Emotional Indicators in Children's Human Figure Drawings: An Evaluation of the Draw-A-Person Test.

Michelle Catte (B.A. (Hons) Dunelm)

A thesis submitted for the degree of
Doctor of Philosophy

University of York, Department of Psychology

September, 1998

Abstract

The aim of this thesis is to evaluate the Draw-A-Person test and its clinical validity. In a series of 11 studies, the human figure drawings (HFDs) of children with emotional/behavioural difficulties (EBD) were compared to normally adjusted children, matched either for chronological age (CA), mental age (MA) or Goodenough-Harris (GH) scaled scores. Several different measures were considered: original Koppitz (1968) emotional indicator scores, revised indicator scores based on new normative data, an intuitive method of identification and ratings of bizarreness.

The original and revised emotional indicators, and both expert and novice judges using the intuitive method, failed to discriminate between mildly disturbed children's HFDs and those of controls matched for CA or MA. The indicators and judges were successful, however, using more severely disturbed children's drawings compared with both CA and MA matched children's HFDs. The severely disturbed children's drawings were also rated as more highly bizarre than the CA and MA controls.

When the severely disturbed children's drawings were compared to GH matched control HFDs, the emotional indicator differences disappeared and the judges were no longer able to discriminate the drawings successfully. No differences were found between the clinical and GH matched HFDs for ratings of bizarreness. Also, the ratings did not alter when the indicators were removed from the drawings. Visual differences were found between subsamples of drawings which were classified as disturbed or normal. These differences relate to variables involved in the GH scale and a factor discovered by Adler (1970) measuring cognitive maturity.

The results of this thesis have implications for the Koppitz indicators and the use of the DAP test. Questions are raised over the interpretation of HFDs for emotional health. The influence of cognitive maturity on the drawings of disturbed children is considered and the results are discussed in the light of drawing theories which consider either the internal representation or production process as paramount.

Acknowledgements

I should like to thank the children who produced drawings for this thesis, and their teachers, for the assistance they gave. My gratitude goes to the art therapists who gave their time and shared their enthusiasm for children's artwork; Derek Roger and Peter Bull for their advice on my research committee; and my supervisor, Maureen Cox, for her guidance and support. I should also like to acknowledge my friends in 'The Zoo' who made the experience more enjoyable, and my partner, Jasper, for always believing I could do it.

Contents

List of figures

List of tables

Study 1

CHAPTER ONE

CHILDREN'S DRAWINGS IN PSYCHOLOGY

Historical Overview

The study of children's drawings in psychology dates from the latter part of the last century, when childhood itself began to be seen as a distinct stage of development. Children, therefore, became viable subjects of research and drawing development was included in case studies such as the 'baby biography' of Darwin's own son (1877), as well as involving larger scale collections (e.g. Lamprecht, 1906; Claparède, 1907; Ivanoff, 1909; Rouma, 1913, cited in Harris, 1963).

Early work established the developmental character of drawing, and investigations were primarily descriptive in nature. The first of the developmental models was proposed by Cooke (1885, cited in Goodenough, 1926), who sought to influence the art education of children through knowledge about the development of drawing. There were many more studies describing the nature of children's drawings and establishing their developmental character, culminating in 1921 with Burt's model. Seven stages were defined altogether, from scribbling to visual realism and repression in young teenagers, to artistic revival in adolescence, a stage never achieved by many people.

Luquet (1927) proposed perhaps the most significant of the developmental models that is still considered relevant today, though it does bear some similarities with that of Burt. Luquet's five stage model contained an underlying theory which made it stand out from the others at the time. He believed that the child based the drawing on his/her internal model of the object, intending it to be realistic at the same time. Various factors such as the drawing medium and artistic ability would influence how this internal model was drawn. Luquet greatly influenced the work of Piaget who was very influential in developmental psychology during the middle part of the 20th century. Piaget's interest in children's drawing was mainly as an illustrative technique for his theories of spatial cognition and his interest in children's drawing was not integral to his theory of cognitive

development. Whilst Piaget's influence held, the study of children's drawing was neglected for many years.

The popularity of the human figure drawing (HFD; see glossary, page 245, for list of abbreviations used in the thesis) in children's spontaneous artwork was realised early on in surveys (e.g. Maitland, 1885; Lukens, 1896; Ballard, 1912; Luquet, 1913; Hurlock, 1943, cited in Harris, 1963). The use of children's HFDs for assessment purposes originated in the belief in the developmental nature of drawing and Luquet's theory that a child's drawing of an object was related to his/her concept of that object and could therefore be used as a measure of mental development. Schuyten (1904, cited in Goodenough, 1926) had been one of the first investigators to try to devise an objective measuring scale for children's HFDs, based on age related normative data. Goodenough (1926) quantified this development for the psychometric study of intelligence in her book 'The Measurement of Intelligence by Drawings' which was revised and updated by Harris (1963).

Goodenough devised the Draw-A-Man (DAM) test which credited a child with a score for the number of features and correctness of proportions which were included on his/her drawing of a man. Though attempts were made to validate this point-scale system for adolescents (e.g. Cohen, 1933; Levy, 1931), the test was found unable to sufficiently discriminate differences in intellectual functioning beyond 11 or 12 years old (Harris, 1963). It later came to be regarded as a measure of the child's intellectual maturity rather than intelligence due to its measurement of the child's actual rather than potential level (Cox, 1993). Psychometric properties of the DAM test showed that the Goodenough IQ correlated reasonably well with other intelligence test scores such as the Stanford-Binet (Goodenough, 1926; Yepsen, 1929; Williams, 1935), and showed good inter-rater and test-retest reliability (McCarthy, 1944). The DAM test was revised in 1963 with Harris, becoming the Goodenough-Harris test, with scales to assess drawings of both men and women.

Since the Goodenough-Harris DAM test was developed, other researchers have developed tests of intelligence using the child's drawing of the human figure. Koppitz (1968) used her Draw-A-Person

test, scored for items considered 'exceptional' and 'expected' based on age-related normative data, to arrive at a broad score of intellectual functioning. Correlations with standard intelligence tests showed this method to be as valid as the DAM test, correlating reasonably well with the Stanford-Binet and WISC (Koppitz, 1968). This method is quick and simpler than the Goodenough-Harris method, but gives rather vague and very broad final scores which have limited use compared with the more comprehensive DAM test (Gayton, Tavormina, Evans & Schuh, 1974).

More recently, Naglieri (1988) has developed the 'Draw-A-Person: Quantitative Scoring System' which is very similar to the way that the DAM test works, on a point scale method crediting features included and proportions used, with norms gathered during the 1980s. Reliability coefficients showed good internal consistency and retest reliability has also been found to be good over a four week period (Naglieri, 1988). This new test, however, though demonstrating good psychometric properties, has no independent evidence supporting it and has yet to gain popular status with clinicians and educational psychologists.

The clinical-projective approach to the use of children's drawings stemmed from the work of Goodenough and the Psychoanalytical movement and examined the emotional status and personality of the artist rather than his/her cognitive functioning. The psychoanalytic tradition already used projective techniques such as the Rorschach to assess personality and emotional health, and in her research on the measurement of intelligence, Goodenough had anticipated an interpretative use of children's drawings after noticing that some aspects of children's drawing seemed unrelated to their intellectual level and more to do with their personality (Harris, 1963). Early interpretation of spontaneous pictures and paintings utilised projective principles (e.g. Alschuler & Hatwick, 1947) but it was the Draw-A-Person (DAP) test which quickly became used as a projective technique. A person's body image or self-concept was seen as being unconsciously projected onto the figure drawn (Machover, 1949), allowing for interpretation of the artist's emotional functioning.

Individual features of a drawing were given clinical significance such as the stance of the figure drawn reflecting the emotional stability of the subject, or the head being symbolic of intellectual power (Machover, 1949). Research assessing the reliability and validity of the claims made regarding this sign interpretation of HFDs found mixed results (Swensen, 1957, 1968).

Using collective features was found to be more valid and reliable and total scores of 'emotional indicators' (EI) were found to be higher in clinical populations (Koppitz, 1968). Global ratings of adjustment or bizarreness and assessment of clinical status using drawings also achieved more success. The clinical-projective approach has been criticised for lacking a theoretical base (Harris, 1963; Mortensen, 1991) and positive evidence (Motta, Little & Tobin, 1993) though the use of children's drawings as a diagnostic tool still prevails today.

Contemporary research into children's drawings began with a change of focus in this area in the 1970s as the study of children's drawings came to include the *process* of drawing itself and not just the finished product. The idea of drawing as a problem solving task rather than a 'window on the mind' (Buros, 1972) focused research on the strategies by which the child represents a concept using his/her graphic skill. Previous reliance on the finished product ignored the procedural problems faced by the child when drawing (Thomas & Silk, 1990). Freeman (1980) investigated the task demands involved in the process of drawing, analysing the drawing in terms of the cues children use when constructing a picture. Investigations into children's drawing as an artful and constructive activity focused on the performance breakdowns that affect the child's drawing. Thus a large oversized head on the human figure can be seen as the result of the lack of forward planning by the child rather than a sign of the importance the child places on the head, as in the clinical-projective approach. Similarly, 'transparencies' can be explained in terms of the sequence of drawing rather than being symbolic of emotional disturbance. Thomas and Silk point out that this approach to children's drawing helps in the understanding of planning and organising skills in general.

