

Evaluation of a Puppet Interview to Measure Young Children's Self-Reports of Temperament

Janet H. Roth

Griffith University, Australia

Mark R. Dadds and John McAloon

University of New South Wales, Australia

This study developed and evaluated a puppet interview that allows children to self-report on temperamental constructs. Structured child self-report measures are rarely utilised in clinical assessment of young children under the age of 7–8 years. Given that clinical assessment is often characterised by low convergence between raters, such a measure may offer important contributions. The present study developed and evaluated a measure based on items from the Colorado Childhood Temperament Inventory and reports two studies with child participants aged 4 to 5 years. Independent observations of the children were also made. Results showed moderate levels of internal consistency and stability, and convergence between child self-report and teacher/parent raters was low, but similar levels of agreement were achieved between adult informants. The puppet interview thus showed some potential but highlighted the difficulties of self-report in young children within a multiple informant framework in clinical assessment.

Assessment of behavioural and emotional problems in children is best informed by data drawn from multiple informants. A number of problems, however, have been associated with including self-report data from children younger than 7–8 years (Ablow et al., 1999; Achenbach, 1995; Offord, 1996). As a result, young children have generally been excluded as informants (e.g., Edelbrock, Costello, Dulcan, Conover, & Kalas, 1985). To date, few attempts have been made to develop self-report instruments for assessing emotional problems in children under the age of 6 (e.g., Lugt-Tappeser & Kriependorf, 1992; Spence, 1998). The present study evaluates a new self-report measure for assessing temperamental constructs in young children.

A number of concerns have been raised with regard to the cognitive capabilities of children under 7–8 years old. Many children have not developed sufficient cognitive ability to provide valid responses concerning either traits or the presence or absence of symptoms relevant to diagnostic criteria (Achenbach, 1995; Sturgess & Zivaniani, 1995). Recent evidence, however, suggests that by the age of 3 years, children are sufficiently cognitively developed to organise information about themselves (Eder & Mangelsdorf, 1997). In addition, young children have been demonstrated to be sufficiently cognitively developed to

Address for correspondence: Professor Mark R. Dadds, School of Psychology, University of New South Wales, Sydney NSW 2052, Australia. Email: m.dadds@unsw.edu.au

assess, interpret and elaborate on the reasons for and consequences of emotional experience (Lagattuta & Wellman, 2002; Adams, Kuebli, Boyle, & Fivush, 1995). Thus, the ability of young children to report accurately on their emotions and behaviour remains controversial.

Attempts to extend adult models of assessment to children have seen structured diagnostic interviews developed for children aged as young as 6 years. However, when test–retest data are collected on such measures, pronounced attenuation effects have been demonstrated (Edelbrock et al., 1985). It is noteworthy that similar effects of smaller magnitude have been demonstrated in adult samples completing structured diagnostic interviews (Robins, 1985). The attenuation effects demonstrated by children have been taken to support the argument that children's cognitive abilities are insufficiently developed to warrant the use of their self-report data (Edelbrock et al., 1985). The possibility remains that it is the format of the structured diagnostic interview and not the children's cognitive ability that results in unreliable self-report data. For example, evidence suggests that young (preschool) children are able to comprehend their own emotions if a similar emotion is overtly expressed as a prompt, whereas older (early school age) children have been found to rely more on internal states and social interactions in their self-representations (Harris, Olthof, & Terwogt, 1981; Carroll & Steward, 1984; Nannis & Cowan, 1987).

Little work has been done specifically on young children's ability to report on their own temperaments. One study (Lugt-Tappeser, 1994) correlated direct observations and multiple reports from parents, teachers and structured interviews with 4- to 6-year-old children to determine if anxiety could be successfully identified in this age group. A 26-item standardised interview for children, the Marburg Anxiety Scale (Lugt-Tappeser & Kriependorf, 1992), was administered to 128 children. Children responded to questions about various forms of anxiety relevant to their age (e.g., darkness, separation, general worries) as well as general satisfaction, social competence and perceptions of peer acceptance. The children were interviewed individually and also rated on a scale of 1 to 5 for observed situational anxiety during the interview. The children's interviews correlated moderately with teachers' ratings and to some extent with the direct observations. However, there was wide variability in the relationship between maternal and child reports, according to the situation and type of anxiety. The reliability of the children's interviews was also questionable, partly due to the small number of questions representing each category of anxiety. Although these results must be interpreted cautiously, they do provide encouragement for further research into the potential of young children to reliably report on their psychological states.

In an attempt to address both developmental factors and methodological issues, Ablow et al. (1999) used a puppet interview with community and clinical samples of 4- to 8-year-old children. Internal consistency of subscales related to a range of internalising and externalising problems ranged from .36 to .89, centring around .6 to .7. Test–retest reliabilities ranged from .42 to .76, centring around .6, over a 7- to 10-day period. Finally, reasonable discriminant validity was shown in that mean scores discriminated between various diagnostic groups of children. The applicability of these results to the subsample of preschool-aged children included in the present study is not known.

Eder (1990) also used a structured puppet interview; however, the participants were slightly younger than in Ablow et al.'s (1999) study, (3½ to 7½ years as

opposed to 4 to 8 years). The aim of this study was to assess self-reports of personality characteristics (e.g., achievement, control) rather than specific symptoms. Eder presented two puppets, matched by gender to the child being tested. Statements made by the puppets represented either end of a continuum for each trait characteristic. For example, "I like to climb high" versus "I do not like to climb high"; "I usually play with friends" versus "I usually play by myself". Similar to Ablow et al. (1999), measures showed reasonable internal consistency across the trait characteristics, and were moderately consistent over 1 month, especially for the eldest children (Eder, 1990). Eder's work did not examine the children's self-reports with respect to multiple informants.

On the basis of this research, it appears that the use of structured puppet interviews may hold promise for gaining reliable and valid self-report data from children. The aim of the present study was to develop and evaluate an interview measure for 4- to 5-year-old children based upon the Colorado Childhood Temperament Inventory (CCTI; Buss & Plomin, 1984). Given the previous findings by Ablow et al. (1999), Eder (1990), and Lugt-Tappeser (1994), the use of puppets was chosen to provide an appropriate assessment tool for exploring young children's ability to report on their temperamental traits. As part of a larger longitudinal study on prevention of internalising disorders, the puppet study focused more closely on traits implicated in the development of problems with internalising disorders in childhood. The CCTI was chosen because it is one of the most commonly used and best-validated research tools for assessment of temperament in young children. To overcome previous limitations of assessment of young children, assessments included parent reports of children's temperament, teacher reports of social and emotional adjustment and independent observations, as well as drawing from a larger sample of children than previous studies.

