

Draw-A-Person: Screening procedure for emotional disturbance

An investigation of the sensitivity of this method to internalising and externalising behavioural problems identified by the Rutter Parent Questionnaire at age 7 in the 1958 National Child Development Study

Miranda Crusco

CLS Cohort Studies

Working paper 2013/7

August 2013

Draw-A-Person: Screening Procedure for Emotional Disturbance

An investigation of the sensitivity of this method to internalising and externalising behavioural problems identified by the Rutter Parent Questionnaire at age 7 in the 1958 National Child Development Study

Miranda Crusco

August 2013

First published in 2013 by the
Centre for Longitudinal Studies
Institute of Education, University of London
20 Bedford Way
London WC1H 0AL
www.cls.ioe.ac.uk

© Centre for Longitudinal Studies

ISBN 978-1-906929-67-1

The Centre for Longitudinal Studies (CLS) is an ESRC Resource Centre based at the Institution of Education. It provides support and facilities for those using the three internationally-renowned birth cohort studies: the National Child Development Study (1958), the 1970 British Cohort Study and the Millennium Cohort Study (2000). CLS conducts research using the birth cohort study data, with a special interest in family life and parenting, family economics, youth life course transitions and basic skills.

The views expressed in this work are those of the author and do not necessarily reflect the views of the Economic and Social Research Council. All errors and omissions remain those of the author.

This document is available in alternative formats.
Please contact the Centre for Longitudinal Studies.
tel: +44 (0)20 7612 6875
email: info@cls.ioe.ac.uk

Contents

Abstract.....	2
Acknowledgements	3
1. Introduction	4
1.1. Background.....	4
1.2. The National Child Development Study.....	4
1.3. Draw-A-Person: Screening Procedure for Emotional Disturbance (DAP:SPED)	5
1.3.1. Reliability and validity of the DAP:SPED	5
1.4. Rationale.....	6
2. Method.....	8
2.1. Design and participants.....	8
2.2. Measures	8
2.2.1. DAP:SPED.....	8
2.2.2. Rutter Parent Questionnaire (RPQ).....	9
2.3. Sampling.....	10
2.4. Procedure	12
3. Results.....	13
3.1. Distribution analysis	13
3.2. Mean comparison of DAP:SPED scores between the samples using pairwise comparisons	15
4. Discussion and conclusion	17
4.1. Limitations.....	18
4.2. Conclusion	19
References.....	20
Appendix A. Original behavioural domain items in the Rutter Parent Questionnaire and modified NCDS (1965) items	22

Abstract

The National Child Development Study is a longitudinal study that recruited 17,500 babies born across England, Scotland and Wales in one particular week during 1958 and collects educational, physical and social data at regular intervals. The current study is concerned with the sweep conducted when the cohort was aged 7 (1965). As part of this sweep they completed a test booklet in school which included the Draw-A-Man (Goodenough, 1926) test. Over the last century, tests have been devised to assess children's drawings of a man for cognitive ability and emotional functioning, one of which was the Draw-A-Person: Screening Procedure for Emotional Disturbance (DAP:SPED) (Naglieri, McNeish and Bardos, 1991). The main aim of this study was to investigate whether children identified as having internalising (emotional) or externalising (behavioural) problems as indicated by the Rutter Parent Questionnaire included indications of such problems in their drawings of a man as identified by the DAP:SPED. A mean DAP:SPED score comparison between the three groups found no significant differences between the normative and both behavioural problem groups, however a borderline significant mean difference of 1.09, $p = .073$ was found between the behavioural problem groups. These findings suggest that the DAP:SPED has limited sensitivity to the behavioural problems as identified by the Rutter Parent Questionnaire in this sample.

Non-technical summary

This study investigates how well the already established Draw-A-Person: Screening Procedure for Emotional Disturbance (DAP:SPED) can identify children with emotional and behavioural problems. Children's drawings have formed part of diagnostic tools and children's tests since the mid-20th century, including in the National Child Development Study (NCDS), a longitudinal birth cohort study. In 1965, the cohort members (aged 7) completed a test booklet which included the Goodenough (1926) Draw-A-Man test to assess the children's creativity, mental age and visual-motor intellectual maturity. This was achieved by coding a drawing of a man and assigning points to the presence of certain attributes such as ears; the quality of the drawing (for example how the lines meet and whether they are rigid); and the proportionality of the head, feet, hands, etc.

These drawings were manually accessed at the Centre for Longitudinal Studies by the current author and recoded using a more recent coding system, the DAP:SPED. This coding system uses an objective approach where the features of the child's drawing of a man are physically measured and the inclusions/omissions of particular features are assigned points. If the drawing scored above the benchmark, this would indicate emotional disturbance. This method contrasts the subjective approach commonly used by clinicians where the drawing would be interpreted and given contextual meaning.

To assess whether the DAP:SPED could identify children with emotional or behavioural problems, groups were formed based on another test within the same sweep of the NCDS: the Rutter Parent Questionnaire (RPQ). This is typically completed by the child's mother and identifies children with emotional disturbance,

behavioural disturbance or neither. By grouping the sample in this way, it was possible to compare the children's DAP:SPED scores between the three groups. It was found that the only significant difference in drawing scores was between the two disturbance groups. The children who were identified by the RPQ as not having such problems did not score much lower on the DAP:SPED than the two disturbance groups.

The findings of this study suggest that the DAP:SPED was not particularly successful at telling the difference between a child with emotional disturbance and a child without. This paper concludes by discussing the potential limitations of the current study, particularly the use of the RPQ to form the three groups for DAP:SPED score comparisons. For a more reliable test of the DAP:SPED's ability to identify children with emotional disturbance, a sample of children who have been clinically diagnosed should be used, avoiding the issues around using potentially biased parental judgements.

Acknowledgements

I would like to thank Dr John Done and Jörg Schulz for their supervision and constant support on this project. I would also like to thank the Centre for Longitudinal Studies for allowing me access to their ESRC-funded scanned document archive, without which I would not have been able to recode the original children's drawings.

