Disponible en ligne sur www.sciencedirect.com

ScienceDirect

Pratiques psychologiques 22 (2016) 61–73

Pratiques psychologiques

www.em-consulte.com

Psychological assessment

The Unfair Card Game: A promising tool to assess externalizing behavior in preschoolers

Le jeu de cartes truqué : un instrument prometteur pour évaluer les comportements externalisés chez le jeune enfant

I. Roskam *, M. Stievenart , E. Brassart , M. Houssa , L. Loop , B. Mouton , A. Volckaert , N. Nader-Grosbois , M.-P. Noël , M.-A. Schelstraete

Psychological sciences research institute, université catholique de Louvain, 10, place du Cardinal-Mercier, 1348 Louvain-la-Neuve, Belgium

Received 17 February 2015; accepted 11 September 2015

Abstract

The assessment of externalizing behavior (EB) in preschoolers is crucial in developmental psychopathology. In the absence of any gold standard measure, new tools contribute to a multi-method and multi-informant approach. The aim of the current study is to present and validate a new observational paradigm, the Unfair Card Game (UCG), intentionally structured to increase the likelihood that negative affect, agitation and inattention will emerge during a video-recorded task. It was administered to 268 young children, and the results were validated by means of factorial analysis, reliability analyses, inter-rater agreement, test–retest, discriminant analyses and external validation with the Child Behavior Checklist (Achenbach & Rescorla, 2004). The validity of the UCG is supported by the findings.

Keywords: Observation; Psychometric properties; Preschoolers; Externalizing behavior

Résumé

L'évaluation des comportements externalisés chez les jeunes enfants est un enjeu crucial en psychopathologie du développement. En l'absence d'un instrument de mesure optimal, la validation de nouveaux dispositifs

* Corresponding author.
E-mail address: isabelle.roskam@uclouvain.be (I. Roskam).

http://dx.doi.org/10.1016/j.prps.2015.09.004

contribue à l'approche multi-méthodes et multi-informateurs. L'objectif de cette étude est de présenter et de valider un nouveau paradigme d'observation, le jeu de carte truqué (UCG). Ce jeu a été conçu pour favoriser l'apparition des affects négatifs, de l'agitation et de l'inattention au cours d'une tâche vidéo filmée. Il a été administré à 268 jeunes enfants et les résultats ont fait l'objet d'une validation par analyse factorielle, indices de consistance interne, accord inter-codeur, fidélité test–retest, analyses discriminantes et d'une validation externe avec le Child Behavior Checklist (Achenbach & Rescorla, 2004). La validité de l'UCG est soutenue par les résultats.

Mots clés : Observation ; Propriétés psychométriques ; Comportements externalisés

1. Introduction

Behavior in young children can be considered as an extremely powerful indicator of well-being and mental health. Behavioral assessment reveals the goodness-of-fit between the child and his/her environment (Chess & Thomas, 1999; Churchill, 2003). In particular, it shows how the child with his/her own features, such as his/her intellectual and cognitive functioning, temperament, somatic and physiological functioning, copes with environmental requirements. And these requirements come from specific settings where the child interacts with caregivers themselves characterized by particular educational and emotional skills, values, expectations and dreams about the child (Super & Harkness, 1986). As such, behavioral assessment can reveal if the child's psychological balance between personal features and environmental requirements is overdue or broken. Or it can inform us that the child is going well. For this reason, its valid assessment is needed in any psychological or psychotherapeutic approach.

From a clinical point of view, some specific behaviors are of primary interest. This is the case for externalizing behavior (EB), which is characterized by agitation, opposition, aggression, provocation, and transgression of social norms (Achenbach & Rescorla, 2004). The identification of EB is very important since EB may interfere with the child's personal, social and academic development (Calkins, Blandon, Williford, & Keane, 2007; Owens & Shaw, 2003). The aim of the current study is to present and validate a promising new observational paradigm of children's EB, the Unfair Card Game (UCG).

As it will be exposed below, behavioral assessment can be obtained in preschoolers with two main methods, i.e. questionnaire and observational paradigm. Each method has both advantages and shortcomings. On the one hand, questionnaires are few time-consuming and they provide information about children's behavior in various daily situations. However, informant's bias can influence the results of questionnaire-based assessment. On the other hand, observations provide rich qualitative and objective information but they are time-consuming and limited to a very specific context. In the absence of a "gold standard measure", it is strongly recommended that a multi-method and multi-informant behavioral assessment be conducted (De Los Reyes & Kazdin, 2005; Kraemer et al., 2003; Roskam, Meunier, & Stievenart, 2013). In this way, UCG is presented here as a supplement to existing questionnaires and observational paradigms.