The overall aim was to assess the validity and reliability of the children's reports. Study 1 examined the internal consistency of the puppet interview (PI). Inverse relationships were anticipated between the traits of shyness and sociability, and between emotionality and soothability (Buss, 1995). Second, test-retest reliability of the PI was assessed. Finally, construct validity was examined by assessing the degree of convergence between the children's reports and those of parent, teacher and observer raters.

The issue of construct validity is somewhat problematic. Previous research has shown that parents and teachers show low (albeit statistically significant) convergence on measures of child temperament (Offord et al., 1996). Similarly, teachers and children (Dadds, Spence, Holland, Barrett, & Laurens, 1997), and parents and children, show low rates of convergence on anxiety-type measures (Rapee, Barrett, Dadds, & Evans, 1994). Further, the younger the child, the lower the convergence. Thus, at best, we would expect low (but statistically significant) convergence comparing these young children's reports with other raters.

On the basis of the results obtained in Study 1, a second study was undertaken. In this study a refined version of the puppet interview was developed in order that the Shyness subscale not be contaminated by items that related to sociability. An additional Social Desirability subscale was developed on the basis of social desirability items from the revised Children's Manifest Anxiety Scale (Reynolds & Richmond, 1979). It was also determined, on the basis of the results obtained in Study One, that the observation methods used were not valid measures for preschool children and, therefore, were not included in Study 2.

STUDY 1

Method

Participants

Within the metropolitan area of Brisbane, Australia, three state government preschools and one private Catholic preschool participated in the study. Schools were drawn from a range of socioeconomic levels with classifications ranging from working class to upper-middle class. The children were predominantly Caucasian. Of the 132 families in these four schools, 118 (90%) consented to involve their children in the puppet interview. Nine children were absent at the time of the assessment, and one boy was excluded because he could not speak English, leaving a sample of 108 children aged between 50 months and 69 months, including the 30 children in the pilot phase. Girls comprised 44% of the final sample with a mean age of 4.8 years ($SD = .33$), and boys (56%) had a mean age of 4.9 years ($SD = .35$). The children had been at the preschools for between 5 and 7 months at the time of the study, in order to ensure that teachers were adequately familiar with the children when responding to questionnaires.

Measures

Parent and Teacher Reports

The Colorado Childhood Temperament Inventory (CCTI; Buss & Plomin, 1984), designed for children aged 1 to 6, examines maternal perceptions of children's characteristics (Sociability, Shyness, Emotionality, Soothability, Persistence, Activity) using a 5-point Likert scale. Internal consistency and test-retest reliabilities were reported as moderately high (.73–.88) for all scales with the exception of Soothability.

The Preschool Play Behaviour Scale (PPBS; Coplan & Rubin, 1998) is an 18-item teacher rating scale, comprised of five factors: Social Play, Reticence, Solitary-passive, Solitary-active, and Rough-play Behaviours. A strong relationship was found between reticence as measured by the PPBS and internalising problems. Internal consistency estimates for the factors ranged from .76 to .96 (Coplan & Rubin, 1998). Moderate to high stability of the subscales was observed across 6 months. Convergent validity was significantly demonstrated between the PPBS and the CCTI.

Puppet Interview: Conceptual Basis, Structure, and Item Selection

Items for the puppet interview (PI) were based upon the CCTI (Buss & Plomin, 1984). The items were simple self-descriptive opposing statements, describing activities that were hypothesised to be reflective of temperamental constructs. The statements were designed to assess children's perceptions of their own behavioural expressions of traits. For example, "I have lots of friends to talk to" versus "I don't have many friends to talk to" is a Sociability item. Activity descriptions were designed to circumvent, as much as possible, variance due to differences in interpretation.

A total of 38 pairs of statements comprised the five subscales of the pilot PI: Sociability (9 items), Shyness (8 items), Emotionality (7 items), Soothability (7 items), Activity (7 items). These original 38 items were tested with a subsample of 30 children. Items with poor internal consistency correlations were deleted, leaving 24 items to be used in the PI, which was reduced to four subscales: Sociability (7 items), Shyness (5 items), Emotionality (6 items), and Soothability

(6 items). The Activity items resulted in a negative alpha ($-.80$), and this scale was deleted from the Puppet Interview.

The puppets used were two identical teddy bear hand puppets. Teddy bears were chosen because of their unisex nature and because both boys and girls like teddy bears. Literature also shows that children respond better to soft and cuddly animals such as bears (Caputo, 1993). A small puppet theatre, with teddy bear pictures on it, was constructed to allow the puppeteer to sit behind and hide most of her body, thus preventing body language from influencing the children's responses.

Procedure

Two young female researchers introduced themselves and the puppets to the children in the early morning "show and tell session" in the classroom. Each of the puppets said "Hello boys and girls" and asked "Do you like teddy bears?" This was followed by an excited "Yes!" from most children. The researcher explained that some children would be playing a game with the teddy bear puppets all by themselves in the teacher's office, but not everyone would be able to play with the puppets because "The teddy bears just get too tired to see so many kids. However, children who did not have a turn in the office could say 'hello' to the puppets in a special group puppet show just for them". Consequently, after all children with parental consent had finished the PI the researchers performed a special puppet show for the other children.

Children with parental consent were individually asked if they would like to play with the puppets. Only four children declined to play with the puppets. Once the children were escorted to the office they sat in a designated chair and again were introduced to the researchers. The children were told that one adult (the independent observer) was there to watch the puppets because she liked teddy bears. When the puppeteer was ready the teddy bear puppets appeared in the puppet theatre and each said "Hello (child's name)". The children were asked by the puppeteer if they were ready to play with the puppets. A "yes" response was taken as voluntary consent. The instructions given to the children were as follows: "In this special game, the teddy bears will tell you about themselves. (pause until the child's eyes focus on both teddy bears) I want you to listen to what the teddy bears are saying (pause) and then choose the teddy that's most like you. (pause). Remember, you have to listen to what the teddy bears are saying and then choose the teddy bear that's most like you".

A practice session ensured that the children understood the forced-choice format. There were three paired-items in the practice session and getting all three correct denoted that the child understood the forced-choice format. The practice always began with "I go to preschool" versus "I don't go to preschool". The puppeteer then asked "Which teddy bear are you most like?" Children could respond by speaking or pointing. The second and third practice statements were related to what the child was wearing, such as "I am wearing shoes" versus "I am not wearing shoes". If children did not understand after the initial three practice pairs, more items were presented until the child responded correctly to three consecutive practice pairs. It is important to note, that practice items had objective, observable criteria (e.g., shoes/no shoes), whereas the Puppet Interview items were subjective (e.g., I like/I don't like).