1. Introduction

1.1. Background

The significance of children's drawings has been explored extensively since the late 19th century, and they are thought to provide indications of visual-motor development, levels of cognitive functioning and intellectual maturity, projections of personality and self-concept, and assessments of emotional state and disturbances. Drawing is an activity that children tend to enjoy and they willingly produce spontaneous scribbles and drawings from a young age (Koppitz, 1968), making it an attractive method to use in clinical settings. The method can be employed informally, where the child is asked to draw whatever they like in order to make them feel at ease, or to provide grounding for a clinical interview. Alternatively, drawing can be used as part of a formal diagnostic tool and a range of drawing tests are used for this purpose, focusing on two core elements of the drawing process: cognitive development, and a symbolic projection of the unconscious (Bekhit, Thomas and Jolley, 2005), which allows for an assessment of emotional and behavioural problems.

One of the earliest drawing tests was the Draw-A-Man test devised by Goodenough (1926) to assess children's creativity, mental age and visual-motor intellectual maturity by coding features of their drawing of a man (Goodenough, 1926; Knoff, 1990). Points were assigned according to the presence of particular attributes such as ears; the quality of the drawing, for example how the lines meet and whether they are rigid; and the proportionality of the head, feet, hands, etc. (Goodenough, 1926).

1.2. The National Child Development Study

The NCDS is a longitudinal study which was initially set up to study perinatal mortality, identifying social and obstetric factors that may contribute to neonatal death and stillbirth. The birth survey was conducted on approximately 17,500 babies who were born in one particular week during 1958 in England, Scotland and Wales. Although this initial survey was not intended to be a longitudinal study, it was decided by the National Children's Bureau to conduct a follow-up sweep when the cohort had reached age 7, to collect data concerning educational, physical and social development (Davie, Butler and Goldstein, 1972). The information collected about the cohort members is from several sources such as parents (particularly the mother), school teachers, medical professionals and the cohort members themselves (Davie et al., 1972). Part of the 1965 sweep included a test booklet completed in school by the cohort members which included tests to assess the children's cognitive ability. One of these tests was the Goodenough (1926) Draw-A-Man.

Following an inspection of the scoring sheet for these drawings, it seems that the scheme used was Harris' (1963) updated version of Goodenough's original test. Recent research using the NCDS (1965) Draw-A-Man test (see Schoon et al., 2002; Lynn and Kanazawa, 2011; Butler, Peckham and Sheridan, 1973; Kanazawa, 2011) used the original scores available in the dataset and it would seem that the drawings (which are accessible electronically at the Centre of Longitudinal Studies) have not been recoded using a more recent and up-to-date drawing scheme.

1.3. Draw-A-Person: Screening Procedure for Emotional Disturbance (DAP:SPED)

The literature suggests that using children's drawings to assess cognitive ability is an out-dated method, due to high levels of misclassification, i.e. false positives and false negatives and the fact that there are more comprehensive, standardised IQ tests available (Aikman, Belter and Finch, 1992; Plbrukarn and Theeramanoparp, 2003; Willcock, Imuta and Hayne, 2011). Towards the end of the 20th century, drawing tests were focused on the assessment of emotional and behavioural functioning (Koppitz, 1968; McNeish, 1989; Naglieri, McNeish and Bardos, 1991). The most recent and psychometrically advanced human figure drawing test for assessing emotional problems is the Draw-A-Person: Screening Procedure for Emotional Disturbance (DAP:SPED), developed by Naglieri et al. (1991). The test was developed as a modification of existing objective approaches to scoring human figure drawings (HFD), based on physically measured structural items and inclusions or omissions of particular content items. Items that indicated disturbance occurred in less than 16 per cent of the normative sample's drawings, and items with above 80 per cent inter- and intra-rater consistency were included in the scale. When matched for age, race and intelligence measure score, McNeish (1989) found that the set of items was able to differentiate significantly between the normative group and a group of children who had been formally diagnosed with emotional and behavioural disorders.

1.3.1. Reliability and validity of the DAP:SPED

Trevisan (1996) and Flanagan and Motta (2007) conducted reviews of studies investigating the psychometric properties of the DAP:SPED, and both reviews found evidence that with proper use, accurate and reliable information about children's emotional and behavioural functioning can be obtained. The reviewed studies demonstrated the DAP:SPED's reliability using various measures of consistency. Computations of Cronbach's alpha had a range of .67 to .78, indicating a good level of internal consistency, particularly for a projective measure that is based on a rater's judgement of items present in a drawing.

To assess whether the DAP:SPED has an adequate level of discriminant validity, validation studies have been carried out by Naglieri and Pfeiffer (1992), McNeish and Naglieri (1993), Wrightson and Saklofske (2000), and Matto, Naglieri and Clausen (2005). The clinical samples consisted of special education students with serious emotional and behavioural disturbance, as identified by state and federal guidelines, matched with control students on age, sex, race and geographic location, to eliminate potential confounding effects. These studies found that the DAP:SPED was able to significantly discriminate between regular students and special education students.

In an examination of whether the DAP:SPED could predict emotional and behavioural functioning, Matto (2002) used the Shortform Assessment for Children (SAC), which is a shortened version of Achenbach's Child Behaviour Checklist (1991 as cited in Matto, 2002), and the Child and Adolescent Adjustment Profile (CAAP; Ellsworth and Ellsworth, 1998 also cited in Matto, 2002) as concurrent measures. The SAC is completed by the child's parent or guardian, and measures internalising and externalising behaviours. The CAAP is a screening tool for psychosocial

adjustment problems which is also completed by the child's parent or guardian, and has two subscales (Withdrawal and Hostility) that provide information about internalising and externalising behaviour (respectively). The sample used in the study were children referred to outpatient clinical services or residential treatment facilities for concerns such as behavioural and learning problems or disorders such as attention deficit disorder/attention deficit hyperactivity disorder, oppositional defiant disorder and conduct disorder. The results indicated that the DAP:SPED was not a significant predictor of externalising behaviour disturbance when controlling for the Hostility subscale, however the DAP:SPED significantly accounted for 10.8 per cent variance in internalising behaviour disturbance, when controlling for CAAP Withdrawal subscale. The findings provide support for the instrument's validity as a screening tool for child behavioural functioning, particularly internalising behaviour. Matto (2002) stresses the importance of this finding, as internalising behaviour is more difficult to detect than externalising behaviour which is more overt in nature. As this study did not find the DAP:SPED to significantly predict externalising behavioural disturbance, it is suggested that future studies should investigate this using alternative behavioural measures.

