The Bird's Nest Drawing and Accompanying Stories in the Assessment of Attachment Security

Plain-Language Summary

The relationship people have with others can be greatly affected by the childhood experiences they have with those caring for them and can cause what art therapists call 'attachment difficulties'. The Bird's Nest Drawing, first developed in America, provides a way for therapists (and researchers) to understand people's attachment difficulties. Drawings can express what is difficult to say in words. The person draws a bird's nest and writes a short story about it. Inclusion of the story is intended to provide richer information about the client and his or her drawing.

It is important that the drawing and accompanying story provide an accurate view of people's attachment experiences. This is more likely if they use a tested set of instructions to rate elements in the drawing, such as how the nest is positioned on the tree. Instructions for rating drawings already exist, however no instructions existed for rating the stories people tell. In this study, we used 136 drawings and stories from a previous study with American university students, and also participants' scores (from the same study) on well-established questionnaires about attachment relationships. We devised a rating system for the stories. Then we checked how reliable and accurate the ratings were for both the drawings and the stories. Part of this checking involved comparing the ratings with the questionnaire scores.

We found that neither the existing drawing ratings nor the new story ratings were sufficient when judged by established principles for good design of assessment measures. For example, the ratings did not sufficiently fit the pattern of scores on the established questionnaires. Although drawings can be helpful in art therapy when people find it difficult to use words, we suggest that therapists and researchers exercise caution in how they use the Bird's Nest Drawing (even with stories) until it can be further developed as an assessment measure.

Abstract

Art therapists recognise the importance of clients' early attachment experience. The Bird's Nest Drawing (BND) is an attachment security measure with potential for research and practice, developed by American art therapists. The client draws a bird's nest and writes a corresponding narrative, the latter assisting an integrative evaluation. A systematic rating procedure currently only exists for the drawings. The present study was conducted to determine the psychometric properties of the BND after adding new ratings of the BND story using data from an earlier study with university students. The current investigators established a systematic method for coding BND stories ($N = 136$). Content analysis yielded five ratings, which were examined in conjunction with the corresponding drawings and existing drawing ratings, and two self-report attachment questionnaires. Four of the 11 drawing rating items and the overall attachment rating had acceptable psychometric properties, as did two of the new story ratings. However, the seven items did not inter-correlate consistently, nor yield improved internal reliability. Construct validity using the established attachment scales was not established convincingly. Results suggest a need for further development of the BND taking greater account of psychometric principles of scale development, with implications for art therapy research and practice.

Keywords: Art therapy assessment, Assessment research, Attachment theory, Attachment security, Bird's Nest Drawing, Content analysis, Psychometric properties, Integrative assessment approaches

Introduction

The advantages of process- and performance-based over self-report or questionnaire-based testing are well established in the psychological assessment literature (Betts & Deaver, 2019; Bornstein, 2011; Finn, 2012). Further, 'performance-based personality tests are very useful in part because they tap right-hemisphere and subcortical brain functioning and provide information that clients cannot directly report' (Finn, 2012, p. 440). Presenting problems can appear more clearly in drawings than by verbal description alone and can aid in building rapport between client and therapist (Kaiser, 1996). Although UK art therapists are less inclined to use formal art-based assessments, there is recognition that clients' attachment insecurity may be both the source of many of their difficulties and impact therapeutic processes (Greenwood, 2011). Beginning in 2010 the British Association of Art Therapists run an annual conference on 'attachment and the arts' to explore the relevance to art therapy of new understandings in the neurobiology of attachment (Springham, Thorne & Brooker, 2014). Greenwood (2011) drew on Wilkinson (2010) in explaining the role of implicit memory in early childhood before left-brain verbal processing comes to the fore. Thus the earliest attachment experiences may be held in nonverbal memory.

According to Bowlby's (1969. 1980) seminal attachment theory, young children have an innate need for safety and security, and remaining close to effective care givers. Early experiences with caregivers are internalized as 'internal working models' (IWM) of attachment, an internal mental representation of self and others, which includes beliefs and expectations about behaviours in attachment-relevant contexts (Mayseless, 1996). Attachment theory is empirically well-grounded, with findings that attachment security in infancy correlates with

numerous life variables, including emotional health, academic achievement, suicide risk, and quality of life (Cassidy & Shaver, 2008).

The knowledge of attachment and other affectional bonds across the life cycle is useful in understanding how early attachment styles with caregivers can affect future relationships (Ainsworth, 1989; Hinde, Marris, & Parkers, 1991). The caregiving system in the early stages of life has an impact on bonds with others such as friends, siblings, and significant others. The applications of attachment concepts and subsequent findings have proven clinical significance and are beneficial for understanding parental functioning and other influences on the security of the individual (Goldberg, Muir, & Kerr, 1995).

Studies on children with insecure attachment have demonstrated that the early attachment styles predict certain behaviours such as impaired social, psychological, and neurobiological functioning over time (Schore, 1994, 2001). Insecurely formed early relationships with the caregiver increase the chances for children to develop psychopathology throughout their lifespan (Betts, 2003; Da Costa et al. 2000; Greenberg & Speltz 1988; Sroufe, 1988). Insecure attachment has also been linked to depressive symptoms and less constructive responses in stressful situations (Feeney et al. 2003). Finally, children with insecure attachments are less likely to be cheerful, often finding intimate relationships difficult, and tend to be vulnerable in conditions of adversity (Bowlby 1988).

In order to assess secure and insecure attachment, family drawings have been used to evaluate family dynamics and the individual's perception of the family (Kwiatkowska, 1978) in a number of countries. The American art therapist Donna Kaiser developed the Bird's Nest Drawing (BND) in 1996 (Kaiser & Deaver, 2011), and suggested that a benefit of the BND is that participants usually perceive it as non-threatening and symbolically distant. Kaiser (1996)

asserted that the symbolic imagery of a bird's nest provides information relevant to security. Kaiser's original study collected BND drawings from a sample of 41 mothers and used the established Attachment to Mother (ATM) scale from the Inventory of Parent and Peer Attachment (IPPA; Armsden & Greenberg, 1987). The participants were divided into a more securely attached group and a less securely attached group using the established scale. The results showed no statistically significant association between ATM scores and nine indicators from the BND rating. However, Kaiser (1996) used five new picture elements not previously measured (for example, presence or absence of parent and baby birds), and the effect sizes for the association between two of this second set of five indicators and ATM were between medium and large (Cohen, 1988), with all three others (though not statistically significant) in the expected direction. Kaiser (1996) suggested that her study may have had low statistical power and might have had more significant findings with a larger sample and given these results this may be a fair suggestion. Despite the limitations of Kaiser's preliminary study, it provided the groundwork for further studies on the BND's validity and reliability in measuring attachment security.