Theories of Drawing

Stage models

The study of children's representational drawings was one of the first areas of child psychology to become established as a field of research. The sequential nature of children's drawing development was established early on, with Luquet's model becoming the accepted and popular version. Rigid sequencing was seen as orderly and secure and though psychologists since then have criticised parts of this stage model for its basis in a single case study and mentally ill patients (Paine, 1992), its influence has held. This may be due to the fact that Luquet's five stage model incorporated a theory, as well as a description of drawing development, and also because of its use by Piaget as an illustrative framework for his theory of spatial representation.

Luquet hypothesised that the ultimate aim of the developmental progress of drawing is realism, so the "goal of drawing would be a realistic translation of the visual properties of objects into graphics" (Krampen, 1991, p. 38). Children's drawing is seen as a direct reflection of their internal model of the object and thus the child aspires towards a visually realistic depiction of that object. The classical idea of ideal forms that are used as a standard is still widespread and the most popular opinion of drawing development is that of a journey towards the universal objective of accurate representation.

Pre-Representational Drawing At stage one, the scribbling phase (18 months to 2;6 years), the child makes marks on paper and seems more involved in the *activity* of drawing than the finished product. Scribbling is normally seen as the starting point of drawing development. Kellogg (1970, 1979) was one of the first influential writers describing children's drawing development since Luquet. She catalogued 20 different kinds of scribble forms which she considered to be the building blocks of future drawing development. Development occurred as the 20 basic scribbles became 'combines' and 'aggregates', eventually becoming representational drawings.

Scribbling can be seen as primarily action on the medium, rather than the build-up of schemas (Matthews, 1983). There is no intentional representation at this stage and the motor movement of the writing

instrument on the paper expresses the meaning of the drawing better than the finished product. Burt (1921) also included this stage in his formulation, describing it as motor expression and 'purposeless pencillings' (cited in Harris, 1963). Early drawing was seen as 'mark making' as children's body actions in the centre of a spatial layout leave visible traces of their happening (Golomb, 1992). 'Action representations' as Matthews called them are a combination of motor action and representation that act as an undifferentiated behaviour. For example, the child may scribble spirals in a continuous overlap in order to depict something going round a corner (Golomb, 1992), rotational whirls become an engine and a vehicle in motion, or dots across the page may become the jumping movement of an animal. Verbalisations can also be symbolic of the object which is being drawn, for instance pre-schoolers who, when asked to draw an 'angry' house, growled whilst drawing (Scarlett, Fucigna & Finkelstein, 1980, cited in Winner & Gardner, 1981).

Neither Kellogg's nor Matthews' theories can wholly explain how scribbling activities develop into representational activity. The problem with Kellogg's theory is that it is an adult view of the children's drawings and ignores the meaning ascribed by the child to the pictures. It was also found that only around 4% of the children produced the 'aggregate' figure which Kellogg believed preceded the human figure (Golomb, 1992). Matthews' theory puts emphasis on action symbolism in pre-representational drawing but it is unclear how this links into the ability to draw a meaningful shape and thus into representational drawing (Golomb, 1992).

The second pre-representational stage, fortuitous realism (2;6 to 5 years), is where children assign a representation to the drawing after it has been drawn or whilst it is in the process of being drawn, in a fortuitous fashion. These post-hoc decisions can mean that the drawing goes through several representations before the picture is finished. At this stage, children often recognise something in their drawing which was unrelated to any prior intention to draw that object and then proceed to alter the drawing with regard to the newly recognised object. Children are willing to reinterpret the work rather than look for an alternative method of drawing which might make the

figure more like the adult standard and obey the constraints of realism (Golomb, 1992).

Transition to representation Scribble lines alone cannot sustain meaning, as once the scribble drawing is separated from the creation process itself, the lines become meaningless. For example, the rotational swirls used to show the roar and movement of an aeroplane in flight become unintelligible when the movement and verbalisations are no longer present and only the scribble is seen. Scribbles therefore are not necessarily the beginnings of graphic representation.

A brief transitional phase exists between scribbling and representational drawing where the location of marks on the paper correspond to the features of the drawing in a spatial format. The transition to representation is marked by the child showing awareness of the possibility of representing objects by the use of shape in the two dimensional medium. This transition is characterised by the production of a final representation on paper that can be understood independently of the motor action used to create it. The transition from pre-representation to representational drawing is considered to be best marked by the advent of the circle. This closure of a single line is achieved by most children during their scribbling experience and the figure-ground characteristics of the circle mean that it is useful for representational purposes (Golomb, 1992). Studies have shown that this beginning, however, is not necessarily linked to past experience with scribbling. Children in remote villages who had never had experience of paper and pencil before, could be seen progressing quickly from pre-representational scribbles to early representations of the human figure (e.g. Harris, 1971; Haas, 1978, cited in Golomb, 1992). Millar (1975) and Kennedy (1980, 1983) found congenitally blind children producing representational forms of the human figure with no previous scribble experience. Prior visual experience may be a facilitating but not necessary condition for drawing development (Gardner, 1985). Using a drawing-on-dictation task, Golomb (1974) found that children could produce recognisable forms without much previous scribble experience. Scribbling therefore is neither necessary nor sufficient for representational drawing but can be seen as

acquainting the child with the tools necessary for future drawing development.

Representational Drawing There is evidence that young children have a great deal of constructive ability but have a lot of problems using it in different situations. The co-ordination of various parts of the drawing often fails in children of the third stage (3;6 to 5 years) and they have a problem positioning parts to make up a whole configuration. This was described by Luquet as failed realism.

Once children do begin to co-ordinate their drawing, they often use simple schemas that may be adapted for many uses. For example, a circle becomes the head and body of the human figure as well as the nose and eyes. The classic tadpole figure typical of this age can be adapted to represent animals as well as humans. The symbolic nature of the drawings children make at this age is evident in the way they tend to draw highly stylised schemas rather than attempting visually realistic pictures. Luquet explained part of the problems of this stage as being due to a general clumsiness on the part of the child and a lack of attention.

In the fourth stage, intellectual realism (5 to 8 years), children intentionally draw something but often fail to make it visually realistic. In attempting to depict what they know is there, they draw things that in reality cannot be seen. Differing viewpoints appear in the same picture as children grapple with how to draw a real recognisable object. The phrase that the child "draws what he knows, rather than what he sees" (Goodenough, 1926, p. 12) is used to describe this stage.

Some evidence exists for a separate purely symbolic stage of development distinct from, and previous to, intellectual realism. Barrett and Light (1976) asked children to draw an object from imagination and from a model, then told a story which drew their attention to particular aspects of the model object. When asked to re-draw the model object, the younger children did not alter their drawing from before the story, producing a symbolic representation of the object in both instances. The intellectual realists, however, altered their second model drawing based on the features pointed out in the story, showing that they were affected by what they subsequently knew about

the object. The visual realists drew the object as it was presented in both trials. It is unclear, however, how this purely symbolic stage differs from the highly stylised symbolic schemas seen in the failed realism stage.

Children in the fifth and final stage (>8 years), attempt to portray depth and a single viewpoint in their drawing and make the picture conform to visual realism. Highly stylised depictions of cartoons and comic style pictures appear as children's drawings become more conventional. The older children become increasingly dissatisfied with their attempts to draw in a visually realistic way and their inability to produce at the level of their expectations.

No translation of Luquet's work has been published, and the version of Luquet that became popularised was through the writings of Piaget and Inhelder (1956). In their book, 'The Child's Conception Of Space', they used the development of drawing as a framework for establishing the development of spatial representation. Luquet's failed realism stage of development became Phase 1 in Piaget and Inhelder's scheme that he characterised by 'synthetic incapacity'. Phase 2 of Piaget and Inhelder's theory was also characterised by 'intellectual realism', as children draw what they know about a scene rather than what is visually apparent. Phase 3 was also characterised by 'visual realism'. The relationships which children used to organise their drawings were assumed to reflect spatial relationships in the mind, following the idea that the drawing reflected the child's internal representation of the object.

Luquet has been taken as representing stage theorists in general and much research is based on the assumption that the development of children's drawings passes through distinct stages (Costall, 1989). The stage theorists assume that perspective is natural in visual experience, but also that depiction reflects the internal model, two contradictory ideas, since perspective as a system of depiction used by artists occurs towards the end of development. Deviations from visual realism in drawing, such as intellectual realism, therefore require explanations, usually by higher mental processes. Sense perceptions were seen as being corrupted by intelligence (e.g. Sully, 1895) and linguistic symbolism (e.g. Buhler, 1930). The phrase that 'language first spoils

drawing and then swallows it up completely' (Buhler, 1930, cited in Costall, 1989) referred to the diminishing of the early impetus towards drawing and its replacement by writing. Support can be found in examples of precocious drawing ability associated with communication deficits (e.g. Selfe, 1983).