1.4. Rationale

The traditional method of assessing whether a child has an emotional or behavioural problem is by administering a behavioural checklist for the child's parent or teacher (or both) to complete, with items pertaining to the child's behaviour over a period of time. A widely used checklist in the behavioural domain is Achenbach's (1991 as cited in McConaughy, 1992) Child Behaviour Checklist, which is based on the internalising and externalising dimensions of behaviour. Similarly, Rutter, Tizard and Whitmore (1970) devised parent and teacher behavioural questionnaires to identify children with behavioural problems within the same dimensions. As both of these measures (and others similar within the domain of behavioural problems) are completed by individuals other than the child themselves, any conclusions that are made based on such measures cannot be conclusive, as some facets of behaviour may not be identifiable by the child's parent or teacher.

To obtain diagnostic information directly from a child, several issues may be encountered such as language boundaries, limited understanding of emotion and inability to verbalise feelings, or even responding in an acquiescent way. To counteract such issues, drawing tests which involve the objective interpretation of features were developed to allow for an unbiased representation of the child's emotional profile, including the most recent DAP test, the DAP:SPED. The DAP:SPED's psychometric properties have been shown to be more advanced than previous HFD tests, as it provides a reliable and consistent method to scoring drawings. Nonetheless, there still appears to be gaps in identifying the types of disturbance the DAP:SPED is able to detect.

The nature of the NCDS is such that it allows for the investigation of a nationally representative dataset both concurrently and longitudinally. The 1965 sweep contains both children's drawings and a measure of social adjustment, therefore it is possible to assess whether drawings can be used to identify children with current emotional or behavioural problems, using the most recent scoring system, the DAP:SPED. As the drawings were originally coded using the Goodenough-Harris

(1963) scheme, a subsample were manually accessed at the Centre of Longitudinal Studies and recoded using the DAP:SPED procedure. This made it possible to test whether children identified as having internalising (emotional) or externalising (behavioural) problems by the Rutter Parent Questionnaire include indications of emotional or behavioural problems in their drawings of a man as assessed by the DAP:SPED.

2. Method

2.1. Design and participants

A secondary data analysis was conducted on a sample of the data collected for the National Child Development Study (NCDS). The current study was concerned with the 1965 sweep of the NCDS, when the cohort members were 7 years old; the appropriate dataset containing this information was obtained from the UK Data Service (available to researchers at www.ukdataservice.ac.uk). Three samples were drawn from this cohort, to form a normative group ($n = 80$) which was stratified on gender, and two extreme disturbance groups which were stratified on gender and the Rutter Parent Questionnaire's (RPQ) internalising and externalising behaviour subscales to form the behavioural disturbance group ($n = 45$) and the emotional disturbance group ($n = 46$).

2.2. Measures

2.2.1. Draw A Person: Screening Procedure for Emotional Disturbance (DAP:SPED)

The DAP:SPED was developed by Naglieri et al. (1991) to provide a brief screening tool for the identification of children who have emotional or behavioural disturbances. It is a human figure drawing (HFD) scoring system with objectively scored items for which good psychometric properties have been demonstrated. The tool was normed on a relatively recent standardisation sample of 2260 children aged 6-17 years, which was representative of the USA population in terms of gender, geographic region, race and socioeconomic status. With reference to previous HFD scoring systems, in developing the DAP:SPED, Naglieri et al. (1991) stated that the rules for scoring HFDs needed to be better refined and made more objective, where holistic ratings of HFDs related significantly to gross levels of adjustment and maladjustment, rather than to particular pathological issues. It was also proposed that the number of emotional indicators present in a drawing need to effectively differentiate normal from maladjusted individuals, and the Draw-A-Person technique should be used for both cognitive and emotional aspects to provide a rich source of information.

The items included in the DAP:SPED scoring system were first identified following a systematic review of literature on the emotional indicators in HFDs, and were then subjected to an actuarial analysis by age to determine which items occurred infrequently (<16 per cent) in the standardisation group. Items demonstrating poor internal reliability were excluded. This resulted in a scoring system with 55 items, recording the presence and absence of emotional indicators in 3 HFDs (man, woman, self), where the items include omissions of body parts, shading, erasures of particular areas and idiosyncratic features. To improve reliability and ease of scoring, a set of transparencies was developed to measure height, orientation and size of the drawing. Once the items are scored for each drawing, they are summed and converted to a standard *T* score (mean = 50, SD = 10) and then compared with the normative group to determine whether further evaluation is i) not needed, ii) needed or iii) strongly indicated. As significant differences were found by gender within each age group, separate norms were generated, and no significant differences were found between mean scores of different races in the standardisation sample.

The instructions provided to the children undergoing the DAP:SPED test are standardised to ensure that all administrations of the test follow the exact same procedure to produce results that can be compared to the norms. As the test can be administered individually or in groups, the instructions were devised in a way to make this possible:

“I’d like you to draw some pictures for me. First I’d like you to draw a picture of a man. Make the very best picture you can. Take your time and work very carefully, and I’ll tell you when to stop. Remember, be sure to draw the whole man. Please begin. (Allow 5 minutes).”
(Naglieri, et al. 1991: 21)

These instructions are then repeated for each drawing with just the change of ‘man’ to ‘woman’ and ‘yourself’. The instructions given in the NCDS were also standardised, however they were instructions to the administrator rather than the children themselves:

“Please ask the child to ‘make a picture of a man’ on the reverse side of this sheet within the rectangular frame. Ask the child to make the best picture he/she can and remind him/her, please, to draw a whole man, not just the head and shoulders.

The drawing should be done in pencil and in conditions free from distraction. If more than one child is doing this task, please ensure that no child can see another’s drawing. The child should be given no help other than the instructions.”