1.1. Assessment of EB with questionnaires

The most widely employed assessment of EB in preschoolers consists of parents' reports via questionnaires containing EB subscales or questionnaires related to externalized syndromes such as ADHD or conduct disorders. The Child Behavior Checklist (Achenbach & Rescorla, 2004) is one of the most frequently used in research. The Strengths and Difficulties Questionnaire (Goodman, 1997) as well as the Preschool Socio-Affective Profile (LaFreniere, Dumas, Capuano, & Dubeau, 1992) and its short form, the SCBE-30 (LaFreniere & Dumas, 1996), are also commonly used. With regard to the syndrome-oriented questionnaires, among the most widely employed are the Conners rating scales (Conners, 1999; Goyette, Conners, & Ulrich, 1978; Kollins, Epstein, & Conners, 2004) and the SNAP-IV (Bussing et al., 2008; Swanson et al., 2001). The importance of the observation of child behavior both within and outside of the parent–child context has been stressed (Wakschlag et al., 2007). The questionnaires are therefore sometimes also completed by preschool teachers, who also have the opportunity to observe the children in various daily situations (Kerr, Lunkenheimer, & Olson, 2007). However, unlike with school-aged children and adolescents, self- and peer-reports can only be used with extreme caution with young children (Clarke-Stewart, Allhusen, McDowell, Thelen, & Call, 2003). In the rare studies, which have considered self-reports, the preschoolers were interviewed by a trained research assistant who reported their responses on Likert-type scales (Perren, Von Wyl, Stadelmann, Bürgin, & von Klitzing, 2006).

1.2. Assessment of EB through observation

The special importance of observation for assessment of EB during the preschool period has been delineated (Wakschlag et al., 2005). An alternative to questionnaires is provided by observational paradigms. These brief methods are administered in a standardized manner. Rather than life-time symptomatology as in questionnaires, they assess current behavior (Wakschlag et al., 2005). Observational paradigms are intentionally structured to increase the likelihood that a range of clinically relevant EB will emerge. Such paradigms are very helpful for collecting data from clinicians whose diagnostic observation has been demonstrated to provide "added value" beyond that which can be gleaned from parent or teacher information (Wakschlag et al., 2007). A first example is the Disruptive Behavior Diagnostic Observation Schedule (DB-DOS), consisting of a structured laboratory observation (Wakschlag et al., 2007, 2005). The child is observed in several common tasks, such as compliance, clean-up and frustration tasks with the examiner and with the parent (Wakschlag et al., 2005). As in the DB-DOS, the Mother–Child Interaction Task (MCIT) consists of a structured parent–child interaction, which involves a series of episodes of free play, clean-up and teaching tasks, as well as a separation and a reunion (Crowell & Feldman, 1988; Crowell, Feldman, & Ginsberg, 1988). The coding system is made up of both child and mother scales assessing for example irritability and aggression. As a third example, the Snap Game (Hughes et al., 2002) is a specific brief context of peer interaction with the child. Since children with EB are known to display overt anxiety and frustration in competition with a familiar peer (Underwood, Hurley, Johanson, & Mosley, 1999), the children are asked during a school visit to pick one of their classmates to play with. The Snap Game consists of a rigged competitive card game between two children designed to expose them to the threat of losing. Success or failure in competitive play is important for children. The Snap Game has been designed to elicit spontaneous agitation, negative affect and aggression in a realistic context. However, the most important limitation of this

observational paradigm is the influence of the classmate on the target child's EB (Meunier et al., 2011).

1.3. Correlations between methods of assessment of EB

Limited agreement between informants and methods has been consistently reported in previous literature (Achenbach, McConaughy, & Howell, 1987; Collishaw, Goodman, Ford, Rabe-Hesketh, & Pickles, 2009; Gross, Fogg, Garvey, & Julion, 2004; Roskam et al., 2013). For example, correlations measured between the DB-DOS and an impairment questionnaire for children from preschool to adolescence, the Children's Global Assessment Scale (C-GAS) (Lavigne et al., 1998; Setterberg, Bird, Gould, Schaffer, & Fisher, 1992; Wakschlag & Keenan, 2001) completed by the mothers ranged from –.12 and –.21 (Wakschlag et al., 2008). Also, the correlations between the Snap Game and the CBCL (aggression, delinquency and externalized behavior scales) ranged from .09 to .16 when completed by mothers and from .16 to .21 when completed by teachers (Hughes et al., 2002). Similarly, the correlation between three of the MCIT child scales (positive affect, anger and effortful control) and the mothers' CBCL (externalized behavior scale) ranged from –.10 to .02 (Robinson et al., 2009).

Variations in EB according to informants and methods of assessment have sometimes been considered as reflecting informant bias or methodological problems. Actually they can be related to several other explanations (De Los Reyes, Henry, Tolan, & Wakschlag, 2009; Wakschlag et al., 2005). First, they can be due to real variations of EB across contexts. Children's behavior in the family setting is not exactly the same as in the school context for example. Second, they may be explained by the focus of the instruments on EB in daily life in the case of the questionnaires and on current EB in the case of the observations. Third, they are due to informants' subjectivity and the quality of the relationships with the child whom they are currently assessing (Kinoo et al., 2009; Roskam et al., 2010). Each source of information being parents' or teachers' reports with questionnaires or clinicians' observations actually provides unique information, indicating the importance of having more than one informant and method when gathering data about preschoolers' EB (Roskam et al., 2013). The Unfair Card Game is therefore presented here as a supplement to existing instruments.