The BND assessment has been conducted with different groups in the USA and has shown significant relationships between participants' attachment security and BND drawings. Francis, Kaiser, and Deaver (2003) conducted a study with problematic substance users (PSU) using Kaiser's (1996) original graphic indicators of the BND. The BNDs of 43 participants in the PSU group, and the BNDs of 27 participants in the comparison group with no known history of problematic substance use were compared. The two groups appeared to differ on marital status, although no statistical test was carried out. The participants' attachment security was established through the Relationship Questionnaire (RQ; Bartholomew & Horowitz, 1991), which examines adult attachment styles, and the BND. Kaiser and Deaver (2003) predicted that the BND would

show differences between people who had different adult attachment styles according to the RQ, and that BND scores of the two groups (PSU versus no PSU) would differ. Two drawing features were significantly different between the PSU and comparison group ($p < .01$): The comparison group used green as the dominant colour more often and included birds drawn in the nest. When the group was divided into secure and insecure using the RQ, four drawing indicators showed high statistical difference between these two groups, with large effect sizes (Cohen, 1988): any birds, birds in the nest, four or more colours, and any colour used. Additionally, the study revealed some common themes through content analysis of the BND stories: 'home and family, the wonder of nature and renewal of life, food or hunger, abandonment, and stories based on personal experiences or with personally significant environments' (Francis, Kaiser, & Deaver, 2003, p. 131). This preliminary study added to the previous findings about BND's graphic indicators and extended the research population. Further, this study provided initial indications for the potential use of the BND story to complement the drawings for attachment assessment.

Kaiser and Deaver's (2009) research outlined statistically significant findings from five previous BND studies and provided additional information from four other studies to support the BND's clinical use. Hyler's (2002) study on 49 elementary school children's BNDs indicated that children with secure attachment used green as the dominant colour in their drawings statistically significantly more than the other participants. It was also demonstrated that the insecure attachment group tended to draw nests in vulnerable positions and the dominant colour used in this group was brown. However, these latter results did not reach statistical significance. Overbeck (2002) researched the BNDs of a high-risk sample of 32 pregnant women. Women who displayed an insecure attachment tended to use minimal paper space, did not include an environment for the bird's nest and used little colour. Finally, a study of 14 adolescents in foster

care (Trewartha, 2004) supported previous studies' graphic indicators of insecure attachment. Some of the nests floated on the page, drawings appeared impoverished, were completed with minimal investment, and used very little space. Results from these studies are not conclusive, especially where there was no comparison group. However, they do lend some support for the BND as an instrument with potential for assessing attachment security.

The Harmon-Walker and Kaiser (2015) study with university students examined individual graphic indicators and overall impression ratings of 136 adult undergraduates' BNDs using two rating scales. Participants completed the BND, the Inventory of Parent and Peer Attachment (IPPA; Armsden & Greenberg, 1987), and the Experiences in Close Relationships Questionnaire (ECR; Bartholomew & Shaver, 1998). BND features were blind-rated by two raters and level of agreement computed using the kappa coefficient, which corrects for chance (two raters choosing the same category by chance alone). According to Fleiss (1981), kappa of 0.75 represents good agreement, 0.4-0.75 fair, and 0.4 or lower is poor. Inter-rater reliability on 10 of the 11 graphic indicators was between 0.362 for nest in vulnerable position, and 0.924 for birds included. Reliability for presence of restarts, cross-outs or erasures was poor (-0.011). Two of 11 graphic indicators reached statistical significance for chi squared test of association with self-report attachment groups as defined by Kaiser and Deaver (2009): secure, dismissing-avoidant, preoccupied, dismissing-fearful). The inclusion of a family of birds was significantly associated with high attachment to mother, and bottomless nests were associated with poor experience of close relationships. An association was also found between overall impression ratings of attachment from BND drawings and self-reported attachment to mother.

The inconsistent results of these BND studies suggest the need for global rating approaches for process-based, constructive assessments rather than use of individual indicators.

Several studies that have examined child attachment using drawings have found stronger correlations with aggregate systems over individual indicator scales (Fury, Carlson, & Sroufe, 1997; Goldner, 2014; Goldner & Scharf, 2012; Pianta, Longmaid, & Ferguson, 1999). The movement to adopt global, composite, formal, or structural variables over isolated drawing variables has been clarified in the US art therapy literature (Betts, 2006; Cohen, Hammer, & Singer, 1988, Gantt & Tabone, 1998). It also concurs with broader psychometric principles (DeVellis, 2017).

The Present Study

This study was conducted to determine the construct validity of the BND as an assessment of attachment security based on features of the BND story in addition to the drawing features. The researchers examined the original data set derived from Harmon-Walker and Kaiser (2015) collected from 136 participants that included Kaiser's BND rating scale (Kaiser, 2012a, 2012b), BND stories, and attachment security data from established scales.

There has been no previous attempt to systematically investigate whether themes from BND written descriptions that accompany the drawings might relate to attachment security. While themes of stories were mentioned in Francis, Kaiser, and Deaver (2003), and Kaiser and Deaver (2009) cited BND story themes, formal analyses of BND stories for the purpose of attachment assessment have not been done. Thus, the investigators set out to develop systematic coding, in keeping with current best practices in performance- and process-based psychological assessment (Betts & Groth-Marnat, 2014). Integration of multiple sources, with consideration of global characteristics of art, observed behaviours, and the client's direct input on the artwork, serves clients better (Betts, 2012; Hinz, 2009). It is clear to most clinicians who use drawing as part of assessment that when they ask a client '…to elaborate on the meaning that…aspects of

their drawings have for him or her, this interactive process…' can enhance the communication of meaning between client and therapist (Betts, 2012, p. 207). This approach is especially relevant when considering problems with traditional approaches to projective testing analysis using analysis of single elements (single sign). The authors expected that the content analysis of the BND stories would demonstrate congruent themes with BND graphic indicators and attachment security.