In essence, these instructions are very similar, and similar guidance is given to the administrator of the DAP:SPED, in that they are to ensure children cannot see each other’s drawings in a groups setting and other than reiterating what is given in the standardised instructions, children are not to be given any other help. It is essential that the drawings were completed using similar procedures for the application of the DAP:SPED scoring system to be valid.

2.2.2. Rutter Parent Questionnaire (RPQ)

Rutter et al. (1970) devised parental and teacher questionnaires in order to screen for behavioural and emotional disturbances in children by making developmental assessments of their behaviour. The original parental questionnaire contains a total of 31 items under subheadings of health problems, habits and behaviours, which are scored additively on an applicability scale, i.e. does not apply (0), applies somewhat (1), and certainly applies (2), referring to behaviours within the last 12 months. The likelihood of disorder is indicated using a cut-off score of 13, and the type of disorder is determined by the subscale with the greater score (equal subscale scores indicate comorbidity). The two main subscales are Conduct Disorder (externalising problems) and Emotional Disorder (internalising problems): Conduct Disorder is characterised by delinquent, antisocial and destructive behaviour, restlessness, fidgeting and overactivity; Emotional Disorder is characterised by subgroups of disorders such as

anxiety disorders (fears and phobias), obsessional/anxiety disorders, depressive disorders and tics.

The NCDS (1965) used a modified version of the Rutter Parent Questionnaire (RPQ) and Goodman et al. (2007) advise that out of the original 18 items pertaining to the behavioural domain, only 13 items closely resemble them in the NCDS modified version. The original and modified items can be found in Appendix A. The modified version also uses a frequency scale as opposed to the original applicability scale, therefore the wording of the items had to be changed accordingly. Goodman et al. (2007) conducted an investigation into whether such modifications would have detrimental consequences on the validity of the scale, and found that it still held good concurrent validity ($r = .73$) with the Strengths and Difficulties Questionnaire (SDQ), which is also a behavioural screening tool. This was also evident for the modified RPQ internalising (3 items) and externalising (4 items) subscales, which correlated with the corresponding subscales of the SDQ with correlation coefficients of .65 for both (Goodman et al., 2007).

2.3. Sampling

Following Goodman et al.'s (2007) recommendation, the NCDS RPQ items that most resembled those in the original Rutter behavioural questionnaire were identified and the scores were re-coded in order to compute meaningful total scores, where the total score ranged from 0 to 26. The items contributing to the total RPQ score are as follows:

1. Is squirmy or fidgety
2. Destroys own or others' belongings (e.g. tears or breaks)
3. Fights with other children
4. Worries about many things
5. Prefers to do things on his/her own rather than with others
6. Is irritable, quick to fly off the handle
7. Is miserable or tearful
8. Has twitches or mannerisms of the face, eyes or body
9. Sucks thumb or finger during day
10. Bites nails
11. Is disobedient at home
12. Has difficulty in settling to anything for more than a few moments
13. Is upset by new situations, by things happening for the first time

Goodman (1994) recommended using the 80th percentile of total RPQ scores to identify disordered children, however it was decided to use the 95th percentile in the current study to identify the children whose total RPQ scores were in the highest 5 per cent, indicating a considerable level of disturbance. This yielded a total cut-off score of 12, and it is from these cases that the two extreme disturbance groups were defined. To identify whether the child was emotionally or behaviourally disturbed, total subscale scores were computed, where items 4, 7 and 13 formed the emotional subscale and 2, 3, 6 and 11 formed the behavioural subscale. The subscale with the higher total score defines the type of disturbance and equal total scores on the subscales indicates comorbidity (Goodman et al., 2007). A crosstable of the total

scores on the subscales was produced to facilitate the identification of the cut-off scores to be used in order to avoid comorbidity.

Table 1. Crosstable of RPQ Internalising and Externalising subscales total scores for the 95th percentile

Internalising	Externalising									Total
	0	1	2	3	4	5	6	7	8	
0	0	0	1	1	1	7	7	3	2	22
1	0	0	0	5	8	10	9	12	5	49
2	0	0	2	18	34	44	35	7	5	145
3	0	2	7	42	55	57	28	9	2	202
4	1	4	14	27	51	40	19	13	2	171
5	2	3	18	28	29	19	14	4	4	121
6	0	2	8	7	8	11	6	3	1	46
Total	3	11	50	128	186	188	118	51	21	756

Table 1 shows the total scores on both subscales, and the extreme groups were selected according to the shaded areas; children with a total score of 6 and above on the externalising subscale and 3 or below on the internalising subscale were identified as behaviourally disturbed and children with a total score of 4 and above on the internalising subscale and 3 and below on the externalising subscale were identified as emotionally disturbed. Those outside the shaded area were considered as having comorbid disturbance. A random sample which was stratified by gender was drawn from both groups; the behavioural disturbance group ($n = 45$) comprises 33 males and 12 females, and the emotional disturbance group ($n = 46$) comprises 23 males and 23 females.

As can be expected when accessing archived historical data, some of the cases were missing or the quality of the scanning of some drawings too poor to be able to effectively score. Therefore further sampling was carried out to replace the missing cases, resulting in the scoring of a total of 176 drawings; normative sample $n = 80$, behavioural disturbance group $n = 50$, emotional disturbance group $n = 46$.

2.4. Procedure

A secondary data analysis was carried out on the historical longitudinal data which is held by the UK Data Service (available to download by researchers at www.ukdataservice.ac.uk). Once confidentiality agreements had been undertaken, the data from the NCDS 1965 sweep was obtained and the groups were sampled. The children's drawings from this sweep are held by the Centre of Longitudinal Studies (CLS) as part of an electronic archive that holds scans of completed questionnaires and other survey instruments, and on microfiche. They can only be accessed directly in a secure computing environment at the CLS offices.

Prior to accessing and scoring the drawings using the DAP:SPED system, competency training was carried out, as is recommended by Naglieri et al. (1991) to ensure accurate scoring. The training consists of 5 practice drawings which serve as an aid to learning the procedure, as it is possible to compare one's own scoring with the correct scoring and descriptions provided by the authors. Following the training, a competency exam was undertaken which consists of 5 drawings (2 self, 1 woman, 2 men) including a range of different features, and drawn by a range of ages. The standard of accuracy outlined by Naglieri et al. (1991) is 90 per cent correct; the level of accuracy achieved for this study was 97 per cent correct.