The PI, consisting of 24 pairs of statements, immediately followed the practice session. To avoid children identifying or favouring either the left- or right-hand puppet, alternate hands began each pair of statements, thereby alternating the more socially desirable statements between puppets. Voice tone remained equally

enthusiastic for each statement to offset approval-seeking by the children. After each pair of statements the puppeteer would lift her head up above the puppet theatre and ask the child, "Which one are you more like"? (child responds) "OK". The puppeteer periodically gave encouraging statements such as "You are doing very well", "Just keep going". If the child appeared to be losing concentration s/he was encouraged to put on her or his "special imaginary listening ears". The interview usually lasted between 10 and 15 minutes, depending upon the number of practice sessions. The puppeteer praised the child's efforts and thanked the child for his or her help. Each child received a star sticker before returning to class.

Observations

Observations of the children occurred simultaneously with the PI. An independent observer sat on a chair outside of the child's view. The observer scored the child's forced-choice response, as well as recording latency to respond, measured in seconds using a hand-held stopwatch. Scoring also included whether the child used speaking or pointing to respond to each of the 24 questions, as well as frequency of anxious (i.e., fingers in mouth, touching body parts, biting lip, twisting hair), avoidant (i.e., look away when asked to pay attention, refusal to respond, pulling faces, stamping, or other non-compliance), or distressed (crying, saying "no") styles of responding during each of the 24 questions, using a partial interval time sample. That is, the presence or absence of each category was scored for the interval from the end of the puppeteer question to the start of the next question. There were no distressed responses, and this category was deleted from further analyses. Data for each of the styles of responding were calculated as percentages by multiplying the frequency by 100 and dividing by 24 (the number of questions). A global rating of shyness on a scale of 1 to 5 was assigned by the observer at the conclusion of the PI: (1) spontaneously talkative, with no anxious or avoidant behaviours; (2) somewhat talkative, not anxious or avoidant; (3) few, if any, anxious or avoidant behaviours; (4) not talkative, some anxious or avoidant behaviours; and (5) anxious or avoidant most of the time, and talks very little. Speaking and pointing were mutually exclusive categories. Speaking was only coded if the child repeated all or part of the statement. Only the pointing category will be reported in the results.

Interrater reliability checks occurred for 21% of the interviews by having a second observer present during the PI. Observers coordinated their use of stopwatches and had previously been trained to criterion agreement on definitions of response styles prior to beginning the observations through pre-assessment trials with children outside of the project.

Results

Table 1 shows means and standard deviations for the PI. Using independent paired *t* tests and an alpha level of .05, no significant gender differences were found for the PI.

Internal Consistency of the Puppet Interview

Internal consistency was computed on the 24 items of the PI ($n = 79$). The coefficient alphas reported are equivalent to the Kuder-Richardson 20 coefficients for dichotomously scored items. Alphas for PI scales are reported in Table 2, and overall, suggest moderate internal consistency consistent with previous studies. The alpha levels (.47-.64) were very similar between the 38- and 24-item versions with

TABLE 1
Means and Standard Deviations by Gender for Puppet Interview

	Study 1				Study 2			
	Child		Child		Parent		Teacher	
Puppet Interview	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Sociability								
Boys	5.85	1.40	5.55	1.27	5.83*	1.67	6.07	1.39
Girls	5.90	1.40	5.61	1.44	6.54*	.76	6.23	1.03
Shyness								
Boys	.94	1.12	2.10	1.99	2.83	1.95	2.07	2.17
Girls	1.14	1.11	2.03	1.59	1.50	1.33	1.35	1.70
Emotionality								
Boys	1.38	1.52	1.69	1.39	1.90**	1.68	1.31	1.47
Girls	1.95	1.76	1.57	1.63	.54**	.95	1.35	1.62
Soothability								
Boys	5.05	1.20	4.20*	1.52	5.31	.89	4.76	1.24
Girls	4.50	1.48	5.03*	1.11	5.69	.47	4.77	1.34
Social Desirability								
Boys			5.41	2.69	2.76	1.80	3.86	2.39
Girls			5.42	2.50	2.11	1.68	4.38	2.89

Note: * $p < .05$, ** $p < .01$

only minimal improvement for Sociability and Shyness, and reduced alpha levels for Emotionality and Soothability.

Pearson product moment correlations were used to examine the convergence between the subscales of the PI, split by gender (see Table 3). As hypothesised, negative correlations emerged between Sociability and Shyness, although more strongly for girls, and between Soothability and Emotionality, again more strongly

TABLE 2
Internal Consistency (*a*) and Test-Retest (*r*) for Development of Puppet Interview (PI) and Colorado Childhood Temperament Inventory (CCTI)

Subscale	Study 1				Study 2			
	Pilot	PI		PI — R	PI — Parent	PI — Teacher	CCTI — Parent	CCTI — Teacher
	(<i>n</i> = 30)	(<i>n</i> = 79)	(<i>n</i> = 10)	(<i>n</i> = 55)	(<i>n</i> = 10)	(<i>n</i> = 55)	(<i>n</i> = 55)	(<i>n</i> = 55)
	<i>a</i>	<i>a</i>	<i>r</i>	<i>a</i>	<i>r</i>	<i>a</i>	<i>a</i>	<i>a</i>
Shyness	.41	.47	.31	.60	.29	.63	.75	.84
Sociability	.59	.61	.30	.43	-.33	.72	.62	.68
Emotionality	.67	.64	.59	.56	.55	.70	.72	.72
Soothability	.67	.52	.86	.53	.84***	.19	.51	.84
Social Desirability				.78	.80**	.55	.79	

Note: ** $p < .01$, *** $p < .001$

TABLE 3

Convergence between Child (PI), Parent (CCTI), and Teacher (PPBS) Measures: Study 1

Measure	Subscales	Puppet Interview (PI)							
		Sociability		Shyness		Emotionality		Soothability	
		Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
PI	Sociability	—	—						
	Shyness	-.42	-.59**	—	—				
	Emotionality	-.48**	-.12	.43**	.49**	—	—		
	Soothability	.38*	.51**	-.45**	-.50**	-.35*	-.49**	—	—
Parent CCTI	Sociability	-.01	.35*	.10	-.33	.15	-.24	-.37*	.22
	Shyness	.19	-.01	-.26	-.06	-.16	-.07	.24	-.19
	Emotionality	-.07	.08	-.07	-.09	-.07	.08	.14	-.10
	Soothability	.04	.12	.18	-.09	.07	-.21	-.14	.25
Teacher PPBS	Social Play	-.08	.04	.39*	-.25	.08	-.36*	-.33*	.17
	Reticence	-.01	-.16	-.23	.33	-.30	.21	.26	-.14
	Solitary	-.03	.04	-.37*	-.28	-.24	.00	.29	.15

Note: * $p < .05$, ** $p < .01$

for girls. Consistent with current temperament theory (Buss, 1995), the results also indicated a negative relationship between Emotionality and Sociability, more strongly for boys, and a moderate positive correlation was found between Sociability and Soothability, more strongly for girls.