The current study sought to add information about the BND's (Kaiser, 1996) validity in measuring the construct of attachment security. However, in order for any measure to be useful it must have acceptable reliability as well as validity. According to DeVellis (2017), an even earlier step before examining these scale properties is to examine individual scale items for their range, skew (lop-sidedness) and central tendency. Items with limited range, too much skew or insufficiently centred around the mid-point are not good at discriminating between respondents on the variable being measured. The present study set out to examine individual item properties, then inter-rater reliability and internal consistency reliability, and finally construct validity.

Methodology

Design

Data from Harmon-Walker and Kaiser (2015) were used to examine the inter-rater reliability of drawing ratings, and item properties. The new data from the present study, comprising our ratings of the stories gathered by Harmon-Walker and Kaiser (2015), were then examined in the same way. In order to ascertain whether the addition of story items enhanced the existing BND, a correlation matrix was computed to examine the inter-correlations between all items (existing and new), and the internal consistency of the combined items was examined using Cronbach's alpha coefficient. Finally, the plan was to examine the construct validity of the

combined scale by examining its correlation with attachment to mother and attachment to peers from the established IPPA scale, and overall security of attachment using the ECR.

Participants

Participants in Harmon-Walker and Kaiser's (2015) study were recruited via convenience sampling and consisted of 136 students (*M* age = 19.3 years, age range = 18-23, 55 males, 83 females) from a small liberal arts college in southern California, US. Two of the participants did not complete questionnaires, and their drawings were eliminated from analysis. The ethnicity of participants included European American (*n* = 83, 61%), Hispanic/Latino (*n* = 23, 16.9%), Asian American (*n* = 11, 8.1%), African American (*n* = 2, 1.5%), and Other (*n* = 17, 12.5%). Students who participated in the study received research credits for their classes.

Instruments

Instruments used for the study were the Bird's Nest Drawing (BND; Kaiser, 1996), the BND story rating system (developed for the present study by *authors' names withheld for blind review*), the Experiences in Close Relationships Questionnaire (ECR; Bartholomew & Shaver, 1998), and the Inventory of Parent and Peer Attachment (IPPA; Armsden & Greenberg, 1987).

The Experiences in Close Relationships Questionnaire (ECR). The ECR (Bartholomew & Shaver, 1998) assesses self-reported beliefs about current romantic love and relationship patterns. Currently, the ECR is the most commonly used self-report measure of adult attachment style and is psychometrically strong (Graham & Unterschute, 2015; Pilkonis, Kim, Yu & Morse, 2014; Wei, Russell, Mallinckrodt, & Vogel, 2007). Mean Cronbach alpha has been shown to be high (between 0.892 and 0.898) across different studies (Graham & Unterschute, 2015), and self-report scores concur well with clinician-rated attachment (Pilkonis et al., 2014).

The Inventory of Parent and Peer Attachment (IPPA). The IPPA (Armsden & Greenberg, 1987) was developed to assess adolescent attachment security in relation to parents and peers through self-report. Three subscales of the IPPA (Attachment to Mother, Attachment to Father, and Attachment to Peers) have high reliability, with Cronbach's alphas of above .90 (Armsden & Greenberg, 2009, cited in Harmon-Walker & Kaiser, 2015; Pace, San Martini & Zavattini, 2011). Pace et al. (2011) reported that it had overall factorial validity for self-report of overall security of attachment in each of the domains of mother, father and peers. Harmon-Walker and Kaiser's (2015) study excluded the Attachment to Father after pilot testing in order to maintain consistency with previous BND studies.

The Bird's Nest Drawing (BND). The BND (Kaiser, 1996) is a process-based, constructive assessment designed to assess attachment security. Participants are provided with a standardized set of drawing materials and, in a 15-minute period, asked to 'draw a picture of a bird's nest' (Kaiser, 2012b, p. 4). After completion, they are asked to '…write a story about the drawing that includes at least 2-5 sentences.'

BND rating scales. Harmon-Walker and Kaiser (2015) applied the BND Two Category Checklist (BND-TCC; Kaiser, 2012b) and the BND Four Category Overall Impression Form (BND-OIF; Kaiser, 2012b) to rate drawings. The BND-TCC (Kaiser, 2012b) contains 11 graphic indicators rated on a scale of one to five; one indicating 'strongly disagree,' and five, 'strongly agree.' The categories are: one or more birds are depicted; a bird family is included; an environment is included; four or more colours are used; green is the dominant colour; brown is the dominant colour; the nest is tilted 45 degrees or more such that it appears any contents would fall out; the nest lacks a bottom such that it appears the contents would fall out; the nest is

depicted in a vulnerable position; and, there are restarts, erasures, or areas crossed out; there are unusual, bizarre, incoherent, or disorganized elements or approaches to the drawing.

Kaplan and Main's (1986) phrase 'overall impression' was borrowed to describe the approach for ascertaining attachment classifications in children's drawings. The BND-OIF includes operational definitions ascribing the four categories of child attachment to the four groups of drawing characteristics with which they are theoretically associated. Trained raters used the BND-OIF to categorize drawings as one of four overall impression categories (Harmon-Walker & Kaiser, 2015): (1) secure, (2) avoidant, (3) preoccupied, and (4) fearful. An example of 'secure' overall impression incudes 'cheerful, realistic, and/or calm; communicates a generally welcoming quality' whereas 'avoidant' overall impression includes 'isolation or emptiness, absence of connectedness, and/ or absence of events/ movements' (Kaiser, 2012b).