The drawings were then scored according to the criteria outlined by Naglieri et al. (1991) in the DAP:SPED Examiner's Manual, using the provided transparencies and checklist of items. To maintain objectivity, the scoring procedure was followed literally without any attempts to interpret what the child was aiming to portray in their drawing. Additionally, to minimise any bias when scoring the drawings, the scorer was blind to the case grouping.

As the Draw-A-Man test used in the NCDS 1965 sweep was the Goodenough-Harris (1963) version, certain adaptations were made in order to be able to score the drawings using the DAP:SPED. First and foremost, the NCDS Draw-A-Man test only required the drawing of a man unlike the DAP:SPED, which requires a drawing of a man, woman and self. This resulted in obtaining raw scores for one drawing, meaning that the subsequent NCDS drawing scores could not be converted to the standardised scores used by the DAP:SPED, as this conversion relies on the sum of the raw scores for the three drawings. Additionally, as the NCDS drawings are kept in an electronic archive they were scored on a computer screen, therefore utmost effort was undertaken to ensure that these drawings were to original scale to maintain objectivity when measuring proportions using the transparencies. As none of the DAP:SPED items concerned the weight of pencil-lines, it was possible to score all items using an electronic version of the drawings.

3. Results

To find out whether the DAP:SPED identified indicators of emotional or behavioural problems in the sample of NCDS drawings, mean comparisons between the RPQ normative, emotional and behavioural groups were computed. The following subsections outline the distribution of the DAP:SPED, preceded by a mean comparison of DAP:SPED scores across all three groups.

3.1. Distribution analysis

The distribution of the DAP:SPED was analysed in total and according to the gender stratification, and the descriptive statistics of the children's total scores organised by group membership can be found in Table 2.

Normative Group

The distribution of the overall normative group's DAP:SPED scores resembles that of a normal distribution, with relatively low skewness and kurtosis figures. When broken down by gender, the normative group's mean scores on the DAP:SPED are slightly higher for the males than females, however both mean scores are relatively close to their respective median. The standard deviation for both genders is very close, indicating the scores are spread equally. The females have a higher max score of 11 and although the skewness for both genders is low, the female scores are modestly more positively skewed. This may be as a result of the two outliers that can be seen in the boxplot in Figure 1. The kurtosis of the male scores is negative indicating the peak is somewhat flattened around the mean, however as with the positive peak of the female scores, this is not to the extent that would indicate a major deviation from a normal distribution.

Behaviourally Disturbed Group

When considering the overall DAP:SPED scores in the behaviourally disturbed group, it can be seen in Table 2 that this group has the highest mean score, and the data is more widely spread than the normative and emotionally disturbed groups. The skewness and kurtosis figures indicate that the scores in the behaviourally disturbed group resemble a normal distribution, which has a minimally flattened peak around the mean.

The mean score for the females is higher than the males' mean score, and this difference can also be observed in the respective medians. The difference in standard deviation would indicate that the female scores are more widely spread than the males', and the females also have a higher max score. The skewness and kurtosis for both genders are relatively low and therefore approximate normal distributions; the slightly more elevated skewness of the male scores in comparison to the female scores can be observed in the boxplots (Figure 1).

Emotionally Disturbed Group

Of all three groups, the emotionally disturbed group has the lowest overall DAP:SPED mean score. However as can be seen in Table 2, the distribution of

scores is moderately more positively skewed and peaked than the other two groups. As this group has a higher range of scores, the skewness can be accounted for by the outliers (one of which is extreme) identified in the boxplots in Figure 1.

Unlike the behaviourally disturbed group, the mean score for the males in the emotionally disturbed group is somewhat higher than the mean score of the females. The standard deviation of the scores indicates that the male scores are spread more widely than the female scores, as can be expected on observing the considerably higher male max score. The skewness of the male scores is moderate and as can be observed in the boxplot (Figure 1) the distribution is affected by two outliers and one extreme outlier. The kurtosis of the male scores is also considerably high, therefore indicating a deviation from a normal distribution. The skewness and kurtosis of the female scores are comparatively low, suggesting a closer resemblance to a normal distribution.

Table 2. Descriptive statistics of DAP:SPED total scores for the normative and behavioural problem groups

		n	Mean	Median	SD	Min- Max	Skew	Kurtosis
Normative	Males	41	5.01	5.00	2.30	1-10	.20	-.97
	Females	39	4.72	5.00	2.36	1-11	.85	.51
	Total	80	4.88	5.00	2.32	1-11	.51	-.36
Behavioural Disturbance	Males	37	5.49	5.00	2.46	2-11	.40	-.49
	Females	13	6.31	6.00	3.33	1-12	.20	-.56
	Total	50	5.70	6.00	2.70	1-12	.41	-.39
Emotional Disturbance	Males	22	5.00	4.50	2.76	2-13	1.73	2.96
	Females	24	4.25	4.00	1.73	1-7	-.09	-.49
	Total	46	4.61	4.00	2.29	1-13	1.56	3.93