1.4. The Unfair Card Game (UCG)

The UCG is inspired by an adult paradigm focusing on perspective-taking (Bukowski & Samson, 2012). Like the Snap Game, it is based on a context of peer interaction, except that in this case, the child faces a virtual peer. The UCG can thus be played in a range of contexts, such as at home, at school or in a lab session. The flexibility in its administration is a great advantage for both researchers and clinicians over the DB-DOS and the MCIT, which need to be administered in a lab, and over the Snap Game, which need to be administered in the school context.

The UCG is a computerized version of a moving card game. Two cards appear on the screen, one of which contains the image of a candy. The two cards flip over and start moving. When they stop, the child has to point to the one with the candy picture. The child is invited to play with a virtual partner (a child of the same age range): each time the child points to the correct card, a real candy is 'given' to the virtual partner and, similarly, each time the virtual partner points to the correct card, a real candy is given to the child. The child starts by playing five rounds of the moving card game, and the virtual partner then plays another five rounds. The trick is that the

child will always be correct, thus allowing the virtual partner to win five candies (winning phase) whereas the virtual partner will be a poor player and allow the child to receive one candy only (losing phase).

1.5. Current study

The validation of the UCG paradigm was the main objective of the current research. Validation was based on classical factorial analysis, reliability analyses, and inter-rater agreements. Comparisons between the winning and the losing phases of the game were also conducted in order to demonstrate the ability of the UCG to elicit frustration so as to improve the likelihood of EB in the second phase, which is the core principle of such observation assessment tools. Furthermore, as observation assessment often relies on a limited sample of child behavior, which is considered to be a great advantage both in research and clinical contexts, it is necessary to validate such an assessment procedure with test–retest reliability analyses. Discriminant analyses were computed in order to test the UCG's sensitivity to age- and gender-related differences. Higher EB was expected among boys and among younger participants (Bartels et al., 2004; Bayer et al., 2012; Bongers, Koot, van der Ende, & Verhulst, 2004; Ensor, Hart, Jacobs, & Hughes, 2011). External validation was conducted with the preschool version of the CBCL (Achenbach & Rescorla, 2004). In line with previous research on the agreement between informants and methods (e.g. Achenbach et al., 1987; Roskam et al., 2013), only modest to moderate agreement between UCG and the CBCL was expected.

2. Method

2.1. Sample

This study was part of the Hard-t(w)o-Manage (H2M) Children research program conducted at the university of Louvain in Belgium which received the approval of the Ethics Committee of the Psychological Sciences Research Institute. Data were collected from a community sample of 268 children (59% girls) aged from 45 to 75 months ($M = 59.5$, $SD = 7.09$) in the French-speaking part of Belgium.

2.2. Procedure and analysis strategy

Four research assistants in the Psychological Sciences Research Institute of the university of Louvain who had been intensively trained in sampling and data collection procedures, undertook the data collection. Parents were informed about the research program through leaflets distributed by surrounding schools, posters and a website and Facebook page created for this study. The UCG was administered to 139 children (52%) during a lab session and to 129 (48%) during a school visit. The factorial analysis, reliability analyses and discriminant analyses were computed on this sample ($n = 268$). The test–retest analysis was performed with a subsample of 51 participants. External validation was conducted on a subsample of 201 children whose mothers were asked to complete the preschool version of the CBCL (Achenbach & Rescorla, 2004).

2.3. Instruments

2.3.1. Instrument to be validated

The Unfair Card Game (UCG) is a computerized game lasting approximately 10 minutes. It has been designed to elicit spontaneous agitation, inattention, negative and positive affect in the context of play interaction with a virtual peer. The complete progress of the UCG is presented in details in Appendix A. It contains the complete instructions given by the virtual examiner who have been previously recorded, by the real examiner who is present with the target child, and by the virtual peer who have also been previously recorded. The Appendix A further contains the standardized comments given by the real examiner as well as the behaviors which are coded during the different rounds of the game.

The game starts with instructions presented to the child by a virtual examiner (a previously video-recorded adult). In the first phase, the child plays five rounds of the game. The game is designed to ensure that the child is successful in every round, allowing the virtual partner to win five candies. Accordingly, after each round, the child puts a candy in a bowl containing a photo of the virtual peer. The real examiner's role is strictly standardized. At the end of each round, he/she verbally reinforces the child for his/her good performance and comments on how this is good news for the virtual peer, who has received the maximum number of candies.