Test-Retest Reliability

A subsample of 10 children repeated the Puppet Interview after a 1-week period. The test-retest reliability for Soothability ($r = .86$) was high. Test-retest for Emotionality ($r = .59$) was moderate. However, Sociability ($r = .30$) and Shyness ($r = .31$) were lower (see Table 2). Examination of the scores for the 10 children showed that when scores changed at retest, children tended to report themselves as being more social, less shy, and more emotional than at the initial testing.

Construct Validity: Convergent and Discriminant Validity

Correlations shown in Table 3 suggest that the level of agreement between parents, teachers, and children was very low (.00 to .39) with only a few significant correlations. Given the large number of correlations, some of these significant correlations could be due to chance. Therefore, rather than report the significance levels in text, the general patterns of correlations are interpreted. As the sample is split by gender, creating smaller subsamples, the significance level reflects the number in the sample, as well as the relationships between subscales. Therefore, patterns are discussed by categorising correlations as low (up to .40), moderate (.40 to .60), or high (above .60). For readers who wish to use the traditional significance cut-offs, a sample of 50 with a one tailed test at $p = .05$ and $p = .01$ would be $r = .23$ and $r = .33$ respectively (Hoel, 1971).

Convergence between parent reports of children's temperament and children's self-reports was analysed separately for boys and girls. A pattern of low correlations

was anticipated, yet, overall, there was little or no relationship between parents' and children's reports (see Table 3). Similarly, near zero and low correlations were found between children's and teachers' reports, and in the opposite to that expected for Shyness. Thus, there was no clear and tenable relationship between children's reports and teachers' reports.

Observations

There were no significant gender differences for either anxious or avoidant behaviours, pointing or latency to respond to the PI questions. The interrater reliability of the observation categories, including global ratings, was moderate to high. Cohen's kappas were .61 (anxious), .92 (average latency), .92 (percentage pointing) and .87 (global rating). No cases of avoidant behaviour were coded for the interrater observations. Correlations between the observations and other measures were generally low, or showed no relationship with other reports, with the exception of a moderate relationship between girls' self-reports of shyness and observed latency to respond (see Table 4).

Summary

The children's responses showed a moderate degree of internal consistency. Convergence between the constructs of sociability and shyness, emotionality, and soothability within the PI were in the anticipated direction. The test-retest reliability was variable (.30–.86). Finally, convergence between adult raters and children proved to be quite low.

A problem in establishing convergence between measures in this study was the use of different measures for different informants. By creating a common assessment across parents, teachers, and children, a more accurate portrayal of convergence would be expected. This was addressed in Study 2. Also, a larger pool of items may provide more consistent responses (Marsh, Craven, & Debus, 1998). Examination

TABLE 4

Convergence of Puppet Observations with Parent, Teacher, and Child Reports for Shy-type and Social-type Constructs: Study 1

Subscales	Observations									
	% Anxious		% Avoidant		Average Latency		% Point		Global Ratings	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Parent Reports										
Shyness	.15	.13	.07	.26	-.24	.07	.04	.19	-.00	.08
Sociability	.05	-.07	.29	-.26	.06	-.34*	-.29	.18	.17	-.23
Teacher Reports										
Reticence	.32*	.09	.08	.34*	.04	.30	.02	.20	-.04	.07
Social Play	-.15	.03	.01	-.20	.20	-.38**	-.30	-.16	-.06	-.14
Child Reports										
Shyness	.05	-.06	.04	.30*	.35**	.51***	-.12	-.19	-.21	.02
Sociability	.07	.12	-.20	-.02	.04	-.33*	-.15	.03	.25	.07

Note: * $p < .05$, ** $p < .01$, *** $p < .001$

of the predictive validity from child reports versus adult reports would help in understanding relative strengths and weaknesses of both reports. However, this does not overcome the developmental problem that young children may only be capable of reporting on their immediate experiences, rather than long-term trait-like experiences.

Given the low convergence found between children and adults in Study 1, it was hypothesised that a social desirability scale could be used to aid interpretation of the validity of the children's self-reports. Social desirability scores typically measure the tendency to fake-good in a socially desirable direction (Hagborg, 1991). Dadds, Perrin, and Yule (1998) found that social desirability scores interacted with gender to show different patterns of association with adult ratings of children's anxiety. Furthermore, the social desirability scores decreased with age, presumably reflecting normal developmental changes in children's ability to provide accurate self-reports. However, preschoolers' tendency to fake-good was not tested. Therefore, in Study 2, a Social Desirability scale was added to the PI.

To further address the low convergence between raters, identical measures for all informants were introduced and modelled after the forced-choice format of the PI. The independent observations of children's behaviour during the PI were not included in Study 2.

STUDY 2

Method

Participants

Eighty-seven consent forms were distributed between three new preschools and parental consent was obtained for 74 children. Of these participants, five children were excluded from the study due to incomplete parent questionnaires, one child chose not to continue with the interview after completing five items, three children refused to play with the puppets, three children were excluded due to intellectual impairment, a further five children were absent during data collection, and two children had been attending the preschool for less than 4 months. The sample thus consisted of 55 preschool children, including 26 girls and 29 boys. The age range of the girls was 4.0 to 5.9 years ($M = 4.9$, $SD = .55$), and boys' aged ranged from 4.1 to 5.9 years ($M = 5.1$, $SD = .47$). All children were of English-speaking background.

Measures

Parent and Teacher Reports

As for Study 1, the CCTI was included in the assessments, although in Study 2, teachers, as well as parents, completed this form. In addition, a forced-choice questionnaire identical to the children's PI-R was included for parents and teachers. Wording of statements was changed from "I like ..." to "Child likes ..." resulting in a 36-item, pencil and paper Puppet Interview — Parent/Teacher Version (PI-P/T). Scoring was identical to the PI-R.

