychology*, *72*, 809–820.
- Ialongo, N. S., Edelsohn, G., & Kellam, S. G. (2001). A further look at the prognostic power of young children's reports of depressed mood and feelings. *Child Development*, *72*, 736–747.
- Jaddoe, V. W., van Duijn, C. M., Franco, O. H., van der Heijden, A. J., van IJzendoorn, M. H., de Jongste, J. C., et al. (2012). The Generation R Study: Design and cohort update 2012. *European Journal of Epidemiology*, *27*, 739–756.
- Kim, T. E., Guerra, N. G., & Williams, K. R. (2008). Preventing youth problem behaviors and enhancing physical health by promoting core competencies. *Journal of Adolescent Health*, *43*, 401–407.
- Klein, D. N., Dougherty, L. R., & Olino, T. M. (2005). Toward guidelines for evidence-based assessment of depression in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *34*, 412–432.
- Kovacs, M. (2001). *Children's depression inventory (CDI)*. North Tonawanda, NY: Multi Health Systems Inc.
- Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry*, *160*, 1566–1577.
- Kroes, G., Veerman, J. W., & De Bruyn, E. E. J. (2003). Bias in parental reports? Maternal psychopathology and the reporting of problem behavior in clinic-referred children. *European Journal of Psychological Assessment*, *19*, 195–203.
- Kuijpers, R. C. W. M., Otten, R., Krol, N. P. C. M., Vermulst, A. A., & Engels, R. C. M. E. (2013). The reliability and validity of the Dominic Interactive: A computerized child report instrument for mental health problems. *Child & Youth Care Forum*, *1*, 35–52.
- Lavigne, J. V., LeBailly, S. A., Hopkins, J., Gouze, K. R., & Binns, H. J. (2009). The prevalence of ADHD, ODD, depression, and anxiety in a community sample of 4-year olds. *Journal of Clinical Child & Adolescent Psychology*, *38*, 315–328.
- Luby, J. L. (2010). Preschool depression: The importance of identification of depression early in development. *Current Directions in Psychological Science*, *19*, 91–95.
- Luby, J. L., Belden, A., Sullivan, J., & Spitznagel, E. (2007). Preschoolers' contribution to their diagnosis of depression and anxiety: Uses and limitations of young child self-report of symptoms. *Child Psychiatry and Human Development*, *38*, 321–338.
- Luby, J. L., Si, X., Belden, A. C., Tandon, M., & Spitznagel, E. (2009). Preschool depression: Homotypic continuity and course over 24 months. *Archives of General Psychiatry*, *66*, 897–905.
- Mash, E. J., & Hunsley, J. (2005). Evidence-based assessment of child and adolescent disorders: Issues and challenges. *Journal of Clinical Child and Adolescent Psychology*, *34*, 362–379.
- Measelle, J. R., Ablow, J. C., Cowan, P. A., & Cowan, C. P. (1998). Assessing young children's views of their academic, social and emotional lives: An evaluation of the self-perception scales of the Berkeley Puppet Interview. *Child Development*, *69*, 1556–1676.
- Merikangas, K. R., He, J., Burstein, M., Swendsen, J., Avenevoli, S., Case, B., et al. (2011). Service utilization for lifetime mental disorder in U.S. adolescents: Results of the National Comorbidity Survey-Adolescent Supplement (NCS-A). *Journal of the American Academy of Child and Adolescent Psychiatry*, *50*, 32–45.

- Morcillo, C., Duarte, C. S., Sala, R., Wang, S., Lejuez, C. W., Kerridge, B., et al. (2011). Conduct disorder and adult psychiatric diagnoses: Associations and gender differences in the U.S. adult population. *Journal of Psychiatric Research*, *46*, 323–330.
- Morris, A. S., Silk, J. S., Steinberg, L., Sessa, F. M., Avenevoli, S., & Essex, S. J. (2002). Temperamental vulnerability and negative parenting as interacting predictors of child adjustment. *Journal of Marriage and Family*, *64*, 461–471.
- Mutsaers, K. (2009). *Het herkennen en diagnosticeren van depressieve stoornissen*. Retrieved, August 20, 2012, from <http://www.nji.nl/nji/dossierDownloads/Instrumenten%20depressie.pdf>.
- O'Neill, K. A., Conner, B. T., & Kendall, P. C. (2011). Internalizing disorders and substance use disorders in youth: Comorbidity, risk, temporal order, and implications for intervention. *Clinical Psychology Review*, *31*, 104–112.
- Reef, J., Diamantopoulous, S., Van Meurs, I., Verhulst, F., & Van der Ende, J. (2010). Predicting adult emotional and behavioral problems from externalizing problem trajectories in a 24-year longitudinal study. *European Child and Adolescent Psychiatry*, *19*, 577–585.
- Ringoot, A. P., Jansen, P. W., Steenweg-de Graaff, J., Measelle, J. R., Van der Ende, J., Raat, H., et al. (2013). Young children's self-reported emotional, behavioral and peer problems: The Berkeley Puppet Interview. *Psychological Assessment*. doi:10.1037/a0033976.
- Rubin, K. H., & Mills, R. S. L. (1990). Maternal beliefs about adaptive and maladaptive social behaviors in normal, aggressive, and withdrawn preschoolers. *Journal of Abnormal Child Psychology*, *18*, 419–435.
- Scheeringa, M. S., & Haslett, N. (2010). The reliability and criterion validity of the diagnostic infant and preschool assessment: A new diagnostic instrument for young children. *Child Psychiatry and Human Development*, *41*, 299–312.
- Scheeringa, M. S., & Zeanah, C. H. (2008). Reconsideration of Harm's way: Onsets and comorbidity patterns of disorders in preschool children and their caregivers following hurricane Katrina. *Journal of Clinical Child & Adolescent Psychology*, *37*, 508–518.
- Stone, L. L., Giletta, M., Brendgen, M., Otten, R., Engels, R. C. M. E., & Janssens, J. M. A. M. (2013a). Friendship similarities in internalizing problems in early childhood. *Early Childhood Research Quarterly*, *28*, 210–217.
- Stone, L. L., Otten, R., Janssens, J. M. A. M., Soenens, B., Kuntsche, E., & Engels, R. C. M. E. (2013b). Does parental psychological control relate to internalizing problems in early childhood? An examination using the Berkeley Puppet Interview. *International Journal of Behavioral Development*, *37*, 309–318.
- Valla, J. P. (2000). Instruction manual for the Dominic Interactive. In J. P. Valla (Ed.), *The Dominic Interactive*. DIMAT: Montreal, Canada.
- Van Leeuwen, H. M. P. (2002). Het diagnostisch interview met het kind. In T. Kievit, J. A. Tak, & J. D. Bosch (Red.), *Handboek psychodiagnostiek voor de hulpverlening aan kinderen* (pp. 125–144). Utrecht: De Tijdstroom.
- van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D. A., & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child and Adolescent Psychiatry*, *12*, 281–289.
- Verhulst, F. C., Van der Ende, J., & Koot, H. M. (1997). *Handleiding voor de Teacher's Report Form (TRF)*. Rotterdam: Afdeling Kinder- en Jeugdpsychiatrie, Sophia Kinderziekenhuis/Erasmus MC.