In the first phase of the game following the instructions, the child watches the virtual peer playing five rounds of the UCG game. However, his performance is poor, and the child consequently gains only one candy, which the real examiner puts into a bowl with the child's name on it. In this second phase, the examiner's role is also strictly standardized. At the end of each round, he/she verbally comments on the poor performance of the virtual peer and the low number of candies being won for the child. At the end of the UCG, the virtual peer apologizes for his bad performance and proposes to split the gains equally with three candies each.

The administration of the UCG is video-recorded and coded following standardized guidelines set out in a manual (Brassart et al., 2012). The child's behavior is coded according to four scales, i.e. positive affect, negative affect, agitation and inattention. The coding of these scales is made with regard to both the frequency and the intensity of child's behavior on 5-point Likert scales ranging from 1 (neither frequent nor intense) to 5 (very frequent and intense). The four scales are coded once for all the rounds during the first phase in which the child is successful. They are then coded for each of the four rounds in the second phase of the UCG in which the child observes the virtual peer completing the task unsuccessfully (see Appendix A). For example, positive affect is coded 1, i.e. no positive affect, if the child is neutral or displays negative affect and 5, i.e. intense positive affect, if the child smiles all the time during the round and laughs on at least one occasion. Negative affect is coded 2, i.e. low negative affect, if the child exhibits angry behavior once or clutches his/her head once, and 4, i.e. high negative affect, if angry behavior lasts for at least half the round. Agitation is coded 3, i.e. moderate agitation, if the child shows two behaviors such as wriggling in his/her chair, biting his/her lips or getting physically closer to the screen. As a final example, inattention is coded as 3, i.e. moderate inattention, if the child displays three episodes of distraction in the same round, such as staring into space or turning away physically from the task.

In sum, the UCG provides ordinal scores ranging from 1 to 5 for positive affect, negative affect, agitation and inattention during the first phase of the game, i.e. the winning rounds and during each of the four losing rounds of the second phase of the game. The psychometric properties of the UCG are extensively presented and commented on in the results and discussion sections.

2.3.2. External validity

The preschool version of the Child Behavior Checklist (CBCL) (Achenbach & Rescorla, 2004) was used as another assessment of EB. It was completed by the mothers ($n = 201$). The CBCL consists in an original set of 100 items in the form of affirmatives as "Destroys things belonging to his/her family or other children" or "Too shy or timid" for example. The CBCL provides 3-point Likert scales: not at all present, moderately present, or often present. It is organized according to seven first order scales encompassing Emotionally Reactive, Anxious Depressed, Somatic Complaints, Withdrawn, Sleep problems, Attention problems, and Aggressive Behavior, and two-second order scales, i.e. Internalizing Behaviors (Emotionally Reactive, Anxious Depressed, Somatic Complaints, and Withdrawn), and Externalizing Behaviors (Attention problems, and Aggressive Behavior). The CBCL has been widely used in clinical and developmental studies (Roskam et al., 2014) as well as for cross-informant and cross-cultural research (Achenbach et al., 2008). For the external validation of the UCG and to remain within the scope of this study, we particularly focused on the "Attention Problems" and "Aggressive Behavior" scales included in the second order "externalizing behavior" scale. In our sample, the internal consistency was good, with $\alpha = .88$ for the "aggressive behavior" scale, $\alpha = .72$ for the "attention problems" scale, and $\alpha = .89$ for the "externalizing behavior" scale.

3. Results

3.1. Factorial and reliability analyses

Since the four scales are coded once for all in the first phase of the game, Exploratory Factorial Analysis (EFA) was performed only in the second phase. EFA was computed with principal axis factoring and varimax rotation on the 16 scores provided by the coders, i.e. positive affect, negative affect, agitation and inattention in the four losing rounds. Four factors were extracted that fit the four scales perfectly, explaining 45.20% of the variance. Positive affect explained 13.92% of the variance, agitation 11.04%, negative affect 10.22% and inattention 10.01%. All the rounds loaded on the expected factor (> .37) with no cross loadings. Table 1 presents the results of the EFA as well as the results of the reliability analysis. In sum, the four scales, i.e. positive affect, agitation, negative affect and inattention, are clearly delineated and appear to be consistent.

3.2. Inter-rater agreements

Ten percent of the video-recorded UCGs were coded separately by two trained independent coders. The agreement between the two coders was first computed with intraclass correlations. For the winning rounds, it was .84 for positive affect, .89 for inattention and .52 for agitation. Note that in the winning phase of the game for the subsample under consideration, negative affect was consistently coded as 1 on the 5-point Likert scale in the subsample considered for the inter-rater agreement. The intraclass correlation was therefore impossible to calculate. For the losing rounds, it was .94 for negative affect, .89 for inattention, .84 for positive affect and .77 for agitation. In addition to the intraclass correlations, a weighted Kappa coefficient was also computed for the coding by the two independent coders of the four observations of positive affect, negative affect, agitation and inattention during the winning phase and the 16 observations of positive affect, negative affect, agitation and inattention during the four losing rounds. This weighted Kappa coefficient was .72. Note that a four-hour training session was necessary to ensure that the coders

Table 1
Results of the Exploratory Factorial Analysis (EFA) and Cronbach's α.