Puppet Interview—Revised (PI-R): Conceptual basis, Structure, and Item Selection

The PI-R was conceptually identical and structurally similar to the PI used in Study 1. However, two changes distinguished the PI-R from the PI: (a) a new

Shyness scale uncontaminated by sociability items and (b) a Social Desirability scale. In the previous study, Shyness items included a preference for being with others. The criteria for item selection for the Shyness subscale on the PI-R were (a) situations specific to shyness (e.g., meeting strangers); (b) both the affective (tension, worry) and the behavioural (awkward behaviour, gaze aversion) components of shyness would be represented; and (c) no preference for being with people would be made. The following new items were added to the shyness scale: (a) I feel nervous/don't feel nervous when I am with people I don't know well; (b) I like/don't like to talk to new people; and (c) When I meet someone new I look/don't look at them.

The Social Desirability scale consisted of nine social desirability items from the Revised Children's Manifest Anxiety Scale (Reynolds & Richmond, 1979), such as "I always have good manners/ I sometimes have good manners". Thus, the PI-R in Study 2 consisted of 36 pairs of statements, with a score of 1 or 0 for each item to indicate presence of absence of the trait. Overall scores for Sociability, Shyness, Emotionality, Soothability, and Social Desirability scales were obtained by summing scores for the items.

Procedure

As for Study 1.

Results

Gender differences, displayed in Table 1, were examined for the PI-R and PI-P/T using independent paired t tests. Girls reported greater soothability on the PI-R self-report, $t(53) = 2.29, p < .03$. On the PI-P/T, parents reported that girls were more sociable, $t(40) = 2.07, p < .05$, and boys were more emotional, $t(45) = -3.75, p < .01$, (degrees of freedom adjusted for unequal variances). No gender differences were reported by teachers on the PI-P/T.

Internal consistency of the Puppet Interview-Revised

Internal consistency was computed on the 36 items of the PI-R ($N = 55$). The coefficient alphas reported are equivalent to the Kuder-Richardson 20 coefficients for dichotomously scored items. Alphas for PI-R scales are reported in Table 2, and overall, suggest moderate internal consistency, with the Social Desirability scale somewhat higher. The Social Desirability scale will be considered separately. The alpha levels (excluding Social Desirability scale) (.43–.60) were very similar between the PI-R and PI with only minimal improvement for Shyness, and reduced alpha levels for Emotionality and Sociability. The internal consistency of the Social Desirability scale was high for children and teachers, and moderate for parents.

The internal consistency (excluding the Social Desirability scale) of the parents' (PI-P) and teachers' (PI-T) versions tended to be somewhat higher than for children, ranging from .21 to .79 for the PI-P and from .53 to .75 for the PI-T (see Table 2). The only exception was the lower alpha level of Soothability in the parents' reports. Pearson product moment correlations were used to examine the convergence between the subscales of the PI-R and PI-P/T, split by gender. As hypothesised, within raters, negative correlations emerged between Sociability and Shyness (see Table 5) and between Soothability and Emotionality (see Table 6).

Test–Retest Reliability

A subsample of 10 children repeated the Puppet Interview after a 1-week period. The test–retest reliability for Soothability ($r = .86$) and the Social Desirability scale ($r = .80$) was high. Test–retest for Emotionality ($r = .59$) was moderate. However, Shyness ($r = .29$) was substantially lower (see Table 2). A negative reliability coefficient for Sociability ($-.33$) indicates that children were reversing their responses at retest. Overall, this suggests that children responded consistently over 1 week to only one of the four constructs. Thus, reliability shows a similar pattern to the test–retest coefficients of Study 1 (i.e., stability of reports of Soothability, moderate stability of Emotionality, and lack of stability in reports of Shyness and Sociability).

Construct Validity: Convergent and Discriminant Validity

As for Study 1, patterns are discussed by categorising correlations as low (up to .40), moderate (.40 to .60), or high (above .60), with significance levels reported in Tables 5 to 7. The relationship between parents and teachers was examined for the PI-P/T (see Tables 5 and 6). Between raters, moderate correlations were found for Shyness for both boys and girls, and low correlations for Sociability. For Emotionality, a moderate relationship between raters was found for girls only. Parents’ and teachers’ reports showed little or no relationship for Soothability. The relationship between the children’s self-reports (PI-R) and the parents’ and teachers’ reports (PI-P/T) was examined next. Correlations shown in Tables 5 and 6 indicate little or no relationship between child and adult informants, with only a few significant correlations between parent and teacher reports of Shyness. In the context of the generally low convergence observed between adult informants, this finding is not unexpected. However, given the large number of correlations, these few significant correlations could be due to Type I error.

For girls, parents’ and teachers’ reports on the CCTI showed moderate correlations for Shyness and Sociability and no relationship for Emotionality or Soothability. For boys, parents’ and teachers’ reports on the CCTI showed little

TABLE 5

Convergence Between Child, Parent, and Teacher Puppet Interview (PI): Shyness and Sociability Study 2

Measure	PI — Child (PI-R)		PI — Parent				PI — Teacher			
	Sociability	Shyness	Sociability	Shyness	Sociability	Shyness	Sociability	Shyness		
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls		
PI—R										
Sociability	—	—								
Shyness	-.24	-.31	—	—						
PI—Parent										
Sociability	-.14	.20	.20	-.08	—	—				
Shyness	.07	-.08	.08	-.14	-.61**	-.24**	—	—		
PI—Teacher										
Sociability	-.25	.06	.27	.11	.33	.14	-.33	-.49*	—	—
Shyness	.06	.12	.24	-.18	-.27	-.24	.44*	.45*	-.27	-.69***

Note: * $p < .05$ ** $p < .01$ *** $p < .001$

TABLE 6

Convergence Between Child, Parent, and Teacher Puppet Interview (PI):
Emotionality and Soothability Study 2

Measure	PI — Child (PI-R)		PI — Parent				PI — Teacher		
	Sociability		Shyness		Sociability		Shyness		
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	
PI—R									
Emotionality	—	—							
Soothability	-.56*	-.61*	—	—					
PI—Parent									
Emotionality	.06	.36	.13	-.21	—	—			
Soothability	-.12	-.49*	-.08	.48*	-.29	-.24	—	—	
PI—Teacher									
Emotionality	.17	.35	-.16	-.14	.16	.47*	.20	-.17	—
Soothability	.31	-.08	-.33	-.07	-.27	-.23	-.09	.14	-.64* - .70***

Note: * $p < .05$, ** $p < .01$, *** $p < .001$

convergence, except for a high correlation for Shyness (see Table 7). Thus, convergence between parents and teachers on similar constructs was generally low to moderate even for a well-established measure such as the CCTI.