The answer to the problem of intellectual realism may lie in the double meaning of realism - either the detailed representation of the object or the true optical impression of the object. Luquet saw intellectual realism as an attempt to produce a likeness of the represented object. Once the child begins to succeed at realism, a choice between visual and intellectual has to be made. The child opts for intellectual realism in order to maintain the persisting properties of the object and its practical significance. Luquet stressed that intellectually realistic pictures are based on a corresponding internal model rather than the sensation of the image of the object. Both intellectual and visual realism are conventions, two contradictory ways of depicting something. A typical 6 year old child would experience conflict over which one to choose. Contemporary research has shown that techniques such as partial occlusion, traditionally seen as part of the visual realism stage, can be found in much younger children's pictures (e.g. Cox, 1981; Crook, 1985) showing that children do have the ability to produce such techniques earlier than stage theorists such as Piaget would have us believe.

Process Approach

The contemporary approach to research in children's drawings, known as the process approach, acknowledges that early stage theories such as those of Luquet and Piaget had deficient analysis of the task demands of drawing. The process approach attempts to rectify this by looking at the process of drawing and how this affects its development.

Freeman (1980) believed that nothing is 'tipped out' onto paper but everything is laboriously constructed. Researchers therefore need to analyse drawings in terms of the cues the child is responsive to and how those cues are used. Freeman altered the famous statement that the child 'draws what he knows' to 'the child knows more than he draws' (Krampen, 1991). Freeman saw children's drawings like serious

caricatures, in the way identifying features are made to stand out so that the child's representation is recognisable. To children, the blank paper and its salient edges provide a large number of potential degrees of freedom that need to be reduced to a workable order to make a drawing. One solution is to draw a simple and routine configuration. Children choose the best picture that they can produce given that the production process makes different demands upon them compared with the adult. This is a view similar to that expressed by Arnheim (1974) who believed that the drawing satisfied the child's criteria, though not necessarily the adult's. The outcome should therefore be judged by the degree to which the intention has been graphically conveyed. Freeman saw a need to step outside the drawing situation and gain independent evidence of the child's abilities such as memory and cognitive capabilities of planning and monitoring tasks. Experiments have shown that there is a direct relationship between working memory and drawing which is separate from both of their relationships to age (Bensur & Eliot, 1993). Under the process approach, the drawing situation is studied in isolation and seen not only as a constructive task but also a problem solving activity, with differences noted between how younger children approach these tasks compared with older ones.

Development, however, is not just seen as becoming better with age. Simplification and schematisation errors decrease with age whereas orientation biases do not change, with even 8-10 year old children making such systematic errors as re-orienting complex drawings to a baseline (Pemberton, 1990). Factors such as inherent biases, production constraints, perceptual errors, motor or social factors and individual strategy are suggested as reasons for the errors, with orientation biases occurring more often due to their being over-determined by more of these factors than the other errors. Errors in children's drawings may be the result of the children's limited exposure to graphic models in the environment, since children normally only see the finished product and cannot easily see the process by which it was formed. Research has shown that when children are shown the production process, even pre-schoolers can improve (Pemberton & Nelson, 1987, cited in Pemberton, 1990).

Symbols in drawing

Symbolism in drawing was believed early this century to originate in depictions made by other people, rather than an analysis of reality (Oakley, 1930). A picture is not a copy of its referent but is constructed with pictorial schemas, marks created on paper which trigger a pictorial response in the viewer. Symbols in a culture are used in representational drawings in the form of perceptual conventions, which must be learnt (Perry, 1992). Children's drawing development may be the discovery or invention of pictorial schemas and procedures for creating them on paper (Thomas, 1995) and, indeed, teaching children the way to draw an object in terms of pencil strokes has been found more effective than knowledge of the object (Phillips, Inall & Lauder, 1985, cited in Thomas, 1995). The child develops a graphic vocabulary of simple shapes either for aesthetic reasons (Arnheim, 1974; Golomb, 1992) or because they are pictorially useful and easy to make (Freeman, 1980; Thomas & Silk, 1990). Formulae are then used to generate pictures, using this vocabulary.

Education is a process of teaching the notation systems and symbol skills used by a culture which must be acquired by the child. Schemas become elaborated and better proportioned with age. Some development is based on improved motor control and planning ability, most is based on copying pictures made by others; therefore the influence of pictures in the environment must be recognised. The canonical view is more likely to be recognised by the viewer and, therefore, the schema that is acquired will be the canonical view.

Gardner (1985) described a wave of symbolisation at age 3 years called 'analogical/topological mapping'. Symbols bear analogical resemblance to their referents, capturing relative sizes and shapes. Two lines added to a circle form a person. The next wave explores digital or quantitative mapping where the child is intent on getting the number of elements correct. A period of flexibility early on as the child masters the system, is followed by a decline which leads to a reluctance to experiment with symbol systems later on, and may be involved in the child's loss of interest in drawings in later childhood.

Systematic use of rules and symbols for communication in drawings may be similar to that of language (Mortensen, 1991) with the

suggestion that they may share the same conceptual basis (Pemberton, 1990). Both drawing and language start as purely expressive functions, without representational purpose, in forms such as babbling and scribbling. Mortensen thought that in drawing, the individual personally creates the symbols which are used which perhaps makes it more interesting to study those symbols, though as said above, some symbols are common to the culture and are merely copied by the child.

In its early stages, symbol formation is characterised by 'syncretism' and 'physiognomics' (Mortensen, 1991) which means that an attempt is made to adjust various ideas within the symbols and at the same time contain the expressive qualities of the object in the symbol. The difficulty of interpreting symbols lies partly in the subjective nature of their dynamic-affective qualities, chosen because they share these qualities with the referents. Though a consensus may be found as to what qualities a symbol reflects, permitting generalities in interpretation, individual differences can appear. A symbol can convey a mood, feeling or tone as long as the relevant community chooses to interpret the symbol that way (Gardner, 1985). The interpretation of symbols in drawing is implicated in the use of drawings as a projective technique, particularly when specific features are assumed to be diagnostic of particular problems.

The difference between analogic and digital communication characterises the difference between drawing and language (Mortensen, 1991). In digital communication, arbitrary signs are used as symbols and are manipulated by a logical syntax. In analogic language, there is a similarity between the signs and what they represent, such as is the case in drawing. For sharing information about objects, a digital language is good, but for describing relationships, analogic language is better. In the Goodenough (1926) DAM test, drawings are used as a very simple digital language where each feature has an all-or-none function. Drawings by their nature, however, are always analogic in communication form mainly due to the similarity in dynamic-affective qualities between the drawing symbol and its referent. When pressed into a digital language, this leads to a loss of information that may cause problems for the DAM test. The use of drawings as a projective technique is a form of analogic communication. The lack of negation in

this type of language, however, creates serious difficulties in understanding the drawing's message. For example, a drawing of a monster could represent fear or wish fulfilment. The impossibility of expressing time also makes it difficult to give a precise interpretation of this type of drawing.

Summary The stage models of drawing focus on drawing as a reflection of the internal model of a given object. The child's drawing development is seen as passing through various stages on the road towards a visually realistic representation of an object. The problem with this idea is that it ignores the influence of the production process involved in the drawing. Luquet's original work did acknowledge these factors but when his work was adapted by Piaget for the purposes of illustrating his theory of spatial development, the original ideas may have been confused. Luquet is taken as the principle founder of the stage approach, but doubt has been cast on the idea that intellectual realism is merely a stage on the way towards visual realism as the end-product of drawing development. Instead, both forms of realism can be seen as conventions, each with appropriate uses in different circumstances.

Contemporary research into drawing development sees drawing as a construction task, focusing on the performance factors ignored by other perspectives. Analysing drawing in terms of task demands and in the manner of a problem solving activity has resulted in knowledge about the conditions which result in certain types of drawing and has been useful in explaining oddities in children's drawings. The research is more experimental but is also less developmental, taking the child out of context into an isolated drawing situation in order to analyse the activity. The developmental perspective is limited to the differences between what young children are capable of in the experimental situation, and how they deal with the problems they encounter there, compared with the older children.

Drawing can also be seen as using symbols or schemas that are acquired from the relevant culture. Symbol use begins with simple shapes and progresses into more elaborate schemas. Analogical mapping of symbols is replaced by digital mapping and early flexibility is lost as the child masters the symbol system of the culture. The

comparison of drawing with language is useful in highlighting the dangers of treating analogic communication as digital communication, with implications for the Goodenough DAM test, and also for pointing out the problems inherent in the projective use of drawings as a form of analogic communication.

Human figure drawing development

There is a wealth of research that has been undertaken on many aspects of children's drawing, from geometric shapes to the metaphoric use of line, which is reviewed elsewhere (e.g. Cox, 1992; Golomb, 1992; Thomas & Silk, 1990) and is beyond the scope of this thesis. Research and literature on human figure drawings in particular is focused on both as a specific example of children's drawings, and due to the obvious relevance for the DAP test and its interpretation.