	Factor 1	Factor 2	Factor 3	Factor 4
Positive affect round 2	.83			
Positive affect round 3	.79			
Positive affect round 4	.71			
Positive affect round 1	.61			
Agitation round 2		.74		
Agitation round 3		.67		
Agitation round 1		.62		
Agitation round 4		.51		
Negative affect round 3			.73	
Negative affect round 4			.72	
Negative affect round 2			.58	
Negative affect round 1			.38	
Inattention round 1				.66
Inattention round 2				.66
Inattention round 3				.62
Inattention round 4				.49
Eigenvalues	3.53	2.28	1.90	1.63
Cronbach's α	.82	.71	.75	.71

Table 2
Descriptive statistics for the winning and the losing phases of the Unfair Card Game (UCG), results of *t*-tests and effect size.

	Winning rounds	Losing rounds	*t*(264)	Cohen's *d*
Positive affect	1.74 (.73)	1.49 (.50)	6.36***	.39
Agitation	1.61 (.73)	1.83 (.59)	–5.03***	.12
Negative affect	1.20 (.56)	1.58 (.50)	–9.05***	.71
Inattention	1.67 (.84)	1.90 (.66)	–4.97***	.30

*** $p < .001$.

were reliable. In sum, the results of the inter-rater agreement analyses show that the coding system ensure that all independent coders report the same behaviors avoiding informant bias.

3.3. Comparisons between the winning and losing phases of the game

t-test comparisons between the two subsequent phases of the game were conducted to demonstrate the ability of the UCG to elicit frustration so as to improve the likelihood of EB in the second phase, which is the core principle of this observation assessment tool. As expected, positive affects significantly decrease in the losing rounds in comparison with the winning ones, and negative affect, agitation and inattention significantly increase between the first and the second phase of the game. The results of the comparison as well as effect sizes are presented in Table 2. In line with these results which confirm that UCG elicits EB in the second phase, i.e. lower positive affect but higher negative affect, agitation, and inattention, only the four scores obtained in the losing rounds have to be considered as an assessment of EB. Only these four scores were therefore under consideration in the subsequent analyses. In sum, the comparisons between the winning

Table 3
Descriptive statistics for the Unfair Card Game (UCG) scores according to age and gender.

	Younger children *M* (*SD*) (*n* = 129)	Older children *M* (*SD*) (*n* = 139)	Girls *M* (*SD*) (*n* = 159)	Boys *M* (*SD*) (*n* = 110)
Losing rounds				
Positive affect	1.39 (.48)[a]	1.58 (.50)[a]	1.44 (.47)[a]	1.56 (.53)[a]
Agitation	1.80 (.61)	1.86 (.58)	1.81 (.57)	1.86 (.63)
Negative affect	1.56 (.48)	1.60 (.52)	1.57 (.49)	1.60 (.51)
Inattention	2.01 (.66)[a]	1.80 (.64)[a]	1.91 (.64)	1.89 (.68)

[a] Means are significantly different.

and losing phases of the game show that the UCG provides a relevant situation for observing EB in preschoolers.

3.4. Test–retest

Test–retest reliability analyses were conducted with a subsample of 51 children. These analyses were performed after an interval of eight weeks, as it does not make sense to expose the child to a rigged game twice in a short period. Eight weeks corresponds to the time these children were on a waiting list to participate in another part of the research program. The correlations that were displayed were moderate, with .26, $p = .06$, for positive affect, .42, $p = .002$, for negative affect, .41, $p = .003$, for agitation, and .27, $p = .05$, for inattention. In conclusion, test–retest analyses indicate that EB was moderately stable over time when it is observed in the UCG setting. In other words, children's behaviors coded at baseline are not exactly the same than those coded eight weeks later. Change may be due to the familiarity with both the task and the examiner but also to developmental change in children's behavior (Meunier et al., 2011).

3.5. Discriminant analyses

t-test comparisons were computed in order to test whether age- and gender-related differences could be found for the UCG. The children's age in months was dichotomized in order to create two groups containing 50% of the sample each. The first group ranged from 45 to 58 months of age; the second group from 59 to 75 months of age. In line with the good factorial analysis, the scores in positive and negative affect, agitation and inattention were averaged from the four losing rounds. Table 3 presents the descriptive data according to age and gender. In line with the expectation, some EBs were more pronounced among younger participants and girls: the results show that during the losing rounds, younger children displayed lower positive affect, $t(262) = -3.04$, $p < .01$, but higher inattention, $t(262) = 2.50$, $p < .05$, than older ones. Boys also displayed higher positive affect than girls, $t(262) = -3.04$, $p < .01$. In sum, the discriminant analyses show that UCG is sensitive to age- and gender-related differences in children's behavior as they were expected in the literature review (Bartels et al., 2004; Bayer et al., 2012; Bongers et al., 2004; Ensor et al., 2011).