The constructs of Sociability–Shyness and Emotionality–Soothability were expected to be negatively correlated within, as well as between, raters. This

TABLE 7

Convergence Between Parents and Teachers:
Colorado Childhood Temperament Inventory (CCTI)

Measure	CCTI — Parent		Shyness		CCTI — Teacher		Shyness		
	Sociability		Boys	Girls	Boys	Girls	Boys	Girls	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	
CCTI — Parent									
Sociability	—	—							
Shyness	-.51*	-.12	—	—					
CCTI — Teacher									
Sociability	.16	.38	-.49*	-.54*	—	—			
Shyness	-.26	-.18	.64**	.55*	-.67**	-.62**	—	—	
	Emotionality		Soothability		Emotionality		Soothability		
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	
CCTI — Parent									
Emotionality	—	—							
Soothability	-.43	.29	—	—					
CCTI — Teacher									
Emotionality	.13	.07	-.18	-.04	—	—			
Soothability	.00	.01	-.09	-.01	-.65*	-.82*	—	—	

Note: * $p < .05$, ** $p < .01$

hypothesis was confirmed within raters for children and adults on the PI and CCTI (see Tables 5 to 7). However, given the low convergence found on similar constructs between adult raters, the convergence on different constructs between parents, teachers and children was not uniformly reflective of the anticipated relationships. The expected negative relationships were found for Shyness-Sociability between parents and teachers on the PI-P/T and CCTI (Tables 5 and 7). For Emotionality-Soothability a negative relationship was shown between parents and teachers' reports (PI-P/T) for girls, but less clearly for boys, but no relationship was found between the two raters for the two constructs on the CCTI. Children's self-reports were negatively correlated with parents and teachers only for Emotionality-Soothability, more so for girls (see Tables 6 and 7). Correlations for the Social Desirability scale were low between parents and teachers (.12 for boys, .35 for girls), and also low between children and adults (.01 to .20).

DISCUSSION

This study sought to develop and evaluate the Puppet Interview (PI, PI-R), an assessment measure designed to allow young children to report their perceptions of their own temperament. The following discussion will consider the psychometric properties of the interview, including the internal consistency, test-retest reliability, and construct validity found within the PI-R, and through its convergence with parent and teacher reports.

The moderate internal consistency of the PI and PI-R is similar in magnitude to that reported by Eder (1990) and Ablow (1999). Ablow et al. found stronger internal consistency for self-reports of clinic-referred children than for the community sample. The constructs within the children's PI and PI-R all show a moderate relationship with each other and in directions consistent with current temperament theory (Buss, 1995). When parents and teachers completed measures identical to the children's PI, alpha levels were still only moderate, suggesting that the subscales themselves may warrant improvement, both with regard to the adults and children.

There are no clear cut-offs to characterise adequate reliability. Rather, this must be determined within the context of the purpose of the research, the decisions that will be based upon the measure and the construct validity of the measure in relation to other measures as an estimate of the homogeneity of the constructs (Pedhazur & Schmelkin, 1991). Pedhazur and Schmelkin suggest that low to moderate reliability of subscales, as found in the PI, indicate a lack of homogeneity and raises questions about the validity of the measure. Yet, it is also important to consider factors such as the number of items per construct. By artefact, the more items, the greater the alpha levels (Pedhazur & Schmelkin, 1991). Given the exploratory purpose of the measure and the early stages of this research, a greater margin of error may be adopted with regard to reliability (Pedhazur & Schmelkin, 1991).

Typically, assessments for young children are kept short in order to accommodate their relatively short attention span, and prevent fatigue from influencing responses. However, Marsh et al. (1998) found that a longer instrument (64 items) was more effective. The items near the end proved to be more, rather than less, psychometrically sound. Children responded more appropriately towards the end of the instrument than at the beginning. Thus, the common use of short instruments for assessing young children could, in part, contribute to the less than adequate psychometric properties typically associated with measures of child self-report

measures. It is possible that a much longer puppet interview could demonstrate more adequate psychometric properties, although this trend was not evident between the 24-item PI and 36-item PI-R.

In a review, Buss (1995) noted that previous validation studies of the CCTI (parent form) generally reported low negative correlations between shyness and sociability ($-.35$, Asendorph & Meier, 1993; $-.35$, Perry & Buss, 1990 as cited in Buss, 1995; $-.30$, Cheek & Buss, 1981), although one study (Bruch, Gorsky, Collins, & Berger, 1989) noted a correlation of $-.47$ through confirmatory factor analysis. The PI convergence ($-.42$ for boys and $-.58$ for girls), was more in line with the Bruch et al. findings, whereas the PI-R ($-.24$ for boys and $-.31$ for girls) corresponded more to those reported in Buss' review, suggesting that the two traits have a moderate degree of overlap, and are not two clearly distinct constructs. Additionally, emotionality showed the anticipated negative correlations with soothability in both studies. Thus, as an exploratory study with preschool children, the PI-R has shown some construct validity, yet needs further refinement.

Test-retest reliability showed wide variability across the constructs. A number of factors including reporting variance on the part of the children (for instance test-retest attenuation effects), actual changes in the phenomena, or measurement error may contribute to this variability. It is unclear what influence the familiarisation of children with the puppets, the format of the task, the demand characteristics of the task or the presence of experimenters had upon the children's self-report. However, test-retest is commonly low for young children and has shown to improve with age (Marsh et al., 1998; Pellegrini, 1998). It would be worthwhile utilising the present measure in the replication of a previous study (Ablow et al., 1999) in an effort to assess differential rates of responding between clinical and normative samples.

Another way of approaching the low test-retest reliability often demonstrated by young children is from a developmental perspective (Pellegrini, 1998). Temperament is generally conceived of as a trait and, as such, may be expected to demonstrate stability across time. However, it may be that the present measure accounts, at least in part, for state dependent characteristics. For instance, item 24 on the PI-R requires children to discriminate between (a) I have lots of friends to talk to and (b) I don't have many friends to talk to. Young children's responses to items such as this may be significantly influenced by situational or temporary characteristics as opposed to the more stable trait-dependent characteristics. Further investigation is required to ensure that the Puppet Interview is actually measuring stable temperamental characteristics and is relatively free from state-dependent influences.

Convergent validity between children, parents, teachers and trained observers was explored as a means of establishing construct validity. Results of this study suggested that convergence was typically low between all raters. Of note, while convergence between children and adult raters was low, it was generally in the anticipated direction, and greater convergence was demonstrated between girls and adult raters than it was for boys and adult raters.

One explanation for the low convergence between children and adult raters in this study may be that levels are constrained by the low coefficient alphas of the children's subscales (Pedhazur & Schmelkin, 1991). Furthermore, the parents' and teachers' measures showed low to moderate construct validity. Even controlling for Social Desirability scores, there was only minimal improvement on agreement between raters.