An understanding of the normal development of the HFD can shed light on the processes and the problems that are involved in constructing a recognisable representation. The tadpole figure is the earliest form of the human figure to appear in drawings, developing from the circle used to signal 'thingness' (Arnheim, 1974). The figure involves a head contour with facial features and legs attached. It has been described as an animate rather than especially human figure (Golomb, 1992) which is used to represent animals as well as humans. The typical age of a tadpole drawer is around 4 years of age (Ames, 1945).

The puzzle of the tadpole figure is its lack of torso and arms attached to the head. Kellogg (1979) thought that the child was more concerned with the aesthetic appeal of the drawing than making it visually realistic. The sun-schema and mandalas in her scribble theory are radially symmetrical and are also suitable forms for the human figure. Tadpole drawers therefore are more interested in making the drawing achieve the balance they practised whilst scribbling, than the need for a torso. The problem with this theory is that the figure often lacks arms and does not display radial symmetry. Also, the sun-schema and mandalas only occur in a small proportion of children's drawings and appear at about the same time as the tadpole, casting

doubt on the hypothesis that they are precursors to this figure (Cox, 1992).

Arnheim (1974) believed that the circle of the tadpole figure was "an undifferentiated representation of head and trunk" (p. 199). The problem with this theory is that few children claim that their tadpole figure has a body when naming the parts they have drawn. The area at the lower end of the circle is usually labelled the 'head' and not the 'body' (Windybank, cited in Cox, 1993). Investigations by Freeman (1980) also showed that arms were attached not to the body segment as Arnheim would predict, but to the larger circle of a pre-drawn figure. This 'body proportion effect' Freeman classed as a production error that was a systematically biased response to the cue of relative size, rather than location.

Luquet (1927) and Piaget and Inhelder (1956) believed that the child has problems extracting the salient features from an object when creating their internal model which is then reflected in the drawing. Children were thought to first obtain the gross distinctions before analysing in terms of finer details. The head and legs were thought to be important due to the head's relationship to perception and communication and the legs' salience with regard to the height of the figure and its uprightness, though it is less clear why the arms should be less important (Cox, 1993).

Freeman (1980) discusses the phenomenon of the tadpole figure, not in the sense of an under-developed internal model, but in terms of the production problems inherent in the drawing process. The deficiencies of the tadpole figure could be seen as a performance problem, both accessing stored representations, and translating the knowledge into a linearly ordered, sequentially drawn, series of parts. The child's limited cognitive abilities have trouble coping with the required program of action, memory and decision making for the planning and monitoring needed to produce the HFD. Freeman proposed that the vertical axis of the human figure (head to legs) is subject to serial position effects, where precedence is given to the first and last items in a list. The child recalls the head first, then end-anchors on the legs which results in the tadpole figure. The problem is that it is unclear why the legs and not the feet are the last item in the

list (Cox, 1993) and also, 70% of 4 year olds have been found to 'return upwards' in a visual inspection of their figure suggesting the figure completion is independent of the top-to-bottom sequence (Golomb, 1992).

Tadpole drawers have been found constrained to draw the tadpole figure and unable to draw the conventional figure, evidence that a faulty internal representation is to blame. Tadpole drawers choose the tadpole form as the 'best' representation of the human figure rather than the transitional or conventional form (Cox & Stone, cited in Cox, 1992; Taylor & Bacharach, 1981) and a manikin task with reduced demands than drawing, also failed to facilitate drawing of a conventional figure (Cox & Parkin, 1986). When a body is dictated to tadpole drawers, they either fail to complete a conventional figure altogether (Cox & Parkin, 1986), or they have been found to locate it in the head segment or between the legs (Cox & Batra, cited in Cox, 1993). However, it is not that the child does not know that humans have these body parts as all can identify their own 'tummies' and can point them out on a pre-drawn figure (Brittain & Chien, 1983). Tadpole drawers have been found able to construct a conventional figure in manikin tasks using rectangular shaped pieces (Bassett, 1976), though these children were older than those who failed using the circular shaped torso of Cox and Parkin. With features drawn on the pieces even more children succeeded in constructing a conventional figure which suggests that the children do have a complete internal model of the human figure in its conventional form. The drawing task contains more constraints since children have been found able to name more parts of the human figure than they draw (Freeman, 1980). In dictation tasks, the children may have failed to locate the torso as a separate entity for reasons other than a faulty internal model. It may be that the children do not understand at this age that the torso is supposed to have its own contour that goes below the head segment. Equally, they may not have yet gained the graphic skills necessary to draw the body.

Copying an adult model can help the tadpole drawer produce a conventional form, but only if the figure is broken down into its constituent parts for the child and they are allowed to practise several times (Cox, 1993). Though this is not directly comparable with

spontaneous drawing, the conventional form was found in these children's spontaneous productions two days later. Children revealed more knowledge of the human figure via a manikin construction task at all ages compared to free drawing (Celotta, 1973). The children knew more than they could draw, especially at young ages but cueing the children to include anything they might have left out of their HFD had no effect on the finished products. The evidence tends to suggest, therefore, that it is not the internal representation that is at fault as Piaget and Inhelder believed, but the accessing of that representation and the lack of available graphic skills needed to produce the figure.

A second form of the tadpole figure has been noted by researchers (Luquet, 1913; Arnheim, 1974) that consists of longer legs than the traditional tadpole figure with the arms of the figure attached to these legs. The torso of the figure is believed to be located between the legs of the figure. The typical age of this transitional drawer is older than the traditional tadpole drawer, a result found cross-sectionally and longitudinally (Cox & Parkin, 1986). The transitional drawer prefers the conventional form as do the conventional drawers themselves (Cox & Stone, cited in Cox, 1992). They aspired to the conventional form and recognised it as a better form though they could not achieve it in their own drawings.

Usually by the age of 5 or 6 years, most children produce a torso in their drawing of the human figure, the figure then becoming known as the conventional figure (Cox & Parkin, 1986). Placing a horizontal line between the legs of the transitional figure produces a conventional form that differentiates figure and ground for the torso. The large number of circular body segments actually produced casts doubt on the idea that this is how development proceeds. Cox and Parkin (1986) found no clear evidence from longitudinal data that the conventional figure had been adapted from the tadpole form. The advance away from the tadpole figure is associated with a shift towards re-ordering the sequence of the drawing, away from the legs-arms order bias of the tadpole drawers towards the adult strategy of head-torso-arms-legs. The addition of the torso also leads to the inclusion of greater detail.

The frontal aspect of the HFD is the most characteristic and was labelled the canonical representation by Freeman (1980). This form

shows all the relevant criteria for the human figure and adults also draw this view, though they can normally maintain the same viewpoint and perspective throughout the figure, unlike the child's, where, for example, the face may be in front view but the feet may face to the sides. The canonical figure is suitable for giving each body part its own space without the need for partial occlusion, a preference noted in children (Goodnow, 1977).

The development of the figure construction is seen as an addition process, where self contained units are attached or aligned, resulting in the segmented human figure. The contour of the body can be seen as signifying the volume of the figure in three dimensions a well as the outer boundary in two dimensions (Arnheim, 1974). At about 6 or 7 years the child may attempt a single continuous contour that necessitates planning to guide the outline. Creating a figure with the use of the single outline was termed 'threading' by Goodnow (1977) and can sometimes make a drawing seem more immature than the child artist really is, simply due to the difficult nature of planning and constructing a figure in this way. Contouring does not appear due to increased interest in depicting action and movement in their HFDs, since drawing a figure running did not result in more single contour figures being drawn (Cox, 1993). Striving for realism leads to the abandonment of the segmented approach (Fenson, 1985). The shift from segmented to contoured figures occurs as the older child begins to construct not only the constituent parts but also the relationships between them (Goodnow, 1977). The contouring is rarely of the whole body; it usually occurs as either contouring of the arms and upper body segment along with legs and lower body segment, resulting in a figure with a waistline dividing the top half from the bottom half of the figure. Sometimes the legs and foot and/or arm and hand are contoured; sometimes the neck is contoured in with the upper torso. Single contouring usually occurs with the production of clothing and more realistic drawing of the limbs supporting Fenson's claim that children strive for realism.

The development from intellectual to visual realism leads to the ability of children to produce a partial occlusion in their HFD as they attempt to show the figure from one single viewpoint. The appearance

of this graphic technique tends to occur at about the same time as the profile drawing, which inherently demands the use of partial occlusion. There has been evidence, however, for the ability to use this technique much earlier with children as young as 4 able to show partial occlusion in their drawings in certain situations (Cox, 1981).

Profile figures are included when specific themes are used and normally occur in children's drawings at about the same time as single contouring. In the standard draw-a-person task, the occurrence of profile figures is normally quite low (c. 20%), even amongst 9-10 year olds. If asked to draw a figure in action, this figure increases (c. 80%). A significant number of children over 7 years old can alter their drawing to reflect the altered orientation of a model (canonical, profile and back). To represent the profile view, the children drew one leg, one arm, feet pointing to one side and parts of the facial features. Younger children preferred to use the canonical representation whichever orientation was required (Cox & Moore, 1994).