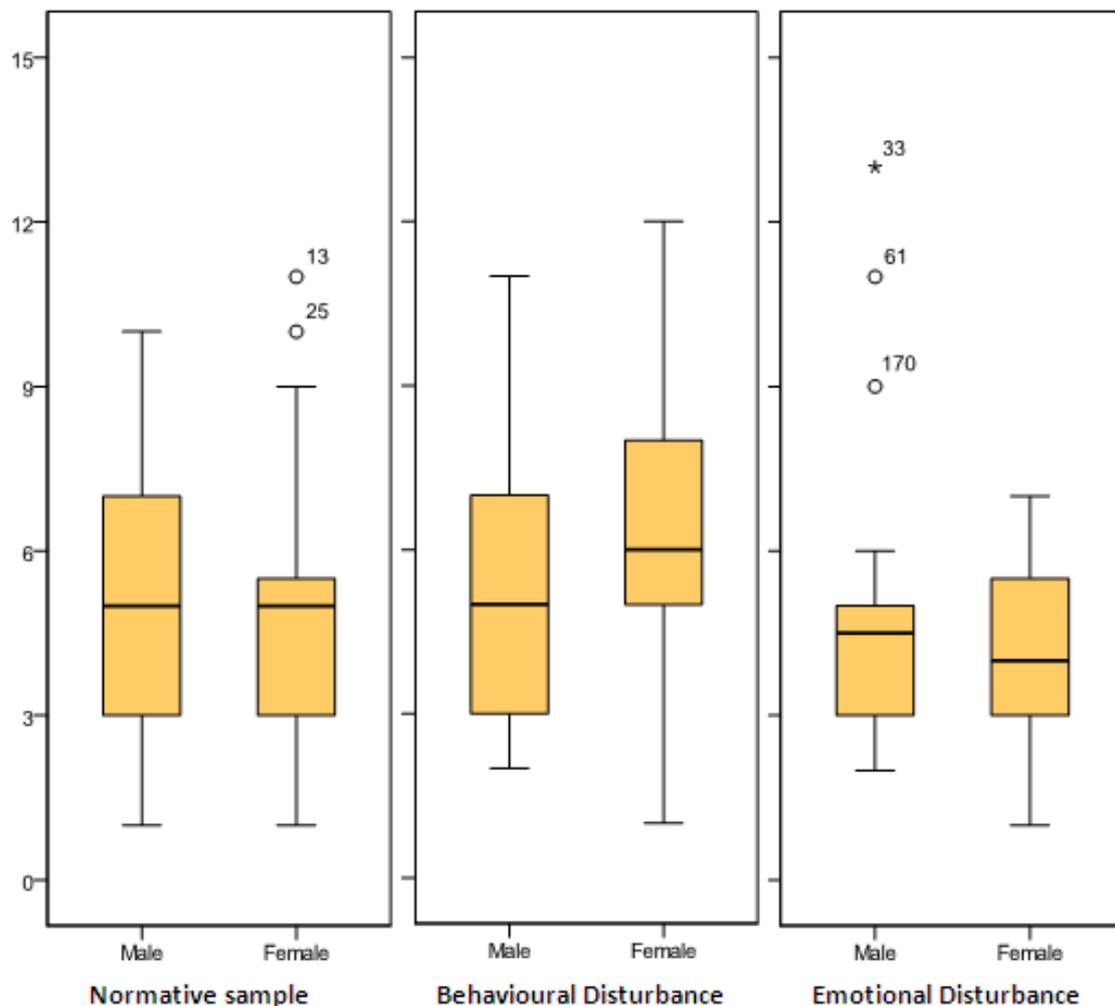


Figure 1. Box plots showing the distribution of the DAP:SPED total scores across all samples.

3.2. Mean comparison of Draw-A-Person: Screening Procedure for Emotional Disturbance (DAP:SPED) scores between the samples using pairwise comparisons

To assess whether there was a difference in the number of emotional indicators in the NCDS drawings between the RPQ groups, the mean DAP:SPED scores of the normative and both disturbance groups were compared. As can be seen in Table 2, in comparison to the mean DAP:SPED score of the normative group ($M = 4.88$), the behaviourally disturbed group had a higher mean score ($M = 5.70$), whereas the emotionally disturbed group had a lower mean score ($M = 4.61$). To find out whether the mean differences in these scores were reliable, a one-way ANOVA was computed.

The result of the Levene's test of homogeneity of variances was $F(2, 173) = 1.83, p = .163$, signifying that the homogeneity assumption of the ANOVA was met. The ANOVA indicated a borderline significant mean difference between the three groups, $F(2, 173) = 16.26, p = .066, \eta^2 = .03$, demonstrating that group membership

accounted for 3% of the explained variance in the DAP:SPED scores. This is a very small effect indicating the DAP:SPED was not particularly sensitive to the differences in behavioural problems between the groups.

To identify significant mean differences between the groups, a Tukey's HSD pairwise comparison was run as a post-hoc test (familywise $\alpha = 5$ per cent). The mean differences in DAP:SPED scores between the normative group and both behavioural problem groups were not significant. A borderline significant finding however was found in the mean difference between the two disturbance groups: mean difference = 1.09, $p = .073$.

4. Discussion and conclusion

The current study found that there were no significant differences in drawings of a man between children with and without behavioural problems as identified by the RPQ. However the behavioural disturbance group was found to significantly include more indicators of disturbance in their drawings than the emotional disturbance group, suggesting the DAP:SPED may be able to identify different types of disturbance. As the mean scores were based on the raw scores of the drawings, the mean difference of 1.09 between the two behavioural problem groups indicates that children presenting externalising behaviours were, on average, including one more indicator of disturbance in their drawings than children presenting internalising behaviours. However this difference in means hinges on the accuracy of the parameter estimates of the two groups, and there is therefore a possibility that this difference may not exist in the population.

Previous studies have not compared the mean differences in DAP:SPED scores between groups with different types of behavioural problems, as the majority of validity studies compared either behaviourally or emotionally disturbed samples to non-clinical samples. The difference between the behavioural problem groups found in the current study can indicate that i) the DAP:SPED is more sensitive to behavioural disturbance than emotional, ii) behaviourally disturbed children include more indicators of disturbance in their drawings, or iii) there was a substantial amount of misclassification of behavioural problems according to the RPQ. This final point may also provide an explanation for the lack of significant mean differences between the normative and behavioural problem groups.

The RPQ uses parental judgement of behaviour in order to identify children who may have potential behavioural problems, either of an internalising or externalising nature. However, by relying on a parent to make such judgements, there is the possibility that certain symptomatic behaviour may be under- or over-reported. By under-reporting symptoms, children who should have been classified with internalising or externalising problems were not, and may have been oversampled in the normative group, and thus inflated the mean number of disturbance indicators in the drawings. An over-reporting of symptoms would result in the inclusion of children who should not have been identified as disturbed in the behavioural problem groups, and the lack of disturbance indicators in their drawings may have contributed to an overall reduction in the group mean.