BND story rating scales. The present study developed a rating scale for BND stories. Two raters, the first two authors of the present study, reviewed 136 stories and agreed on the following five scales: coherence and valence using five-point Likert-type scales, and concrete description, humour, and/or drawing ability using nominal scales. The two main categories, coherence and valence, ranged from -2 (incoherent; negative valence) to +2 (coherent; positive valence). A story was rated more coherent if the elements were logically connected or consistent, and more positive if it included themes of nurturing, feeding, birth, protection, safety and love. In addition, relevant BND stories were rated on three subcategories using nominal scales: concrete description, such as if reference was made to colours and shapes used in the drawing, what kind of art materials were used, or a description of the birds and eggs; humour, such as if the participant's story seemed to include comical and/or bizarre content; and/or drawing ability, for example self-deprecating comments about drawing skill. Stories rated as containing concrete

description, humour, and/ or drawing ability were rated 2, and non-concrete description, non-

humour, non-drawing ability were rated 0 (See Box 1 for examples of BND drawings and

stories). The decision was made not to make these last three also Likert-type, since in practice

the examples of the relevant properties were either present or absent.

Box 1. Examples of BND drawings and stories

Positive Valence in drawing and story	Humour in drawing and story
'Once upon a time, there was a sparrow. She was about to be a momma sparrow so she found a safe place in a tall, sturdy oak tree and, with the help of Mr. Poppa sparrow, she constructed a big nest of twigs and hair and other random findings from about the neighborhood. Finally, one day, she laid 2 eggs. Momma sparrow quickly set off finding food for her babies, with Poppa sparrow, as soon as the first chick hatched. Some would consider them careless parents to leave the nest with one baby to oversee the potential hatching of baby #2. But momma knew they were in a safe place and that baby #1, named Henrietta, could hand the responsibility. The End.'	'An eccentric yet fabulous diva-soprano sings Ave Maria with her pet bird, Angelo, who tweets along as she sings. They travel the world together wow-ing the masses with their collaborative gift of song.'

Vulnerable nest position; Concrete Description in story	Lack of environment in drawing; Drawing Ability mentioned in story
'I wanted some color so I added green leaves and blue eggs. I like how the tree branches came out because of the size and multiple colors. I don't like the red in the	'My drawing is simple. I am not a good artist, but I like to think I express myself well in other ways.'

bird's nest.'

According to Gerlsma and Luteijn (2000), secure attachment may be related to both positive valence (for individuals who have experienced secure parenting in early childhood) as well as negative valence (for individuals with adverse parenting experiences which have earned security). Bartholomew (1991, 1997) asserted that high/positive valence may be a result of secure attachment, whereas low/negative valence may be a result of fearful attachment style. In addition, preoccupied and dismissing attachment style may show as mixed valence. A positive valence system is primarily responsible for responses to positive motivational situations or contexts, such as reward seeking, consummatory behaviour, and habit learning. A negative valence system is primarily responsible for responses to aversive situations or context, such as fear, anxiety, and loss. Thus, the present study utilized this concept to develop a valence scale for the BND stories.

Procedures

Original study. The research was conducted in a university on the west coast of the United States. Following approval by a university ethics review board, participants in Harmon-Walker and Kaiser's (2015) study were asked to complete the tasks either in an introductory psychology class or a quiet dormitory room in groups. Each participant was provided with a 9' x 12' sheet of off-white drawing paper, a 24-count pack of crayons, a 24-count pack of coloured pencils, a 10-count pack of thin markers, a drawing pencil, and a sheet of 8.5' x 11' lined paper (Francis, Kaiser, & Deaver, 2003). Participants were first asked to complete a consent form and a demographic questionnaire and then were asked to complete the Attachment to Mother (ATM), Attachment to Peers (ATP) from the Inventory of Parent and Peer Attachment (IPPA; Armsden

& Greenberg, 1987), and the Experiences in Close Relationships Questionnaire (ECR; Bartholomew & Shaver, 1998). After completing the questionnaires, they were asked to 'draw a picture with a bird's nest' (Francis et al., 2003, p. 128). Finally, participants were asked to 'write a 2-5 sentence story' about their drawings. Participants were given 15 minutes to complete the questionnaires and 15 minutes to complete the drawing and the story (Harmon-Walker & Kaiser, 2015). It is not clear from Harmon-Walker and Kaiser (2015) whether any other instructions were given to participants, for example if they asked questions, or what they were told about the aims of the study.

For rating the drawings, two raters were trained by the first investigator, Harmon-Walker, to use the BND Two Category Checklist (BND-TCC; Kaiser, 2012b) and the BND Four Category Overall Impression From (BND-OIF; Kaiser, 2012b); the raters were blind to the purpose of the study. The procedures for training consisted of a collaborative read-through of the rating manual, followed by practice scoring of sample drawings from the manual. When inconsistencies arose among raters during training, they discussed what items may benefit from clarification in the manual. The raters were given the 136 drawings in random order and were asked to rate them independently after the training. Harmon-Walker and Kaiser (2015) reported good inter-rater reliability (Cohen's kappa = 0.7 or higher) on four of their 11 two-category checklist items for story rating (Birds included, family of birds, nest tilted, and nest has no bottom). Based on Fleiss's (1981) assessment of good, fair and poor agreement, there was fair reliability (0.45-0.65) on four (environment included, four or more colours, green dominant, and bizarre elements). The inter-rater reliability for brown dominant and nest in vulnerable position was poor (0.374 and 0.362), and it was very poor for presence of restarts etc (-0.011), the small negative kappa value indicating disagreement or at worst no agreement between raters. The

variable levels of agreement indicate how difficult it is to create clear criteria that can enable independent judges to decide whether a picture displays certain characteristics. For example, any given picture may contain brown and other colours to varying degrees, and the cut-off between brown being dominant or not is unclear. Regarding 'restarts', a smudge in the corner could indicate a restart or artistic expression.

Present study. Stories from the data set (shared with us following permission from the original research ethics review board) were analyzed by two raters blind to the BND drawings (the first two authors of this paper), which resulted in the development of the BND story coding system. To achieve this, stories were identified as containing either mostly positive or negative themes. After discussion, the raters agreed on the establishment of five categories (see Table 1): coherence (positive or negative); valence (positive or negative); concrete description; humour; and drawing ability. The stories were rated using the five categories, with an interval scale and nominal scale. The researchers discussed any inconsistencies in order to achieve consensus.