The Draw-A-Man Test

Early descriptive studies identified cognitive development as the primary influence on the nature and content of children's drawing (Goodenough, 1926). Drawing processes were urged to be examined as cognitive operations (Oakley, 1930). Goodenough (1926) assumed that intellectual development was the main determining factor in the quality of a child's drawings and thus drawing could be used to assess the child's conceptual development and his/her intellectual maturity. This view gave a new direction to the work on children's drawings as part of the psychometric study of intelligence.

Children add more and more body parts to the HFD up to the age of about 12 years of age (Goodenough, 1926; Harris, 1963; Mortensen, 1991) and the Draw-A-Man test (Goodenough, 1926) took advantage of this normal HFD development. Children were credited for the features and correct proportions that they included on their drawing of a man, this being expected to improve with age. Drawing development was seen as an index of conceptual development and intellectual maturity, with the HFD reflecting the developing internal representation of a person. An evaluation of this early scale acknowledged its advantages

for reliability and ease of use, but also saw problems in diagnosis due to subjective scoring and individual variability over time (McCarthy, 1944). The influence of preceding activities and affective states was also seen as problematic for the DAM test.

Subsequent research by Harris (1963) developed and extended the test to include a drawing of a woman and self as well as updating the norms upon which the test is based. GH scores correlated well with IQ scores and achievement tests for kindergarteners (Vane & Kessler, 1964) but were lower for child and adolescent psychiatric inpatients (Aikman, Belter & Finch, 1992). Aikman et al. cautioned against substituting other well established IQ and achievement tests for the DAM test. Good retest reliability was found over two weeks (Brown, 1977) but poorer reliability was found with longer time intervals, especially when the retest sessions were on an annual basis (Vane & Kessler, 1964). Requesting 'good' drawings obtained higher scores on the GH scale than spontaneous pictures. Spontaneous drawings commonly involved action and were more experimental, whereas the requested ones were static and therefore earned higher scores (Goodnow, Wilkins & Dawes, 1986).

Koppitz (1968) developed a drawing test as a measurement of intellectual maturity. From Draw-A-Person test normative data, features of a child's drawing found to be either expected (i.e. on >85% of drawings) or exceptional (i.e. on <15% of drawings) were used to score a drawing to determine a broad level of intellectual functioning. Recent work confirmed and extended the 30 developmental items to younger children (Groves & Fried, 1991). Increased detail at every age in this later sample, compared with Koppitz, was attributed to widely reported correlations between intelligence and detail, since the later sample was 1.5 standard deviations above average for intelligence.

No differences have been found between the GH point scale and the Koppitz system in terms of inter-scorer reliability and concurrent validity, but the Koppitz method has met with scepticism due to the rather broad categories into which the children's scores fall and the lack of specific IQ score (Gayton et al., 1974). The Koppitz developmental items are easier to use than the Goodenough-Harris scale that is also more time consuming to score but gives a much more specific final IQ

score. Abell, Von Briesen and Watz (1996) found that both the Koppitz and GH scale drawing scores correlated significantly with WISC and Stanford-Binet IQ results but the longer and more detailed GH scale had significantly higher correlations with performance IQ from the WISC than did the Koppitz system. The GH scale also performed comparably with the person drawing obtained during the House-Tree-Person technique (Buck, 1948), but both methods underestimated IQs (Abell, Heiberger & Johnson, 1994).

The most recent version of the DAM test is the 'Draw-A-Person: Quantitative Scoring System' (DAP:QSS, Naglieri, McNeish & Bardos, 1988) though this is not published in the UK and has yet to gain widespread use. The DAP:QSS is a revision and update of the classic DAP technique and is similar in style to the Goodenough-Harris (1963) method. The norms are updated and based on a stratified sample from 1980 US Census Bureau statistics. Reliability coefficients are generally high and concurrent validity with the GH system is also high (Kamphaus & Pleiss, 1991). Concurrent validity with IQ tests, however, is only moderate and comparisons with other screeners, such as short form IQ tests show the DAP:QSS to be weak. The theoretical foundation of the test also lacks clarity and it is difficult to understand the nature of the construct being measured, a problem for all these tests.

There are several problems with the use of the DAM test. A cyclical pattern of the individual variability in HFD development was found (Rubin, Schacter & Ragins, 1983), with variability increasing at ages 4 and 8 and decreasing variability at ages 5 and 10. Freeman (1975, p. 19) wrote that "individuals do not develop at an even rate - they have spurts and lags and no single drawing can give evidence of that". A child will also include features on request that they may not spontaneously or necessarily have included in their drawing (Golomb, 1973, cited in Cox, 1992). It has been possible to analyse skills such as cognitive planning, analysis and synthesis in handicapped children using a computer to aid the drawing process which it was not possible to do with pencil and paper (Olsen, 1992). This method made it possible to measure cognitive abilities 'under the floor' of what is normally possible using drawings and questions the validity of drawings as a measure of intelligence. Previous art experience has also been

shown to have a significant effect on drawing performance (Brewer, 1995).

The relevant details of the canonical figure also change with time and are different among cultures (Wilson & Ligtvoet, 1992). Harris re-standardised the data collected by Goodenough and noted marked differences in the drawing of the torso in proportion, depiction of hair and finger detail, though no diffcrences were found for major details such as presence of head and legs, eyes and mouth. When Mortensen compared her data with the Goodenough and Harris data, she found similarities mostly with the Harris data, which made sense due to the relative proximity of the time periods. Pfeffer and Olowu (1986) found significant differences for overall shape, inclusion and position of body parts, and clothing/details between low and middle income children's HFDs in Nigeria. The middle class children may have had more access to pictorial materials and better art facilities in school, as well as their parents' help. Low income children tended to have a more traditional style of drawing which is not necessarily worse. The skills which are assessed in the DAM test may be more relevant to Western style education which gave the middle income children an advantage.

Children's own awareness of the development in their drawings is limited at younger ages (Tryphon & Montangero, 1992). Trautner, Lohaus, Sahm and Helbing (1989) found that young children were as accurate in rank ordering (for age) older children's drawings as they were with the drawings of children of their own age. Even a 6 year old child can order a series of pictures in terms of the artists' ages, but when asked to draw a HFD as they would have done at an earlier age and explain the changes which have occurred in-between, younger children fail to understand the evolution of the HFD and were found to have a simple model of 'smaller equals younger'. An additive model of the drawing process was found in slightly older children before a truly qualitative stage model came into play at about 10 years of age when the children understood that the origins of the drawing process can be found in scribbles and tadpoles.

Children's exposure to older children's and adults' work is assumed to influence developmental change. Children are able to consistently and accurately discriminate older and younger pictures

from the age of 5 years (Goodnow et al., 1986). Young children are able to recognise the older children's drawings and judge them as better than theirs (Fayol, Barrouillet & Chevrot, 1995). This difference decreased over age and progress was seen with age through the ability to discriminate developmentally close drawings. This supports Freeman's opinion that the internal model of the child is the same as adults' but young children have problems planning, executing and managing the drawing task. Fayol et al. found there was no correlation between judgement and production of drawings at all ages, which would have supported Luquet's theory that drawings correspond to the child's internal model of the object.

The difference between children's judgement of a drawing as the best representation of an object and their own productions of the object cautions against inferring about internal representations from their drawings (Kosslyn, Heldmeyer & Locklear, 1977). Children did not choose a diagrammatic (intellectually realistic) drawing of an object as its best representation, but most produced this type themselves in their own drawings. Similar results have also been shown with HFDs (Taylor & Bacharach, 1981) where the choice of drawing made by children does not always correspond to their own productions.

Children's perception of their own competence in drawing has been found to be related to higher levels of (rated) realism and originality in drawings and greater perceived competence in domains with visual-spatial components, such as maths and acting (Flannery & Watson, 1991). Disabled adults described their drawings at a level similar to their peers, but the visual impact of the drawings was more similar to children's drawings (Wright & Ashman, 1991). This cautions against assuming that intention matches the finished product.

Adapting the HFD

Children's representational drawings are commonly viewed as stereotyped and 'formula' driven, and therefore resistant to change and not specific to a given model. Luquet and Piaget argued that the children's mental representation of a class of objects would be produced when they drew and therefore specific information about particular members of the object class would not be incorporated into the drawing.

Karmiloff-Smith (1990) believed that the child's mastery in a given behaviour is learned or acquired as an entire procedure which the child cannot either analyse or reflect upon. Changes can only be made at the end of the behavioural sequence, if at all. Evidence was found in drawings of children asked to create a man, house and animal 'that doesn't exist'. Examination of the modifications to the drawings showed that the older children deleted features in the middle of the drawing procedure whereas younger children deleted at the end of the sequence. Younger children who added elements, did so at the end not the middle of the sequence. Two sequential constraints were proposed, at the procedural level in terms of the sequential instructions used to produce the drawings and at first level of redescription where younger children have the redescribed components of the procedure available to them but the order in which they are operated on is constrained by the sequential order. Later redescriptions are not constrained, but change must occur in the internal model or sequential representations in order to be able to develop and alter the drawings.