The externalising and internalising subscales of the RPQ used in the NCDS (1965) are formed on 4 and 3 items respectively, where the parent judged the child's behaviour over the past 3 months and indicated the frequency of the observed behaviour. Externalising behaviour is acted out by a child and is therefore clearly observable, however it is also quite easily over- or under-reported. Items such as 'is disobedient at home' depend on the parent's interpretation of what constitutes disobedience, and possibly even on their parenting and disciplinary style. Similarly, 'fights with other children' may be a type of behaviour that is more often observed at school, or somewhere a parent might not be able to find out about. Internalising behaviour on the other hand, is more difficult to ascertain than externalising behaviour as it concerns behaviours or emotions that may only be known to the child.

Researchers such as Matto (2002) have noted how difficult it is to identify internalising symptoms, and with so few items in the RPQ, parental judgement of this type of behaviour is prone to error. Certain symptoms may not always be noticeable, for example the item 'worries about many things' refers to an emotion that may not necessarily be expressed. Parents may also misjudge a particular occurrence as a symptom, for example the item 'is miserable or tearful' may be applied to different contexts which do not indicate disturbance, such as a child being upset after having fallen over. The added difficulty with such measures is that if a parent believes their child has a behavioural problem, they may be more likely to apply this belief when completing a questionnaire, therefore biasing their response and the subsequent identification of a problem.

The type of measurement error that is brought to parent questionnaires may have systematically contributed to the sampling error of the two behavioural problem groups that were identified using the RPQ. Previous studies found differences in DAP:SPED scores between children with and without a clinical diagnosis of behavioural problems. Typically, these diagnoses would have been made using the International Classification of Diseases (ICD-10) and The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), which are international guidelines for the diagnosis of psychiatric and behavioural disorders. Children identified by the RPQ as having behavioural problems may not have received a clinical diagnosis, contributing to the DAP:SPED's lack of sensitivity to disturbance in this study.

4.1. Limitations

The RPQ that was used in the 1965 sweep of the NCDS was a modified and shortened version of Rutter et al.'s (1970) original, which may have compromised the validity of the identification of behavioural problems, and thus contributed to sampling error of the three groups used for comparison. Using a parent questionnaire may not be an appropriate way to identify disturbance groups to compare DAP:SPED scores; previous studies used clinically diagnosed children, and the lack of this diagnosis and possible classification error could have led to the lack of significant differences between the normative and behavioural problem groups in the current study. Although using a measure that is completed by someone other than the child themselves relies upon external observations of behaviours that may not be ascertained with ease (i.e. internalising behaviours), using a self-report measure may not be appropriate considering the age of the participants and how this can affect the way symptoms are reported. Therefore a more feasible option would involve the triangulation of behavioural observations (using a tool such as that developed by Rutter et al., 1970) made by parents and teachers, and the child can be observed in two environments where some symptoms may be more observable in one or the other. The Rutter Teacher Questionnaire was not completed until the 1974 sweep of the NCDS, when the cohort would have been aged 16. Due to the large gap between the teacher questionnaire and when the drawings were completed, this data could not be used as the HFD screening tools are only valid to assess the child's internal state at that moment in time (Koppitz, 1968).

The procedure for obtaining DAP:SPED scores for mean comparison in the current study was an adaptation of the procedure outlined by Naglieri et al. (1991) in the test manual. The conditions of obtaining the drawings in the NCDS (1965) have been

documented to be similar to those in the procedure of the DAP:SPED, however minor differences may have systematically contributed to measurement error, and only one drawing was produced. The lack of drawings of a woman and the child's self meant that the raw scores for the NCDS drawings could not be standardised and compared with the norms provided in the DAP:SPED manual, as the standardised score is obtained using the sum score of the three drawings. This also meant that the normative scores derived from the British birth cohort sample could not be compared to the norms of the DAP:SPED standardisation sample, which was representative of American children.

In light of the limitations outlined in this section, an initial examination of whether the DAP:SPED items do in fact occur in <16 per cent of British children's drawings, due to the finding of a relatively high mean DAP:SPED score in the current study's normative group, should be conducted. Once this has been established, whether children identified with behavioural problems score higher on the DAP:SPED can be investigated. As there may be substantial measurement error associated with the use of behavioural questionnaires to identify children with behavioural problems, the use of clinically diagnosed samples would be recommended to examine the DAP:SPED's discriminant validity. Although this information is available in a later sweep of the NCDS (when the cohort is aged 16), as with the Rutter Teacher Questionnaire, the time lapse would reduce the accuracy of findings as HFDs are representative of the child's state of mind at the time of drawing (Koppitz, 1968).

4.2. Conclusion

The current study used data from the NCDS to investigate whether children's drawings can be used to predict emotional and behavioural problems. The Draw-A-Person: Screening Procedure for Emotional Disturbance (DAP:SPED, Naglieri et al., 1991) was used to identify whether children who had been identified by the RPQ as having internalising or externalising behavioural problems included indicators of such disturbance in their drawings of a man. It was found that there were no significant differences between the drawings of the children in the normative group compared to the behavioural problem groups. In light of these findings, issues with using a parent questionnaire to identify behavioural problem groups for comparison to a normative group were discussed, as was the matter of adapting a standardised scoring procedure. Although there was a lack of significant findings in the current study, it was possible to analyse potential explanations for these results, providing a rationale for the use of clinically diagnosed samples to eliminate the potential bias in classification caused by behavioural questionnaires completed by parents.