Table 1. The BND Story Rating System

Scale	Operational Definition
Coherent	Elements of the story are integrated, i.e., the elements are logically connected or consistent.
Incoherent	Elements of the story are lacking integration, i.e., the elements are illogical, disconnected and/or inconsistent.
Positive Valence	Story makes reference(s) to positive motivational situations or contexts, reflecting such themes as (for example), reward seeking, consummatory behaviour, and reward/habit learning.
Negative Valence	Story makes reference(s) to responses to aversive situations or context, such as fear, anxiety, and loss.
Concrete Description	The participant describes the drawing using concrete details, instead of creating a story.

Humour	The participant uses sarcasm, hyperbole, or humour in the narrative.
Drawing Ability	The participant describes their drawing/ artistic abilities.

To further test the BND story coding system (interrater reliability), two additional raters (practising art therapists) blind to the study and the drawings engaged in rating the stories using the coding system. Two raters were trained by the first two authors of this paper to use the BND story coding system. The training procedures included a review of the operational definitions and the relevant scales for each category. For clarification, raters were provided with examples for each category. Then the raters practised rating 15 sample stories. When the practice was complete, researchers and the raters discussed any inconsistencies for clarification. After the training session, the raters were asked to rate 30 random stories independently. In addition, they were asked to read all the samples first without rating in order to form a baseline.

Data analysis

Using the data from Harmon-Walker and Kaiser (2015), the skew and kappa coefficient of agreement of BND scored items were computed for the present paper, using the IBM SPSS software package (version 23). Items with a good range of scores, low skew (not lopsided) and means close to the centre of the scale (central tendency) tend to have greater capacity to distinguish between levels of the attribute one aims to measure (DeVellis, 2017). With binary ratings (scoring 0 or 1) the items have greater utility if there is an even distribution than if most participants score as only one of the categories. For further discussion of skew and central tendency, see Field (2018). The properties of the new BND story ratings were examined for central tendency, range and skew, and for agreement between independent raters.

Next the Spearman correlation matrix (Field, 2018) was computed for items from the original BND and story items that met criteria for acceptability regarding the above properties.

This was to establish whether all the items appeared to tap into the same underlying construct of attachment security. The internal consistency (a form of reliability) of the new scale was then checked according to the widely used Cronbach alpha coefficient (DeVellis, 2017). A more reliable scale tends to have a higher Cronbach alpha, indicating that its various items are measuring something consistently.

To test construct validity of the BND drawing and story measures, we computed the associations between three existing self-report indicators of attachment (attachment to mother and attachment to peers on the IPPA (Inventory of Parent and Peer Attachment) and security as indicated by the Experience in Close Relationships (ECR) questionnaire and the BND drawing scale and story ratings. Given that the existing questionnaires assessing attachment have established reliability and validity, there should be at least a moderate positive correlation between them and the BND. These associations were computed using the nonparametric Spearman correlation because some scores were somewhat skewed, which can make the parametric Pearson correlation less accurate. Marked negative skew is where most participants' scores bunch up at the high end with smaller numbers of participants scoring in all ranges below it, and positive skew is where scores bunch up towards the lower end with a long tail of a few participants scoring at the middle and higher end. A skewness of zero represents no skew, where participants' scores are distributed evenly above and below the mean score. Bulmer's (1979) cut-off values of skewness were used to determine low (-0.5 through 0 to +0.5), moderate (-0.5 to -1.0 and 0.5 to 1.0) and high (less than -1.0 or greater than 1.0).

Results

Properties of the existing BND ratings

Using the data from Harmon-Walker and Kaiser (2015), Table 2 shows the skew
computed for the present paper, alongside the kappa coefficient of agreement as reported in
Harmon-Walker and Kaiser (2015). On the basis of these two quantities, the last column in Table
2 suggests that only the first four items are suitable for use in measurement, because they
combine good or fair range and good or fair inter-rater reliability. The OIF-BND item was also
reported by Harmon-Walker and Kaiser (2015) to have good inter-rater reliability (Pearson's r =
0.736). It was found by the present authors to have low skew (0.179 with confidence 95%
interval -0.179 to 0.555).

Table 2. Skewness and inter-rater reliability (kappa) of original BND ratings

Two Category Checklist Item	% present	Skew	Skew 95% confidence intervals[1]	Skew category	kappa	Item quality
Birds included	41.9	-.318	-.680 to .015	Low	0.924	Good
Family of birds included	58.8	.381	.045 to .788	Low	0.750	Good
Environment included	39.0	-.445	-.824 to -.105	Low	0.536	Fair
Four or more colours	31.6	-.788	-1.203 to -.413	Moderate	0.624	Fair
Green dominant	89.0	2.629	1.846 to 4.088	Very high	0.511	Poor
Brown dominant	72.1	.983	.576 to 1.404	Moderate	0.374	Poor
Nest tilted	75.7	1.203	.788 to 1.773	High	0.765	Poor
Nest has no bottom	87.5	2.280	1.639 to 3.289	Very high	0.775	Poor
Nest in vulnerable position	49.3	-.045	-.413 to .288	Low	0.362	Poor
Restarts, crossouts, erasures	97.8	6.556	4.088 to 11.619	Very high	-0.001	Poor
Bizarre, incoherent or disorganized elements	69.6	.863	.477 to 1.300	Moderate	0.487	Poor

Notes: 1 Confidence intervals obtained via bootstrapping (1000 samples)

Properties of the new story ratings

The properties of the new BND story ratings are shown in Table 3. All but the valence
scale had a high level of skew, with less than 10% of the stories being rated as present for
humour and drawing ability.