Van Sommers (1984) saw resistance to change not in the internal mental representation of the object but in the visual goal which was used to guide a drawing. However, once a goal has been set for a particular drawing, it interferes with subsequent drawings and thus limits the changes that could be made to them. The influence of the drawing procedure itself rather than the underlying mental description is also thought to both guide and limit the representation (Thomas, 1995). Early in drawing research, the idea was presented that the dominating factor in the drawing process is the actual figure developing on the paper, which becomes mechanical if repeated (Oakley, 1930). Jones (1972, cited in Van Sommers, 1984) indicated that the drawing process itself could restrict graphic development, an idea similar to that of Freeman (1975) who believed that rules guide every decision point in the drawing process, though the child may not be aware of those rules. The use of schemas may constrain the picture because only a small range of elements is available to use and the child has a limited ability to adapt them (Thomas, 1995).

Evidence from a variety of studies, however, casts doubt on the belief in routine, formula-driven, stereotyped representations that are

unable to be modified. The profile figure, for example, is just one alteration that a child can make to his/her figure to portray action or movement. Movement generally occurs in a left to right direction in children's drawings which may be the result of not wishing to cover the work already done whilst drawing, a reflection of the influence of the drawing process itsclf, rather than underlying mental representations. Various tactics such as flowing hair and clothes and movement lines are also used to depict the desired movement. Techniques such as widening the angle of the legs was found on even the youngest children's figures when the children were required to show walking and running, though the older children altered the arms as well as the torso (Goodnow, 1978). In showing a figure stooping to pick up a ball, the younger children were unable to alter anything except the peripheral parts of the figure; the children over 8 years of age altered the core structure and bent the figure at the waist. This relates to Arnheim's (1974) view about development proceeding according to the law of differentiation whereby development moves from simple to complex forms. Children under 6 years, however, have been found capable of conveying the required information to differentiate between standing and walking figures (Smith, 1993). Walking was shown by turning the head or torso sideways, joining the legs at an acute angle and with a bend in them, having a greater distance between the ends of the legs and drawing the feet pointing the same way. Standing was usually portrayed by drawing the legs vertically. The cues increased in use once they had appeared, and increased expertise was shown by the addition and integration of new differentiating features rather than the replacement of one with another. The order in which drawings were asked for had no effect on the pictures.

Other evidence has shown that by the age of 7, children are able competently to adapt their figure to account for the differentiation in orientation and activity (standing facing, standing profile and running figures) (Cox & Lambon Ralph, 1996). The presence of a model had a limited effect except in the running condition where it resulted in more bent limbs, transparencies and partial occlusions. The model had a negligible effect in the youngest children's figures, however, supporting the view that the younger children use their internal model of a figure

much more and are less likely to take account of an external model, producing a stereotyped representation based on their mental representation rather than the specific item in view. These children also produced similar figures across all three conditions, only using one feature adaptation, if at all, compared with the older children who could use more than one feature for their adaptations.

Given the evidence that even young children have been found able to represent a difference in the activity of the HFD, it seems that children are quite able to adapt their human figures. The evidence questions the reliance on behavioural routines as the flexibility of the process mid-sequence must be recognised and not seen to be as rigid as Karmiloff-Smith thought. The evidence also questions reliance on underlying mental representations in all but the very youngest children, though the view that the internal concepts are perceptual rather than abstract (Arnheim, 1974) may be more successful in explaining how the children can alter them to account for different perspectives. Results from contrasting figures tasks (Smith, 1993) where differences between two consecutive drawings were seen, appears to contradict Van Sommers' theory. If the children were told they were to draw two pictures, they may have set up two goals from the beginning, hence being able to draw two different pictures. This would require a great deal of planning though, which is unlikely, especially in the younger children. If they only planned the second after the first, there is little evidence for the constraining effect of the first drawing.

Inflexibility may be due to habit rather than limits in cognitive capabilities. The way children draw an object depends on the features they assume they should be representing and the precedence they choose to give to one type of information over another (Sitton & Light, 1992). In a communicative context, children aged 7 to 8 years can change their drawing whereas they fail to at age 5 and 6 years, unless the other child is present too. When new information is presented which is related to the object itself and not the method of depicting it, young children stick with their established strategy and adapt by adding detail (Van Sommers, 1984). This should not necessarily be viewed in a negative light if they have found a successful graphic solution and choose to stick with it (Arnheim, 1974). Children failing to

adopt the adult style should not be viewed as wrong. Adults are able to pick up on the task demands and alter their drawings, so it may not be a failure on the part of the children to draw the changes, but to see the changes as necessary (Van Sommers, 1984).

Summary The development of human figure drawing can be seen as a process which is affected by a multitude of factors. The lack of body and arms on the tadpole figure is probably not due to a problem with the child's internal model and it seems more likely that the tadpole figure results from the production problems inherent in the drawing task for the young child.

A transitional form occurs prior to the canonical view though it is less clear exactly how the transitional form acts as intermediary between the tadpole and subsequent forms. With the advent of the canonical view, the child focuses on the addition of parts to the drawing and the relationship between these parts is explored in the technique known as 'threading'. The profile view occurs when the child becomes concerned with visual realism and also becomes interested in portraying action and movement. The peripheral parts of the figure are altered before the core sections in portraying action. Drawing development occurs according to the law of differentiation (Arnheim, 1974) with movement from simple to complex forms.

The Draw-A-Man test and the measurement of intelligence by drawings took advantage of the fact that normal HFD development follows a path of increasing detail and better proportions between parts. Various tests have been developed which have shown moderate reliability. Problems have been found with the validity of these tests and criticism arises from the variability in children's drawings both in time and between cultures. Children's own awareness of their drawing development is limited and exposure to older children's and adults' work is assumed to influence developmental change. The contrast between children's judgement of a drawing and their own productions of the object cautions against inferring about internal representations from their drawings.

Limitations on the adaptations possible in HFDs may be due to the mental representation, behavioural routines, visual goals or

schemas. Evidence from a variety of studies, however, casts doubt on the belief in routine, formula-driven, stereotyped representations that are unable to be modified. Inflexibility may be due to habit rather than limits in cognitive capabilities and the drawing may be influenced by the choice the child makes to give precedence to one type of information over another when deciding what information to represent. Sticking with a successful graphic solution should not be viewed negatively, as the child may simply fail to see that the changes are necessary.

Interpretation of HFDs

An interest in children's drawings as a reflection of emotional health and personality was born out of the work involving the DAM test and drawings as a measure of intellectual maturity. Goodenough (1926) noticed that there were some aspects of the children's HFD that were unrelated to their intellectual maturity and seemed to have more to do with their personality. Drawings achieved interpretative and diagnostic status and were interpreted, not for cognitive functioning, but for emotional health. Early work on the interpretation of children's drawings used spontaneous productions in many media. Alschuler and Hatwick (1947) evaluated nursery children's paintings, believing them to have significant content long before they were recognisable in any realistic sense. The existence of repeated, persistent relationships between the children's problems and their artwork was seen to justify their thesis, though their work was criticised for its lack of statistical analysis (Goodenough & Harris, 1950).

In the 1930s and 40s the HFD in the form of the Draw-A-Person (DAP) test became the most popular way to employ drawings as a projective technique which could give insight into an individual's personality or mental health. School psychologists have been found to show a high use of projective drawings (Prout, 1983; Goh & Fuller, 1983, cited in Joiner, Schmidt & Barnett, 1996). Many different drawing systems have been devised and the HFD is incorporated into many of the other systems but is widely used in its lone form. A recent survey of the psychological test usage patterns (Lubin, Larsen, Matarazzo & Seever, 1985) reported that projective techniques have not declined in use and the DAP test was among the top 10 test

instruments used in the settings surveyed (Lubin, Larsen & Matarazzo, 1984). In a 1982 survey, the DAP test was 8th in frequency of usage with 66% of those surveyed mentioning its use. This compares to the MMPI that came first in frequency and 82% mentioning use. The HTP test came in 10th with 67%.

Body Image Hypothesis

Machover (1949) was the first researcher in this area to formally state the body image hypothesis as a connection between the drawers' body image and their HFD. Koppitz (1968) preferred to believe that the child would draw whoever was important for him/her and considered the HFD a representation of the drawers' self-concept. This was more in line with the views of Hammer (1958) who also believed that children would draw their real or ideal self or some other important person. DiLeo (1973) believed that when children were well adjusted and free from anxiety, their intellect would be free and their behaviour exteriorised. The drawings of such children would not be of their self but their concept of humankind. However, in drawings of children with emotional problems certain symbolic elements would appear and their drawings would be more reflective of their self-concept.

Swensen (1957, 1968) and Mortensen (1991) have reviewed the evidence on either side of this debate. They cite positive findings for the body image hypothesis which came from investigations correlating the drawer's body type and HFD type (Berman & Laffal, 1953), matching photographs with HFDs (Apfeldorf & Smith, 1966) and size of drawer and HFD size (Kotkov & Goodman, 1953; Craddick, 1963). No difference was found, however, between pregnant and non-pregnant women's drawings of themselves (Tolor & Digrazia, 1977, cited in Cummings, 1986) which may reflect a failure to represent their body image. Some significant differences have been found between the HFDs of children with and without physical disabilities (Wysocki & Whitney, 1965). The samples were only group matched for mean age and IQ, however, and it was not clear how the 'area of insult' was depicted. Silverstein and Robinson (1956) collected drawings from chronic cases of polio with residual paralysis in the legs. The children were seen to

represent their disability in their drawings but the judges may have been biased in their perception of the pictures.