3.6. External validation

Correlations were calculated between the UCG and the mothers' CBCL. The correlations were in the same range as previously published interrelations. Aggression was negatively related to

positive affect ($r = -.19$, $p < .01$), and positively related to agitation ($r = .16$, $p < .05$); attention was related to negative affect ($r = .12$, $p < .10$) and agitation ($r = .15$, $p < .05$); externalizing behavior was negatively related to positive affect ($r = -.18$, $p < .01$) and positively to agitation ($r = .16$, $p < .05$). To conclude for external validation, these analyses confirm that the correlations are in the expected direction between the two behavioral assessments suggesting a certain conceptual overlap. However, they show that variations in behavioral assessment happen according to the method employed and the informant involved. This supports the strong recommendation that the best way to assess children's EB is a multi-informant and multi-method strategy as suggested in the introduction section (Roskam et al., 2010, 2011).

4. Discussion

The main aim of the current study was to validate the UCG as a tool to assess EB in preschoolers. The rationale for the development of new assessment methods especially for preschoolers has been presented in the introduction section. Some evidence in favor of the validity of the UCG has been presented in the results section. This will be summarized and discussed before stressing the limits of the current study.

In line with the general principle of observational paradigms which are intentionally structured to increase the likelihood that a range of relevant behaviors will emerge, significant differences were found between the winning and the losing phases of the game, with a decrease in positive affect and an increase in negative affect, agitation and inattention as expected. The factorial analysis was conducted on the coding of the four scales, i.e. positive affect, negative affect, agitation and inattention, in each of the four losing rounds (second phase). It clearly demonstrated that such coding leads to consistent and reliable mean scores explaining a significant amount of variance in the child's behavior. Furthermore, based on video-recording and guidelines in a manual, the UCG scores from the losing as well as from the winning phases were also found to be close to each other when they were provided by independent trained raters. Test–retest showed that over an eight-week interval, which was considered as acceptable for the administration of the same rigged game, the scores were moderately stable. Finally, the discriminant analyses showed that some of the UCG scales were sensitive to age-related differences and to gender-related ones. Differences were in the expected direction, with EB decreasing with age (higher positive affect and lower inattention) and boys displaying higher EB (lower positive affect) than girls.

With regard to the external validation with the questionnaire, we found modest to moderate correlations that were in the same range as those reported in previous studies (Achenbach et al., 1987; Hughes et al., 2002; Roskam et al., 2013; Wakschlag & Keenan, 2001). The correlations were also in the expected direction, with positive affect in the UCG being negatively related to aggressive behavior and externalizing behavior in the CBCL, with agitation being positively related to the three CBCL scales, i.e. attention problems, aggressive behavior and externalizing behavior, and with negative affect being marginally related to attention problems. Unfortunately, the UCG inattention scale did not correlate with any of the three CBCL subscales. It could be that inattention behavior as coded in the UCG was more related to lack of persistence in the game, withdrawal and non-compliance than to executive functioning.

In conclusion, the UCG can be considered as a valid observation method for the assessment of preschoolers' EB, i.e. low positive affect, high negative affect, agitation and inattention when exposed to frustration in the losing rounds. It offers an additional way to rate the target behavior that can be useful both for clinicians and researchers thanks to its flexible administration. Since there is no gold standard measure of children's EB, the UCG has been presented as a supplementary

instrument that can be easily combined with assessment by parents and teachers as well as with alternative methods such as behavior checklists. Its applicability should be tested in future research with clinical samples of children referred for EB but also in different cultural contexts.

Disclosure of interest

The authors declare that they have no competing interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.prps.2015.09.004.

References

Achenbach, T. M., Becker, A., Döpfner, M., Heiervang, E., Roessner, V., Steinhausen, H.-C., et al. (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: Research findings, applications, and future directions. *Journal of Child Psychology and Psychiatry*, *49*, 251–275.

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*, 213–232.

Achenbach, T. M., & Rescorla, L. A. (2004). *The Achenbach System of Empirically Based Assessment (ASEBA) for ages 1.5 to 18 years. The use of psychological testing for treatment planning and outcomes assessment: Volume 2: Instruments for children and adolescents* (3rd ed, pp. 179–213). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Bartels, M., van den Oord, E. J. C. G., Hudziak, J. J., Rietveld, M. J. H., van Beijsterveldt, C. E. M., & Boomsma, D. I. (2004). Genetic and environmental mechanisms underlying stability and change in problem behaviors at ages 3, 7, 10, and 12. *Developmental Psychology*, *40*, 852–867.