Table 3. Properties of story attributes *(N = 136)[1]*

Story attribute (Likert)	Mean	Range	Skew	Skew 95% confidence intervals	Skew category
Coherence	1.3088	-2 to +2	-1.110	-1.569 to -.659	High
Valence	0.6691	-2 to +2	-0.484	-.733 to -.161	Low
Story attribute (binary)	**% present**	**NA**	**Skew**	**Skew 95% confidence intervals**	**Skew category**
Concrete story elements	19.8		1.528	1.079 to 2.194	Very high
Presence of humour	7.4		3.304	2.400 to 4.978	Very high
Reference to drawing ability	2.2		6.581	4.105 to 11.662	Very high

[1] Ratings of first author

The agreement between independent raters for the story attributes were as shown in Table 4. Levels of agreement were good for both coherence and valence. These values were not computed for the binary ratings because their skewness meant they were unlikely to be useful items. On the basis of the coherence scale having high skew but good central tendency and covering the full range of scores, as well as good inter-rater reliability, it was retained. The valence scale was retained on the basis of low skew and fair inter-rater reliability.

Table 4. Levels of agreement for ratings of story attributes (with bootstrap)

Subscale	Spearman's rho (N=3 0)	95% confidence interval
Coherence	0.731	0.501-0.903
Valence	0.687	0.446-0.835

Inspecting all accepted items

Table 5 shows the Spearman correlation matrix for all accepted items from the original BND and the two story items accepted. The matrix shows that story coherence and valence are positively correlated, which would be expected if they tap into the same underlying construct

(attachment security). Regarding the four items from the original BND test, OIF correlated well with all four of the retained drawing descriptors. The four drawing descriptors did not all correlate appreciably with each other.

Table 5. Spearman correlation matrix for all initially accepted BND and BND-story items

	Story valence	Birds present	Family of birds present	Environ-ment present	Four or more colours	OIF secure vs insecure
Story coherence	0.394**	0.196*	0.203*	-0.044	0.057	0.095
Story valence		0.104	0.245**	-0.069	0.070	0.215*
Birds present			0.711**	0.146	0.160	0.239**
Family of birds				0.240**	0.151	0.434**
Environment					0.429**	0.308**
Four or more colours						0.273**

Notes: ** Correlation significant at $p = .01$; * Correlation significant at $p = .05$; BND OIF Overall Impression Form

The correlations between *Family of birds present* and *Birds present* is confounded because any drawing with a family of birds would also be rated as *Birds present*. In terms of the correlations between the original BND test and the new story elements, each of the two new story item elements only correlated with two out of five of the original items (Table 5). If the story items were to enhance the overall scale, the Cronbach alpha value would increase with them added. The Cronbach alpha for the five items retained from the original scale was 0.692, which appears to indicate a reasonable level of internal consistency. However, this index might be inflated by the two 'birds' items not being independent of one another. With either of these items deleted, Cronbach alpha was lower, as expected (0.639 with *Birds present* deleted, and 0.582 with *Family of birds* deleted. Therefore, a four-item scale was retained for drawing rating (henceforth termed 'drawing scale'), comprising the OIF value, *Family of birds*, *Environment*, and *Four or more colours*. The scores of all four items were combined after standardising them

to the same scale (1 = absent/insecure, 2 = present/secure) so as not to inflate the influence of any item. With the two new story items added, the alpha value was also less favourable (0.583). Therefore, rather than attempting an improved scale by combining items, the two new items are treated separately in the next part, which concerns attempting to establish construct validity.

To test construct validity of the BND drawing and story measures, we report the associations between three existing self-report questionnaires on attachment (attachment to mother and attachment to peers on the IPPA and security as indicated by Experience in Close Relationships questionnaire - ECR) and the BND drawing scale and story ratings (coherence, valence). These associations were computed using Spearman correlation analysis (Table 6). None of the correlations was statistically significant, and all effect sizes were small. Positive story valence showed a non-significant trend towards association with attachment to mother.

Table 6. Spearman's correlations between BND measures and self-reported attachment to mother and to peers

Story item	Attachment to mother (*p*)	Attachment to peers (*p*)	Experience of close relationships (*p*)
Drawing scale	0.155 (.072)	0.089 (.304)	-0.003 (.974)
Story coherence	0.001 (.994)	0.084 (.328)	-0.016 (.852)
Story valence	0.158 (.066)[1]	0.144 (.094)	0.149 (.082)

Note: 1 = Non-significant trend

Discussion

Art therapists recognise the importance of clients' attachment history and attachment styles in relation to understanding their difficulties and their potential impact on the therapeutic relationship (Greenwood, 2011; Springham et al., 2014; Kaiser & Deaver, 2009). Formal assessment of attachment difficulties has both research utility and potential benefits for therapeutic practice. However, some attachment difficulties may date back to the earliest

experiences before the full development of left-hemisphere functioning (Wilkinson, 2010). For this reason, an assessment comprising both the visual medium and words, such as the BND, might have the potential to capture what a purely verbal approach cannot.

Prior to the present study, there was scarce research on the BND stories in relation to the BND graphic indicators and attachment security. Kaiser and Deaver (2009) briefly mentioned BND story themes, and Sheller (2007) reported that an insecure attachment group of school-aged children wrote stories that included themes such as; 'a lack of feeling safe within a home base, with or without attachment figures, negative and/or abandonment experiences with attachment figures, seeing the world as a dangerous place, strategies to make up for lack of security, protection and care, and experiencing feeling safer at night' (pp. 122-124).

Despite previous attempts to analyse the BND story themes, a systematic approach in rating the BND stories as indicators of attachment security was lacking. The findings of the current study suggest a need for the development of a comprehensive and global approach to assessing clients' artwork. Whilst utilizing both visual and verbal indicators would be expected to enhance assessment by increasing both the type and number of measurement items (DeVellis, 2017), the present study did not achieve this. Neither the original BND nor the addition of the new items yielded high internal consistency reliability or construct validity. This suggests a need to take a step backwards and to make greater use of basic psychometric principles of scale development in further development of this type of measure.