Support for the hypothesis that HFDs represent the drawer's self concept came from research showing greater similarity between the HFD and 'Actual Self' ratings of institutionalised schizophrenic women (Kamano, 1960). Bodwin and Bruch (1960) also found a positive correlation between DAP score and self-concept as rated by interview. Van Dyne and Carskadon (1978) found significant positive correlations using semantic differential ratings between same-sex figure ratings and ratings of real and ideal self. Also, though logically given the first result, a negative correlation was found between ratings of same-sex figure and least-liked self. Significant differences were found between the HFDs of severe and mild psoriasis sufferers for percentages of undressed figures and omissions of exposed body parts (Leichtman, Burnett & Robinson, 1981).

Bennett (1966), however, failed to find a relationship between DAP scores and self-concept as measured by Q-SORT in 10 year old children. This may reflect a more complicated relationship between body image and HFD in children, or an underdeveloped self-concept in the younger children. Children are also future oriented, and tend to draw HFDs older than their own age (McHugh, 1966), which may also support the idea that the HFD is not necessarily their own body image. Children's awareness of their own bodily attributes compared with those of the other sex did not show up in the accuracy, articulation or quality of drawings of themselves and the opposite sex (Gellert, 1968). This contradicted previous results (e.g. Harris, 1963; Richey, 1965, cited in Mortensen, 1991), but the request for the character to be dressed in a bathing suit may have affected the comparability of the data. No conclusive evidence was found for the hypothesis that self drawings in normal girls were indicative of self-concepts (Fu, 1981) supporting the assertion that normal children tend not to be emotionally involved in their HFD production (DiLeo, 1973).

Other studies attempting to relate HFDs to body image or self-concept have used a measure of self-esteem. Delatte and Hendrickson (1982) found a significant linear relationship between self-esteem and both height and width of the HFD in adolescent participants. Dalby and

Vale (1977), however, found that self-esteem was not related to the size of the drawn figure and Prytula and Thompson (1973) found no support for the body image hypothesis as related to self-esteem. The indirect measure used in these studies may however lessen any relationship that occurred between self-concept and the HFD. Content and global characteristics of the HFD differentiated five different self-esteem categories better than the formal characteristics (Coopersmith, Sakai, Beardslee & Coopersmith, 1976). This research suggested that children's HFDs are more directly related to the child's behavioural manifestations of self-esteem than the subjective estimate of self-esteem. Swensen (1968) concluded from his review that scores on various aspects of the DAP were related to some measures reflecting self image. It is important to note that most positive results came from using adults as participants; most negative results came from using children as participants (Mortensen, 1991).

Klopfer and Taulbee (1976) criticised the use of the DAP as a measure of body image because "the criteria are not clear, and some of the theoretical underpinnings seem to be rather off the mark" (p. 560). The choice of which figure was drawn first was seen as more to do with cultural learning than innate self-concept, and also gender identity and sex choice of figure drawn first were found to have no clear relationship, though later research showed drawing the same-sex in the HFD was found to be the norm except for adolescent girls and young boys (Dickson, Saylor & Finch, 1990). Klopfer and Taulbee concluded with the statement that "many of the hypotheses formed by authors like Machover are at a level clearly not related to either conscious self-concept or behaviour" (p. 561). This argument could be counteracted by the fact that there may be a part of the self which is non-phenomenal and unconscious (Jones, 1992). Also, body image disturbances may not be represented isomorphically, but revealed in other ways (Cummings, 1986). It is still unclear, however, by what mechanism the 'self' becomes projected into the picture.

Sign-Interpretation of HFDs

A pioneer in the use of drawings as a projective technique was Karen Machover who, in her book, 'Personality Projection in the Drawing of

the Human Figure' (1949), discussed the use of HFDs as a projective technique and a tool for personality assessment, although this text was aimed at interpretation of adult patients' drawings. Individual 'signs' in a picture were each given specific interpretations, and Machover attached meaning to most aspects of a drawing according to certain principles. The head, for instance, was seen as symbolic of intellectual power and social dominance and the centre for the control of body impulses, whereas the face reflected the interpersonal relationships of the person. There was a great deal of sexual symbolism in Machover's theory where features such as the mouth represented erotic satisfaction and hair was a sign of virility and all clothing was given libidinal significance. Toes were seen as a sign of aggressiveness that was almost pathological and transparencies revealing anatomy indications were seen by Machover as indicative of schizophrenia/mania. Structural and formal aspects of a drawing were also analysed, for example extreme symmetry was seen to be symbolic of rigidity in the person's personality. The presence of a midline, drawn on the figure commonly in the form of a line of buttons, was considered to reflect somatic preoccupation, body inferiority, emotional immaturity and mother dependence.

Machover (1953) extended her original work to children's drawings, using 1000 drawings from public and private schools, kindergarten to 6th grade. She looked at cultural as well as age and school differences, and compared black and white children's drawings and Jewish children's drawings. There was a psychodynamic orientation again to the interpretations and analysis was mainly in terms of the latency period, self concept and the conflicts involved in sex differences. In this text, Machover ignored media influences on children's drawings and preferred to interpret features such as the drawing of gangsters, for example, as evidence for the emergence from the latency period and indicative of problems with mother figures. The oedipal struggle was also used to explain many features of the boys' drawings. This work suffers from a severe shortage of corroborating evidence, and the features that have remained popular, used for both adult and children's drawings are those from the original 1949 book.

The HFD was also involved in Buck's House-Tree-Person (HTP) technique (1948). This involved the patient making a freehand drawing of a house, tree and person. The drawings could be analysed quantitatively for the purpose of acquiring an intellectual level and qualitatively for personality assessment. A post drawing interrogation (PDI) afforded the subject the opportunity to define, describe and interpret the drawn objects. The drawing was assessed in terms of the details produced, the proportion and perspective, time taken and line quality as well as the subject's attitude, comments, drive and criticality. Similar interpretations to those of Machover are made: for example, the hands of the person representing the means with which the person took defensive or offensive action towards the environment or self. Sexual symbolism was as inherent in Buck's work as it was in Machover's, for instance the drawing of a necktie was seen as a phallic substitute, and over-emphasised erogenous zones on the HFD revealed "psychosexual deviations, fixations and immaturities" (p. 370). Buck relied more on the PDI for the qualitative analysis, allowing for the subject to justify his/her drawing but the subjective nature of the assessment left most of the interpretation to the examiner. Variations of the prototypical house and tree were thought likely to be indicative of personality variables (Soutter, 1994). Validity was found for variables of aggression, impulsivity, immaturity, egocentricity and dependency, using the HTP with deaf adults, by comparing psychologists' ratings of the drawings with trained counsellors' clinical observations of the participants (Oullette, 1988).

The projective use of figure drawings has also been found in the use of mother and child drawings (Gillespie, 1994) assuming that the figures carry projections of the internalised self and other. The interpretations that are made from these pictures are based on intuitive impressionism with ideas such as men's troubled attempts with the tasks being the result of their difficulty internalising women as positive objects. Size displayed in the picture is seen as representing the psychological size of the individual in human relationships. This attempt fails to take into account the normal course of development of children's pictures, for instance a description of 'twin' figures of mother and child by a 5 year old is interpreted as reflecting a lack of

differentiation of the child from the mother, and the author fails to recognise the highly stereotyped and symbolic nature of children's drawings.

The sign-interpretation approach to the DAP was twice reviewed by Swensen (1957, 1968). In the first review, Swensen concluded that the evidence did not support but mostly contradicted Machover's hypotheses about the meaning of content and structural variables in the HFD. For example, no differences were found between schizophrenics and normal samples for head size, presence or absence of facial features, depiction of lips, nose, toes, anatomy indications, portrayal of action or scattering of parts (Holzberg & Wexler, 1950). No differences were found between patients who improved during psychotherapy and those unimproved for depiction of the mouth, ear, hair, arms, hands and fingers, placement on the page, stance and shading (Gutman, 1952). Only a limited number of HFDs from paranoid schizophrenics were found to satisfy Machover's criteria for paranoid trends, but a lack of control drawings made this study problematic (Fisher & Fisher, 1950). Silverstein and Robinson (1956) also failed to find significant differences between the drawings of the paralysed and normal children using various signs including Machover's. Though Swensen concluded that "no considerable empirical support for Machover's hypothesis exists at the present time" (1957, p. 460) he also noted that few studies had explicitly tested Machover's hypotheses.

Other reviews agreed with Swensen's conclusions about the lack of supporting evidence for Machover's hypotheses regarding content and structural variables. These also noted the lack of well designed studies from which any conclusions about the usefulness of the HFD test could be drawn (Roback, 1968) and the need for more consistency in methods (Jones & Thomas, 1961). A review of Machover and Hammer's principles for the DAP test with adults, between 1967 and 1982, revealed that figure drawings are not meaningless, but establishing what they do mean with precision and predictability is difficult because of the inadequacies of the research (Kahill, 1984).