Children scored higher on the Social Desirability subscale than parents or teachers, with teachers having the lowest average scores for boys and girls. The Social Desirability subscale showed high internal consistency for both children and teachers, although correlations between the two informants was low, lending support to the notion that young children have a tendency to portray themselves in a socially desirable direction (Dadds et al., 1998). Scores on social desirability scales decrease as children mature, reflecting normal developmental changes in the ability to report realistically rather than idealistically on behavior (see also Wigfield et al., 1997). Results from the self-reports of children in this study support this view, and may be used to inform interpretation of children's self reports of Social Desirability in clinical contexts.

Moderate to high interrater reliability for the observations and global ratings suggest that the behavioural categories were reliably identified and recorded by independent observers. However, there was an overall lack of convergence found, either within the observation categories or between observations and other measures. As few significant correlations emerged in the current study, it was concluded that the observation methods used were problematic for preschool children, and therefore, were not included in Study 2. This conclusion warrants further investigation.

The observations of the PI showed that girls and boys who reported more shyness tended show longer latency to respond, and this relationship was stronger for girls. Asendorph (1993) also found the behavioural indicator of latency to respond correlated with shyness, more so in girls, and this has been further supported by recent research (Eisenberg, Shepard, Fabes, Murphy, & Guthrie, 1998). Asendorph (1993) measured latency to unsolicited utterance, while the current study measured latency to respond to each question. This may have contributed to the lack of convergent validity. Generally, the literature reports that convergence between informants in studies involving children rarely exceed low to moderate levels (Achenbach et al., 1987; Lengua, West, & Sandler, 1998; Prior, 1992; Stevenson-Hinde & Shouldice, 1995). Another contributing factor is the modest level of reliability of the other measures used to assess convergent validity. Thus, it is not surprising that observations during a 10-minute puppet interview, in a strange setting (the teacher's office), with strange people (the researchers), doing a novel task (the PI) would be relatively unrelated to reports of children's behaviour in other more familiar situations, with familiar people.

The results of the present study reinforce previous findings that taking self-report measures from young children is problematic. The use of a puppet interview to assess temperament achieved only moderate psychometric success and further investigation is needed particularly in this regard. At this point, the use of such interviews could not be recommended for clinical or educational settings. Rather, further research is needed to refine the measurement properties within the context of a broader multi-informant assessment. By way of suggestion it would be useful for research to focus on whether the children's reports are providing unique, rather than convergent, information about the child's developmental progress. By using broader outcomes as "gold standards", it may be found that the style and content of young children's reports offer unique added value. Further, further development of procedural aspects of the interview should be considered. It is still possible that further fine-tuning of the interview may further elevate the reliability of children's responses.

References

- Ablow, C., Measelle, J.R., Kraemer, H.C., Harington, 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 Child and Adolescent Psychiatry*, 38(12), 1580–1590.
- Achenbach, T.M. (1995). Developmental issues in assessment, taxonomy, and diagnosis of child and adolescent psychopathology. In D. Cicchetti & D.J. Cohen (Eds.), *Developmental psychopathology Vol 1: Theory and methods* (pp. 57–80). New York: John Wiley.
- Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 101(2), 213–232.
- Adams, S., Kuebli, J., Boyle, P.A., & Fivush, R. (1995). Gender differences in parent-child conversations about past emotions: A longitudinal investigation. *Sex Roles*, 33, 309–323.
- Asendorph, J.B. (1993). Beyond temperament: A two-factorial coping model of the development of inhibition during childhood. In K.H. Rubin & J.B. Asendorph (Eds.), *Social withdrawal, inhibition, and shyness in childhood* (pp. 265–289). Hillsdale, NJ: Erlbaum.
- Asendorph, J.B., & Meier, G.H. (1993). Personality effects on children's speech in everyday life: Sociability-mediated exposure and shyness-mediated reactivity to social situations. *Journal of Personality and Social Psychology*, 64, 1072–1083.
- Bruch, M.A., Gorsky, J.M., Collins, T.M., & Berger, P.A. (1989). Shyness and sociability reexamined: A multicomponent analysis. *Journal of Personality and Social Psychology*, 57(5), 904–915.
- Buss, A.H. (1995). *Personality: Temperament, social behavior, and the self*. Needham Heights, MA: Allyn & Bacon.
- Buss, A.H., & Plomin, R. (1984). *Colorado Childhood Temperament Inventory*. (Available from the Institute of Behavioral Genetics, University of Colorado, Boulder, CO 80309-0447).
- Carroll, J.J., & Steward, M.S. (1984). The role of cognitive development in children's understanding of their own feelings. *Child Development*, 55, 1486–1492.
- Caputo, R.A. (1993). Using puppets with students with emotional and behavioral disorders. *Intervention in School and Clinic*, 29(1), 26–30.
- Cheek, J.M., & Buss, A.H. (1981). Shyness and sociability. *Journal of Personality and Social Psychology*, 41(2), 330–339.
- Coplan, R.J., & Rubin, K.H. (1998). Exploring and assessing nonsocial play in the preschool: The development and validation of the Preschool Play Behavior Scale. *Social Development*, 7(1), 72–91.
- Dadds, M.R., Perrin, S., & Yule, W. (1998). Social desirability and self-reported anxiety in children: An analysis of the RCMAS lie scale. *Journal of Abnormal Child Psychology*, 26(4), 311–317.
- Dadds, M.R., Spence, S.H., Holland, D.E., Barrett, P.M., & Laurens, K.R. (1997). Prevention and early intervention for anxiety disorders: A controlled trial. *Journal of Consulting and Clinical Psychology*, 65(4), 627–635.
- Edelbrock, C., Costello, A., Dulcan, M., Conover, N., & Kalas, R. (1985). Parent-child agreement on child psychiatric symptoms assessed via structured interview. *Journal of Child Psychology and Psychiatry*, 27, 181–191.
- Eder, R.A. (1990). Uncovering young children's psychological selves: Individual and developmental differences. *Child Development*, 61, 849–863.
- Eder, R.A., & Mangelsdorf, S.C. (1997). The emotional basis of early personality development: Implications for the emergent self-concept. In R. Hogan, J. Johnson & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 209–240). San Diego: Academic Press.
- Eisenberg, N., Shepard, S.A., Fabes, R.A., Murphy, B.C., & Guthrie, I.K. (1998). Shyness and children's emotionality, regulation, and coping: Contemporaneous, longitudinal, and across-context relations. *Child Development*, 69(3), 767–790.
- Hagborg, W.J. (1991). The Revised Children's Manifest Anxiety scale and social desirability. *Educational and Psychological Measurement*, 51, 423–427.