Bayer, J. K., Ukoumunne, O. C., Mathers, M., Wake, M., Abdi, N., & Hiscock, H. (2012). Development of children's internalising and externalising problems from infancy to five years of age. *Australian and New Zealand Journal of Psychiatry*, *46*, 659–668.

Bongers, I. L., Koot, H. M., van der Ende, J., & Verhulst, F. C. (2004). Developmental trajectories of externalizing behaviors in childhood and adolescence. *Child Development*, *75*, 1523–1537.

Brassart, E., Houssa, M., Mouton, B., Volckaert, A., Nader-Grosbois, N., Noël, M.-P., et al. (2012). *Coding system of the Unfair Card Game (UCG)*. Louvain-la-Neuve: Université catholique de Louvain.

Bukowski, H. B., & Samson, D. (2012). *Effects of one's emotional state on perspective-taking and prosocial behavior*. Paper presented at the Annual meeting of the Society for Social Neuroscience New Orleans, USA.

Bussing, R., Fernandez, M., Harwood, M., Hou, W., Garvan, C. W., Eyberg, S. M., et al. (2008). Parent and teacher SNAP-IV ratings of attention deficit hyperactivity disorder symptoms: Psychometric properties and normative ratings from a school district sample. *Assessment*, *15*, 317–328.

Calkins, S. D., Blandon, A. Y., Williford, A. P., & Keane, S. P. (2007). Biological, behavioral, and relational levels of resilience in the context of risk for early childhood behavior problems. *Development and Psychopathology*, *19*, 675–700.

Chess, S., & Thomas, A. (1999). Goodness of fit: Clinical applications from infancy through adult life. *Adolescence*, *34*, 800.

Churchill, S. L. (2003). Goodness-of-fit in early childhood settings. *Early Childhood Education Journal*, *31*, 113–118.

Clarke-Stewart, K. A., Allhusen, V. D., McDowell, D. J., Thelen, L., & Call, J. D. (2003). Identifying psychological problems in young children: How do mothers compare with child psychiatrists? *Journal of Applied Developmental Psychology*, *23*, 589–624.

Collishaw, S., Goodman, R., Ford, T., Rabe-Hesketh, S., & Pickles, A. (2009). How far are associations between child, family and community factors and child psychopathology informant-specific and informant-general? *Journal of Child Psychology and Psychiatry*, *50*, 571–580.

Conners, C. K. (1999). *Conners Rating Scales-Revised*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Crowell, J., & Feldman, S. (1988). Mother's internal models of relationships and children's behavioral and developmental status in mother–child interaction: A study of mother–child interaction. *Child Development*, *59*, 1275–1285.

Crowell, J., Feldman, S., & Ginsberg, N. (1988). Assessment of mother–child interaction in preschoolers with behavior problems. *Journal of the American Academy of Child & Adolescent Psychiatry*, *27*, 303–311.

De Los Reyes, A., Henry, D. B., Tolan, P. H., & Wakschlag, L. S. (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology*, *37*, 637–652.

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, *131*, 483–509.

Ensor, R., Hart, M., Jacobs, L., & Hughes, C. (2011). Gender differences in children's problem behaviours in competitive play with friends. *British Journal of Developmental Psychology*, *29*, 176–187.

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*, 581–586.

Goyette, C. H., Conners, C. K., & Ulrich, R. F. (1978). Normative data on revised Conners parent and teacher rating scales. *Journal of Abnormal Child Psychology*, *6*, 221–236.

Gross, D., Fogg, L., Garvey, C., & Julion, W. (2004). Behavior problems in young children: An analysis of cross-informant agreements and disagreements. *Research in Nursing & Health*, *27*, 413–425.

Hughes, C., Oksanen, H., Taylor, A., Jackson, J., Murray, L., Caspi, A., et al. (2002). "I'm gonna beat you!" SNAP!: An observational paradigm for assessing young children's disruptive behaviour in competitive play. *Journal of Child Psychology and Psychiatry*, *43*, 507–516.

Kerr, D. C. R., Lunkenheimer, E. S., & Olson, S. L. (2007). Assessment of child problem behaviors by multiple informants: A longitudinal study from preschool to school entry. *Journal of Child Psychology and Psychiatry*, *48*, 967–975.

Kinoo, P., Meunier, J. C., Stievenart, M., van de Moortele, G., Roskam, I., Charlier, D., et al. (2009). Variabilité des jugements évaluatifs concernant les troubles du comportement du jeune enfant. *Enfances/Adolescences*, *16*, 33–44.

Kollins, S. H., Epstein, J. N., & Conners, C. K. (2004). *Conners' Rating Scales – Revised. The use of psychological testing for treatment planning and outcomes assessment: Volume 2: Instruments for children and adolescents* (3rd ed, pp. 215–233). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

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. *The American Journal of Psychiatry*, *160*, 1566–1577.

LaFreniere, P. J., & Dumas, J. E. (1996). Social competence and behavior evaluation in children ages 3 to 6 years: The short form (SCBE-30). *Psychological Assessment*, *8*, 369–377.