Findings from Harmon-Walker and Kaiser (2015) suggest the need for global rating approaches for formal art therapy assessments. Furthermore, several studies that examined child attachment through drawing tasks have found stronger correlations with aggregate systems over individual indicator scales (Fury, Carlson, & Shaver, 1997; Goldner, 2014; Goldner & Scharf,

2012; Pianta, Longmaid, & Ferguson, 1999). The movement in adopting global, composite, formal, or structural variables over isolated drawing variables has been emerging in the field of art therapy assessment research in the USA (Betts, 2006; Bucciarelli, 2011; Cohen, Hammer, & Singer, 1988) and this seems a step in the right direction.

Study limitations

Limitations of this research include the sampling method and procedures employed in the Harmon-Walker and Kaiser (2015) study. First, the use of a student sample from one American liberal arts college may have limited the potential for variation in any or all scale items, both those related to the drawings and the new story ratings. This in turn may have led to the elimination of some items that might have been retained had the sample contained people with a broader range of attachment experiences. The convenience sampling method also limits the generalizability of the study results. Thus, further research within the art therapy context and in a therapeutic setting, such as participants from art therapy groups, is recommended.

A further limitation is the order in which the instruments were administered: participants completed the self-report measures first and then the BND (Harmon-Walker & Kaiser, 2015). This order may have undermined the projective nature of the BND. Through the process of completing the questionnaire, participants had the opportunity to consider their attachment dynamics prior to completing the drawing. The instructions for the BND story also may have limited the quality of the stories. Many stories comprised only one or two sentences which made rating difficult. In future studies, we recommend that participants be asked to write more than five sentences.

However, these limitations apply mainly to the use of the BND and BND stories for assessment of attachment in the absence of a therapeutic relationship. Within a therapeutic

relationship, there are further sources of information about clients' attachment strategies. The BND and BND drawings might aid art therapists in understanding a client's attachment security styles, which in turn affect the therapeutic relationship and the course of the treatment. However, they should not be used in isolation to assess an individual's attachment style.

Conclusions and Implications for Art Therapy Practice

The BND is at an earlier stage than needed for context-independent use. Further scale item development is needed, and perhaps the use of additional drawing and story-writing tasks in order to produce a more reliable and valid instrument. Test-retest reliability also needs to be established. Our study findings are perhaps a wake-up call to the field of art-based measures, particularly with consideration to holistic, integrative evaluation approaches. If we want the BND and other similar instruments to be helpful in research and assessment without a great deal of caveats around their use, there needs to be a substantial amount of further development and refinement, with careful consideration of what kind of indicators in drawings and stories might, in combination, capture attachment, both as overall level of security and in terms of the main four strategies commonly discussed. Although our findings suggest that further research is required to validate the BND as an assessment of attachment patterns to inform treatment planning and clinical practice, art therapists are encouraged to pursue use of the BND to further the research. Until the evidence is further developed, the BND should be interpreted with caution, and the full range of other available sources of information should be taken into account.

References

Ainsworth, M. D. S. (1989). Attachments beyond infancy. *American Psychologist, 44,* 709-716.

Armsden, G., & Greenberg, M. (1987). The inventory of parent and peer attachment: Individual differences and their relationship to psychological well-being in adolescence. *Journal of Youth and Adolescence, 16*(5), 427-453.

Bartholomew, K., & Horowitz, L. (1991). Attachment styles among young adults: A test of a four-category model. *Journal of Personality and Social Psychology, 61,* 226–244.

Bartholomew, K., & Shaver, P. R. (1998). Methods of assessing adult attachment: Do they converge? In J. A. Simpson, & W. S. Rholes (Eds.), *Attachment theory and close relationships* (pp. 25–45). New York, NY: Guilford Press.

Betts, D. J. (Ed.). (2003). *Creative arts therapies approaches in adoption and foster care: Contemporary strategies for working with individuals and families.* Springfield, IL: Charles C Thomas.

Betts, D. J. (2006). Art therapy assessments and rating instruments: Do they measure up? *The Arts in Psychotherapy, 33*(5), 422–434.

Betts, D. J. (2012). Positive art therapy assessment: Looking towards positive psychology for new directions in the art therapy evaluation process. In A. Gilroy, R. Tipple & C. Brown (Eds.), *Assessment in art therapy* (pp. 203-218). New York, NY: Routledge.

Betts, D. J., & Deaver, S. P. (2019). *Art therapy research: A practical guide.* New York, NY: Routledge.

Betts, D. J. & Groth-Marnat, G. (2014). The intersection of art therapy and psychological assessment: Unified approaches to the use of drawings and artistic processes. In L.

Handler & A. Thomas (Eds.), *Figure drawings in assessment and psychotherapy: Research and application* (pp. 268-285). New York, NY: Routledge.

Bornstein, R. F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment, 23*(2), 532–544.

Bowlby, J. (1969/1980). *Attachment and loss: Attachment* (Vol. 1). New York, NY: Basic Books.

Bowlby, J. (1988). *A secure base: Parent-child attachment and healthy human development* London, UK: Routledge.

Bucciarelli, A. (2011). A normative study of the Person Picking an Apple From a Tree (PPAT) assessment. *Art Therapy: Journal of the American Art Therapy Association, 28*(1), 31–36.

Cohen, B., Hammer, J., & Singer, S. (1988). The Diagnostic Drawing Series: A systematic approach to art therapy evaluation and research. *The Arts in Psychotherapy, 15*, 11–21.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd Ed.).* Mahwah, NJ: Lawrence Erlbaum

Da Costa, D., Larouche, J., Dritsa, M., & Brender, W. (2000). Psychosocial correlates of prepartum and postpartum depressed mood. *Journal of Affective Disorders, 59,* 31–40.

DeVellis, R.F. (2017). *Scale development: theory and applications (4th Ed.).* Thousand Oaks: Sage.

Feeney, J., Alexander, R., Noller, P., & Hohaus, L. (2003). Attachment insecurity, depression, and the transition to parenthood. *Personal Relationships, 10,* 475–493.

Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th Ed.). London: Sage

Finn, S. E. (2012). Implications of recent research in neurobiology for psychological assessment.

Journal of Personality Assessment, 94(5), 440-449. doi: 10.1080/00223891.2012.700665

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: John Wiley.