References

- Aikman, K. G., Belter, R. W., & Finch Jr, A. (1992) 'Human figure drawings: Validity in assessing intellectual level and academic achievement', *Journal of Clinical Psychology*, 48: 114-120.
- Bekhit, N. S., Thomas, G. V., & Jolley, R. P. (2005) 'The use of drawing for psychological assessment in Britain: Survey findings', *Psychology and Psychotherapy: Theory, Research and Practice*, 78: 205-217.
- Butler, N. R., Peckham, C., & Sheridan, M. (1973) 'Speech defects in children aged 7 years: a national study', *British Medical Journal*, 1: 253-257.
- Davie, R., Butler, N. & Goldstein, H. (1972) *From Birth to Seven: The Second Report of the National Child Development Study (1958 Cohort)*. London: Longman in association with the National Children's Bureau.
- Flanagan, R., & Motta, R. W. (2007) 'Figure drawings: A popular method', *Psychology in the Schools*, 44: 257-270.
- Goodenough, F. L. (1926) *Measurement of intelligence by drawings*, New York: Harcourt, Brace and World
- Goodman, R. (1994) 'A modified version of the Rutter parent questionnaire including extra items on children's strengths: a research note', *Journal of Child Psychology and Psychiatry*, 35: 1483-1494.
- Goodman, R., Iervolino, A. C., Collishaw, S., Pickles, A., & Maughan, B. (2007) 'Seemingly minor changes to a questionnaire can make a big difference to mean scores: A cautionary tale', *Social psychiatry and psychiatric epidemiology*, 42: 322-327.
- Harris, D. B. (1963) *Children's drawings as measures of intellectual maturity: A revision and extension of the Goodenough Draw-a-Man Test*, New York: Harcourt, Brace & World.
- Kanazawa, S. (2011) 'Intelligence and physical attractiveness', *Intelligence*, 39: 7-14.
- Knoff, H. M. (2003): Evaluation of projective drawings. In C.R. Reynolds & R.W. Kamphaus (Eds) : *Handbook of psychological and educational assessment of children: Personality, behavior and context* (pp 91-158). New York: Guilford.
- Koppitz, E. M. (1968) *Psychological Evaluation of Children's Human Figure Drawings*, London: Grune & Stratton.
- Lynn, R., & Kanazawa, S. (2011) 'A longitudinal study of sex differences in intelligence at ages 7, 11 and 16 years', *Personality and individual differences*, 51: 321-324.

- Matto, H. C. (2002) 'Investigating the validity of the Draw-A-Person: Screening procedure for emotional disturbance: A measurement validation study with high-risk youth', *Psychological Assessment*, 14: 221-225.
- Matto, H. C., Naglieri, J. A., & Clausen, C. (2005) 'Validity of the Draw-A-Person: Screening Procedure for Emotional Disturbance (DAP:SPED) in Strengths-Based Assessment', *Research on Social Work Practice*, 15: 41-46.
- McConaughy, S. H. (1992) 'Objective assessment of children's behavioural and emotional problems'. In C. E. Walker & M. C. Roberts (eds), *Handbook of Clinical Child Psychology* (2nd edn), New York: John Wiley & Sons.
- McNeish, T. J. (1989) 'Reliability and validity of a human figure drawing screening measure to identify emotional and behavioural disorders in children', Unpublished doctoral dissertation. Ohio State University. Columbus.
- McNeish, T. J., & Naglieri, J. A. (1993) 'Identification of individuals with serious emotional disturbance using the Draw A Person', *Journal of Special Education*, 27: 115-121.
- Naglieri, J. A., McNeish, T. J., & Bardos, A. N. (1991) '*Draw a Person: Screening Procedure for Emotional Disturbance; DAP: SPED*, Texas: PRO-ED.
- Naglieri, J. A., & Pfeiffer, S. I. (1992) 'Performance of disruptive behavior disordered and normal samples on the Draw A Person: Screening Procedure for Emotional Disturbance', *Psychological Assessment*, 4: 156-159.
- Plbrukarn, R., & Theeramanoparp, S. (2003) 'Human Figure Drawing Test: Validity in Assessing Intelligence in Children Aged 3-10 Years', *J Med Assoc Thai*, 86: S610-S617.
- Rutter, M., Tizard, J., & Whitmore, K. (1970) *Education, health and behaviour*. London: Longmans.
- Schoon, I., Bynner, J., Joshi, H., Parsons, S., Wiggins, R. D., & Sacker, A. (2002) 'The Influence of Context, Timing, and Duration of Risk Experiences for the Passage from Childhood to Midadulthood', *Child Development*, 73: 1486-1504.
- Trevisan, M. S. (1996) 'Review of the draw a person: Screening procedure for emotional disturbance', *Measurement and Evaluation in Counseling and Development*, 28: 225-228.
- Willcock, E., Imuta, K., & Hayne, H. (2011) 'Children's human figure drawings do not measure intellectual ability', *Journal of Experimental Child Psychology*, 110: 444-452.
- Wrightson, L., & Saklofske, D. H. (2000) 'Validity and reliability of the Draw A Person: Screening Procedure for Emotional Disturbance with adolescent students', *Canadian Journal of School Psychology*, 16: 95-102.

Appendix A. Original behavioural domain items in the Rutter Parent Questionnaire and modified NCDS (1965) items

	Standard Rutter	NCDS (1965) Rutter
Item 1	Squirmy, fidgety child	Is squirmy or fidgety
Item 2	Often destroys own or others' property	Destroys own or others' belongings (e.g. tears or breaks)
Item 3	Frequently fights or is extremely quarrelsome with other children	Fights with other children
Item 4	Often worried, worries about many things	Worries about many things
Item 5	Tends to be on own – rather solitary	Prefers to do things on his/her own rather than with others
Item 6	Irritable, is quick to 'fly off the handle'	Is irritable, quick to fly off the handle
Item 7	Often appears miserable, unhappy, tearful or distressed	Is miserable or tearful
Item 8	Has twitches, mannerisms or tics of the face or body	Has twitches or mannerisms of the face, eyes or body
Item 9	Frequently sucks thumb or finger	Sucks thumb or finger during day
Item 10	Frequently bites nails or fingers	Bites nails
Item 11	Is often disobedient	Is disobedient at home
Item 12	Cannot settle to anything for more than a few moments	Has difficulty in settling to anything for more than a few moments
Item 13	Tends to be fearful of afraid of new things or new situations	Is upset by new situations, by things happening for the first time
Time scale	Past 12 months	Last 3 months
Low response	Doesn't apply	Never
Middle response	Applies somewhat	Sometimes
High response	Certainly applies	Frequently
Option to respond 'Don't know/Inapplicable'	No	Yes

Centre for Longitudinal Studies

Institute of Education

20 Bedford Way

London WC1H 0AL

Tel: 020 7612 6860

Fax: 020 7612 6880

Email cls@ioe.ac.uk

Web www.cls.ioe.ac.uk