LaFreniere, P. J., Dumas, J. E., Capuano, F., & Dubeau, D. (1992). Development and validation of the preschool socioaffective profile. *Psychological Assessment*, *4*, 442–450.

Lavigne, J. V., Arend, R., Rosenbaum, D., Binns, H. J., Christoffel, K. K., & Gibbons, R. D. (1998). Psychiatric disorders with onset in the preschool years: I. Stability of diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, *37*, 1246–1254.

Meunier, J. C., Roskam, I., Stievenart, M., van de Moortele, G., Browne, D. T., & Kumar, A. (2011). Externalizing behavior trajectories: The role of parenting, sibling relationships and child personality. *Journal of Applied Developmental Psychology*, *32*, 20–33.

Owens, E. B., & Shaw, D. S. (2003). Predicting growth curves of externalizing behavior across the preschool years. *Journal of Abnormal Child Psychology*, *31*, 575–590.

Perren, S., Von Wyl, A., Stadelmann, S., Bürgin, D., & von Klitzing, K. (2006). Associations between behavioral/emotional difficulties in kindergarten children and the quality of their peer relationships. *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*, 867–876.

Robinson, L. R., Morris, A. S., Heller, S. S., Scheeringa, M. S., Boris, N. W., & Smyke, A. T. (2009). Relations between emotion regulation, parenting, and psychopathology in young maltreated children in out of home care. *Journal of Child and Family Studies*, *18*, 421–434.

Roskam, I., Meunier, J. C., & Stievenart, M. (2013). The comparison and combination of multi-informant and multi-method data on preschoolers' externalizing behaviour. *The International Journal of Educational and Psychological Assessment*, *14*, 79–93.

Roskam, I., Stievenart, M., Meunier, J. C., Van de Moortele, G., Kinoo, P., & Nassogne, M. C. (2010). Comment les parents, les enseignants et les cliniciens évaluent les troubles du comportement externalisé du jeune enfant ? Étude de la variabilité des jugements évaluatifs et de son impact sur le développement de l'enfant. *Pratiques Psychologiques*, *16*, 389–401.

Roskam, I., Stievenart, M., Meunier, J. C., Van de Moortele, G., Kinoo, P., & Nassogne, M. C. (2011). Le diagnostic précoce des troubles du comportement externalisé est-il fiable ? Mise à l'épreuve d'une procédure multi-informateurs et multiméthodes. *Pratiques Psychologiques*, *17*, 189–200.

Roskam, I., Stievenart, M., Tessier, R., Muntean, A., Escobar, M. J., Santelices, M. P., et al. (2014). Another way of thinking about ADHD: The predictive role of early attachment deprivation in adolescents' level of symptoms. *Social Psychiatry and Psychiatric Epidemiology*, *49*, 133–144.

Setterberg, S., Bird, H., Gould, M., Schaffer, D., & Fisher, P. (1992). *Parent and interviewer versions of the children's global assessment scale*. New York: Columbia University.

Super, C. M., & Harkness, S. (1986). The developmental niche: A conceptualization at the interface of child and culture. *International Journal of Behavioral Development*, *9*, 545–569.

Swanson, J. M., Kraemer, H. C., Hinshaw, S. P., Arnold, L. E., Conners, C. K., Abikoff, H. B., et al. (2001). Clinical relevance of the primary findings of the MTA: Success rates based on severity of ADHD and ODD symptoms at the end of treatment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*, 168–179.

Underwood, M. K., Hurley, J. C., Johanson, C. A., & Mosley, J. E. (1999). An experimental, observational investigation of children's responses to peer provocation: Developmental and gender differences in middle childhood. *Child Development*, *70*, 1428–1446.

Wakschlag, L. S., Briggs-Gowan, M. J., Carter, A. S., Hill, C., Danis, B., Keenan, K., et al. (2007). A developmental framework for distinguishing disruptive behavior from normative misbehavior in preschool children. *Journal of Child Psychology and Psychiatry*, *48*, 976–987.

Wakschlag, L. S., Briggs-Gowan, M. J., Hill, C., Danis, B., Leventhal, B. L., Keenan, K., et al. (2008). Observational assessment of preschool disruptive behavior, Part II: Validity of the Disruptive Behavior Diagnostic Observation Schedule (DB-DOS). *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*, 632–641.

Wakschlag, L. S., & Keenan, K. (2001). Clinical significance and correlates of disruptive behavior in environmentally at-risk preschoolers. *Journal of Clinical Child Psychology*, *30*, 262–275.

Wakschlag, L. S., Leventhal, B. L., Briggs-Gowan, M. J., Danis, B., Keenan, K., Hill, C., et al. (2005). Defining the "disruptive" in preschool behavior: What diagnostic observation can teach us. *Clinical Child and Family Psychology Review*, *8*, 183–201.