- Harris, P.L., Olthof, T., & Terwogt, M.M. (1981). Children's knowledge of emotion. *Journal of Child Psychology and Psychiatry*, 22, 247–261
- Hoel, P.G. (1971). *Elementary statistics* (3rd ed.). New York: John Wiley.
- Lagattuta, K.H., Wellman, H.M. (2002) Differences in early parent-child conversations about negative versus positive emotions: Implications for the development of psychological understanding. *Developmental Psychology*. Vol 38(4) 564–580.
- Lengua, L.J., West, S.G., & Sandler, I.N. (1998). Temperament as a predictor of symptomatology in children: Addressing contamination of measures. *Child Development*, 69(1), 164–181.
- Lugt-Tappeser, H. (1994). Konnen Kinder ihre Angste schildern? *Praxis Kinderpsychologie Kinderpsychiatrie*, 43, 194–203.
- Lugt-Tappeser, H., & Kriependorf, P. (1992). *Das standardisierte Interview zur Erfassung der Angstlichkeit bei Vorschulkindern* (Bericht Nr. 107 aus den Fachbereich Psychologie). Marburg, Germany: Philipps-Universitat.
- Marsh, H.W., Craven, R., & Debus, R. (1998). Structure, stability, and development of young children's self-concepts: A multicohort-multioccasion study. *Child Development*, 69(4), 1030–1053.
- Nannis, E. D., & Cowan, P. A. (1987). Emotional understanding: A matter of age, dimension, and point of view. *Journal of Applied Developmental Psychology*, 8, 289–304.
- Offord, D.R., Boyle, M.H., Racine, Y., Szatmari, P., Fleming, J., Sanford, M., & Lipman, E. L. (1996). Integrating assessment data from multiple informants. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35(8), 1078–1085.
- Pedhazur, E.J., & Schmelkin, L.P. (1991). *Measurement, design and analysis*. Hillsdale, NJ: Lawrence Erlbaum.
- Pellegrini, A.D. (1998). Play and the assessment of young children. In O. N. Saracho & B. Spodek (Eds.), *Multiple perspectives on play in early childhood education* (pp. 220–239). Albany, NY: State University of New York Press.
- Prior, M. (1992). Child temperament. *Journal of Child Psychology and Psychiatry*, 33(1), 249–279.
- Rapee, R.M., Barrett, P.M., Dadds, M.R., & Evans, L. (1994). Reliability of the DSM-III-R childhood anxiety disorders using a structured interview. *Journal of the American Academy of Child and Adolescent Psychiatry*, 33, 384–992.
- Reynolds, C.R., & Richmond, O.B. (1979). What I think and feel: A revised measure of children's manifest anxiety. *Journal of Personality Assessment*, 43, 281–283.
- Robins, L.N. (1985). Epidemiology: Reflections on testing the validity of psychiatric interviews. *Achieves of General Psychiatry*, 42, 918–924.
- Spence, S.H. (1998). A measure of anxiety symptoms among children. *Behaviour Research and Therapy*, 36, 545–566.
- Stevenson-Hinde, J., & Shouldice, A. (1995). 4.5 to 7 years: Fearful behaviour, fears and worries. *Journal of Child Psychology and Psychiatry*, 36(6), 1027–1038.
- Sturgess, J., & Ziviani, J. (1995). Development of a self-report play questionnaire for children aged 5 to 7 years. A preliminary report. *Australian Occupational Therapy Journal*, 42, 107–117.
- Wigfield, A., Eccles, J. S., Yoon, K.S., Harold, R.D., Arbretton, A.J., Freedman-Doan, C., & Blumenfeld, P.C. (1997). Changes in children's competence beliefs and subjective task values across the elementary school years: A three-year study. *Journal of Education Psychology*, 89, 451–469.

APPENDIX**Puppet Interview — Revised**

1. a) I like being with other kids.
b) I like being alone.
2. a) When someone steps on my sandcastle, I cry and cry.
b) When someone steps on my sandcastle, I cry a bit, but then I build it again.
3. a) I feel scared when I have to meet strangers.
b) I don't feel scared when I have to meet strangers.
4. a) It's scary talking to new kids.
b) It's not scary talking to new kids.
5. a) I cry a lot.
b) I don't cry much.
6. a) I never lie.
b) I sometimes lie.
7. a) I like to play games by myself.
b) I like to play games with other kids.
8. a) When I am sad and Mummy or Daddy talks to me, I feel better.
b) When I am sad and Mummy or Daddy talks to me, I still feel sad.
9. a) Meeting new kids makes me worry.
b) Meeting new kids makes me happy.
10. a) I'm happiest when I play with a group of kids.
b) I'm happiest when I play with one kid.
11. a) Nothing much makes me scream and yell.
b) Lots of things make me scream and yell.
12. a) I never say things I shouldn't.
b) I sometimes say things I shouldn't.
13. a) When I see other kids, I want to play.
b) When I see other kids, I don't want to play.
14. a) No one can make me feel better when I cry.
b) My friends can make me feel better when I cry.
15. a) I am happiest when I play by myself.
b) I am happiest when I play with a group of kids.
16. a) I feel nervous when I am with people I don't know well.
b) I don't feel nervous when I am with people I don't know well.
17. a) When someone takes my toy, I cry and cry.
b) When someone takes my toy, I don't cry much.
18. a) I tell the truth every single time.
b) I don't tell the truth every single time.
19. a) I like kids to come to my house to play.
b) I don't like kids to come to my house to play.
20. a) If someone takes my toy, I am mad for a short time.
b) If someone takes my toy, I am mad for a long, long time.
21. a) I don't like to talk to new people.
b) I like to talk to new people.
22. a) When I'm hurt, I don't cry much.
b) When I'm hurt, I cry for a long time.
23. a) I am always good.
b) I am not always good.

APPENDIX continued

24. a) I have lots of friends to talk to.
b) I don't have many friends to talk to.
25. a) I always have good manners.
b) I don't always have good manners.
26. a) I can never stop crying when I want to.
b) I can always stop crying when I want to.
27. a) When I meet someone new, I don't look at them.
b) When I meet someone new, I like to look at them.
28. a) I don't get mad a lot.
b) I get mad a lot.
29. a) I am always kind.
b) Sometimes I am not kind.
30. a) At preschool, I don't have many friends.
b) At preschool, I have a lot of friends.
31. a) I like everyone I know.
b) There are some people I don't like.
32. a) A hug does not make me feel better when I am sad.
b) A hug makes me feel better when I am sad.
33. a) When I fall off the swing, I cry a lot.
b) When I fall off the swing, I don't cry much.
34. a) I am always nice to everyone.
b) Sometimes I am not nice to some people.
35. a) I would like more friends.
b) I would not like more friends.
36. a) I never get angry.
b) I sometimes get angry.