Francis, D., Kaiser, D., & Deaver, S. (2003). Representations of attachment security in the Bird's Next Drawings of clients with substance abuse disorders. *Art Therapy: Journal of the American Art Therapy Association, 20*(3), 124–137.

Fury, G., Carlson, E. A., & Sroufe, L. A. (1997). Children's representations of attachment relationships in family drawings. *Child Development, 68*(6), 1154–1164.

Gantt, L., & Tabone, C. (1998). *The Formal Elements Art Therapy Scale: The Rating Manual.* Morgantown, WV: Gargoyle Press.

Gerlsma, C., & Luteijn, F. (2000). Attachment style in the context of clinical and health psychology: A proposal for the assessment of valence, incongruence, and accessibility of attachment representations in various working models. *British Journal of Medical Psychology, 73*, 15-34.

Goldberg, S., Muir, R., & Kerr, J. (Eds.). (1995). *Attachment theory: Social, developmental, and clinical perspectives.* Hillsdale, NJ: Analytic Press.

Goldner, L. (2014). Revisiting the Bird's Nest Drawing: Towards a global approach. *The Arts in Psychotherapy, 41*, 391–399.

Goldner, L., & Scharf, M. (2012). Children's family drawings and internalizing problems. *The Arts in Psychotherapy, 39*, 262–271.

Graham, J. M., & Unterschute, M. S. (2015). A reliability generalization meta-analysis of self-report measures of adult attachment. *Journal of Personality Assessment, 97*(1), 31-41. doi: 10.1080/00223891.2014.927768

Greenberg, M. T., & Speltz, M. L. (1988). Attachment and the ontogeny of conduct problems. In

    J. Belsky & T. Nezworski (Eds.), *Clinical implications of attachment* (pp. 177–218).

    Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Greenwood, H. (2011). Long term individual art psychotherapy. Art for art's sake: The effect of

    early relational trauma. *International Journal of Art Therapy, 16,* 41-51.

Harmon-Walker, G., & Kaiser, D. H. (2015). The Bird's Nest Drawing: A study of construct

    validity and interrater reliability. *The Art in Psychotherapy, 42,* 1-9.

Hinde, J. S., Marris, P., & Parkers, C. M. (Eds.). (1991). *Attachment across the life cycle.*

    London, UK: Routledge.

Hinz, L. D. (2009). *The Expressive Therapies Continuum: A framework for using art in therapy.*

    New York, NY: Routledge.

Hyler, C. (2002). *Children's drawings as representations of attachment* (Unpublished master's

    thesis). Eastern Virginia Medical School, Norfolk, VA.

Kaiser, D. (1996). Indications of attachment security in a drawing task. *The Arts in*

    Psychotherapy, 23(4), 333-340.

Kaiser, D. H. (2012a). *Kaiser's Bird's Nest Drawing and Story Rating Scale*. Unpublished

    manuscript. Creative Arts Therapies Department, Drexel University, Philadelphia, PA.

Kaiser, D. H. (2012b). *Kaiser's Bird's Nest Drawing Two Category Checklist and Four*

    Category Overall Impression Manual. Unpublished manuscript. Creative Arts Therapies

    Department, Drexel University, Philadelphia, PA.

Kaiser, D. H. & Deaver, S. (2009). Assessing attachment with the Bird's Nest Drawing: A

    review of the research. *Art Therapy: Journal of the American Art Therapy Association,*

    26(1), 26-33.

Kaplan, N., & Main, M. (1986). Instructions for the classification of children's family drawings in terms of representation of attachment. Berkeley, CA: University of California at Berkeley. Unpublished manuscript.

Kwiatkowska, H. Y. (1978). *Family therapy and evaluation through art.* Springfield, IL: Charles C Thomas.

Mayseless, O. (1996). Attachment patterns and their outcomes. *Human Development, 39,* 209–223.

Overbeck, L. (2002). *A pilot study of pregnant women's drawings* (Unpublished master's thesis). Eastern Virginia Medical School, Norfolk, VA.

Pace, C.S., San Martini, P., & Zavattini, G.C. (2011). The factor structure of the Inventory of Parent and Peer Attachment (IPPA): A survey of Italian adolescents. *Personality and Individual Differences,* 51, 83-88. doi: 10.1016/j.paid.2011.03.006

Pianta, R., Longmaid, K., & Ferguson, K. (1999). Attachment-based classifications of children's family drawings: Psychometric properties and relations with children's adjustment in kindergarten. *Journal of Clinical Psychology, 28*, 244–255.

Pilkonis, P., Kim, Y., Yu, L., & Morse, J. Q. (2014). Adult Attachment Ratings (AAR): An item response theory analysis. *Journal of Personality Assessment, 96*(4), 417-425. doi: 10.1080/00223891.2013.832261

Schore, A. N. (1994). *Affect regulation and the origin of the self: The neurobiology of emotional development.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Schore, A. N. (2001). Effects of a secure attachment relationship on right brain development, affect regulation, and infant mental health. *Infant Mental Health Journal, 22*(1–2), 7–66.

Sheller, S. (2007). Understanding insecure attachment: A study using children's bird nest
    imagery. *Art Therapy: Journal of the American Art Therapy Association, 24*(3), 119–127.

Springham, N., Thorne, D., & Brooker, J. (2014). Softer: Looking for oxytocin in art therapy.
    International Journal of Art Therapy, 19, 31-42

Sroufe, A. (1988). The role of infant-caregiver attachment in development. In J. Belsky & T.
    Nezworski (Eds.), *Clinical implications of attachment* (pp. 18–38). Hillsdale, NJ:
    Lawrence Erlbaum Associates, Publishers.

Trewartha, S. (2004). *Attachment strategies of adolescents in foster care: Indicators and
    implications* (Unpublished master's thesis). Eastern Virginia Medical School, Norfolk,
    VA.

Wei, M., Russell, D. W., Mallinckrodt, B. & Vogel, D. L. (2007). The Experiences in Close
    Relationship Scale (ECR)-Short Form: Reliability, validity, and factor structure. *Journal
    of Personality Assessment, 88*(2), 187-204. doi: 10.1080/00223890701268041