Swensen's second review (1968) found that more positive findings had been determined by more consistent testing of the hypotheses.

Global judgements of drawings were found to be more reliable and related to personality or behaviour ratings (Lewinsohn, 1965; Kahn & Jones, 1965). Global ratings may be better than individual signs because the reliability of a sign was found to be a function of the amount of behaviour included in the sign (Maloney & Glasser, 1982). Mixed evidence was again reported, however, for the content and structural variables. For example, size of drawing was found to be related to manipulated self-esteem (Gray & Pepitone, 1964), shyness (Koppitz, 1966) and depression (Lewinsohn, 1964) but unrelated to self-esteem measured by Q-SORT (Bennett, 1966) and diagnosis of character disorder (Exner, 1962). For every study showing positive results for the placement of the figure on the page (e.g. Handler & Reyher, 1964) there was a study showing no relationship between placement on the page and personality (e.g. Exner, 1962). Only Machover's hypothesis that neurotics draw small slight figures was supported in a comparison of neurotic traits and conduct disturbances (McHugh, 1966). An assessment of Machover's psychopathic indicators using MMPI classified psychopathic prisoners and college student controls failed to find differences in height or placement. One expected indicator (hand in pocket) failed to appear at all in the psychopaths' drawings and more of the controls drew shading which was against Machover's theory (Craddick, 1962).

Drawings made by kindergarten children, who were rated as poorly adjusted, showed significantly more 'grotesque' figures, 'no body', 'no mouth' and 'no arms' (Vane & Eisen, 1962). Significant differences were found between the percentages of poor adjusters and matched controls showing one or more of the above signs. However, the 'grotesque figure' indicator is not elaborated upon and the subjective nature of this indicator therefore casts doubt upon its validity. Also, at the ages studied, it would not be so unusual to find children still drawing tadpole figures (no body and no arms), especially if those 'poorly adjusted' were delayed in mental age. Poor achieving boys are found to deviate most on Machover's items (Lourenso, Greenberg & Dunn, 1965).

Valid findings were found for the variable 'stance' (e.g. Kahn & Jones, 1965; Handler & Reyher, 1966) which was supposed to reflect

the person's emotional stability. The most reliable findings were for 'distortion' in a drawing that was consistently found to be a sign of severe emotional disruption (Hiler & Nesvig, 1965; Koppitz, 1966; Handler & Reyher, 1964). However, there is insufficient evidence to rule out the possibility that such global ratings as 'distortion' may only reflect the artistic quality of the drawing.

Swensen noted that the validity of a sign was related to its reliability. Retest reliability over one week was assessed by Hammer and Kaplan (1966). For omission of body parts, only fingers were found to be reliable. For the drawing of a head only, this was reliable but was related to age and sex. It is unclear, however, whether this refers to tadpole drawers or children who draw a bust as a human figure. Given the young age range of the sample, it was probably the latter. For placement on the page, no particular section of the page was reliable, casting doubt on the validity of Machover's view that orientation on the page was indicative of social orientation and mood. Buttons were reliable only on drawings of male figures that may be reflective of the cultural influences of men's fashion. Given the variable reliability of the features in a drawing, the authors suggested caution when using HFDs to apply interpretative significance and Swensen believed the unreliable variables were responsible for producing the contradictory findings he reviewed.

HFDs were found to be more affected by withdrawal features than those of aggression in a delinquent population (Daum, 1983). Features that differentiated significantly were: squared shoulders for aggressive delinquents, omission of facial features, omission of arms, and dim facial features for withdrawn delinquents. The features considered collectively had more diagnostic power. Size, detail and line heaviness were not validated against contemporary depression and anxiety scales, though the three indicators could be reliably assessed (Joiner et al., 1996). The authors concluded that drawings are not useful measurement devices but this may be a little too hasty given the fact that the pictures were obtained using instructions other than the standard DAP test and only three 'signs' were assessed. Using the DAP test in children, however, no significant relationship was found between intensity of line, vertical placement and depression (Gordon, Lefkowitz

& Tesiny, 1980). A significant negative relationship was obtained between size of figure drawn and teacher-rated depression in children (Gordon et al., 1980), though depression measured with the MMPI in adult hospital patients and employees did not show a relationship with figure size (Holmes & Wiederholt, 1982).

The HFD has also been assessed as a predictive indication of suicide potential (Pfeffer & Richman, 1991). Previous research showed HFDs identify suicidal behaviour with sensitivity of 72% and specificity of 70% (Richman & Pfeffer, 1977) and it had previously been noted that suicidal people often draw details of the neck (Machover, 1949; Hammer, 1976). Suicidal indicators (slash lines at wrist, neck, body, limbs; tendency to impulsively cross out drawing before starting over again) were found to show a significant difference between suicidal and non-suicidal children's drawings. The presence of suicidal impulses but not necessarily the level of lethality could be detected in the drawings.

Children can also draw their pain experiences which can be categorised in drawings by content and colour (Unruh, McGrath, Cunningham & Humphreys, 1983). Rae (1991) used the HFD to assess the emotional status of children who are hospitalised. The child's drawings are used as a standard for comparison that allows for evaluation in terms of changes in emotional and developmental functioning over time. Rae cautions against ignoring factors that may affect the drawing production, for example emotional trauma, stress, physical illness, hospitalisation, sickness and physical impairment.

Interpretation of attitudes The sign interpretation approach has been extended to other themes, for instance, children's attitudes towards their families can be assessed using the Draw A Family test (Hulse, 1952), or the Kinetic Family Drawing (KFD, Burns & Kaufman, 1970) and in a similar vein, towards their school (Kinetic School Drawing, Prout & Phillips, 1974). Kinetic drawings were seen as producing much more valid and dynamic material than static pictures. Burns and Kaufman describe various features that can be used to assess children's attitudes towards their families with extensive reliance on case studies for supportive evidence. Compartmentalisation in a drawing shows isolation and underlining the drawing is typical of

children from unstable families. The actions of the figures were indicative of their roles and the child's needs. Rivalry was depicted as a force between two family members such as throwing a ball and a light/heat source in a drawing was a common theme, symptomatic of the child's need for love and affectionate 'warmth'.

The KFD is used in custody situations (Schetky & Benedek, 1992) and children's attitudes towards their parents can be evaluated from their drawings for use in custody disputes, because children are not defensive about their drawing (Lyons, 1993). The process in which the drawing is produced with the parents can become as useful as the finished product itself. Family drawings can also be used as a way to understand children's perception of their stepfamily situation (Berger, 1994) or attitude towards divorce in the family (Cordell & Bergman-Meador, 1991). Using drawings in these situations is thought to enable the children to express in an acceptable way, their experiences and emotional distress, to communicate feelings and ideas and explore and work through problems and anxieties.

Children's art is a novel way to gain information on children's thought processes about other people such as the elderly (Weber, Cooper & Hesser, 1996) and the mentally ill (Poster, Betz, McKenna & Mossar, 1986). Poster et al. found inappropriate behaviour, suicide, aggression and self-abusive behaviour were predominant themes in the drawings as well as a male sexual identity being consistently assigned. Smiles on 62% of drawings of elderly people was taken by Weber et al. as evidence that they were shown as happy, though a proportion of children's drawings always show smiles, regardless of intended emotion. Lots of use of bright colours and full face views was taken as meaning optimism. The lack of control drawings and corroboration of these opinions casts doubt on the validity of these findings. Earlier research had shown that drawings of elderly people were more negative in content, showing the degenerative process (Falchikov, 1990).

Drawings and Sexual Abuse The American Bar Association supports the use of drawings to facilitate children's testimony and they have been admitted as evidence in child sex abuse litigation (Malchioldi, 1990). This is due to a belief that projective techniques and drawings

allow the thoughts and feelings of the child to manifest themselves in a manner unattainable in clinical interview (Miller, Veltkamp, Lane & Janson, 1987) and provides relevant information that the child is unable to disclose directly (Babiker, 1993). The artwork of sexually abused children may provide valuable diagnostic indicators of stress assumed to be associatcd with the experience (Burgess, McCausland, & Wolbert, 1981) and children's drawings may be a useful associative tool for assessing and accessing traumatic memories (Burgess & Hartman, 1993).

Many researchers have focused on the study of diagnostic indicators of sexual abuse (Cohen-Liebman, 1995), but there is no consensus as to which specific indicators are useful. Many of the studies have been found to lack required methodology such as strong inter-rater reliability and blind raters, meaning that researcher bias may affect the results (Trowbridge, 1995). A variety of indicators have been proposed, but there are no operational definitions for variables that are very subjective, questioning their validity and reliability. It is also unclear whether the normal stages of development have been taken into consideration in the evaluation of many features. Genitalia are most often cited as relevant (Trowbridge 1995; Riordan & Verdel, 1991; Hagood, 1992) though they occur very rarely. Tests such as the HTP are no longer valid for sexual abuse given the higher level of exposure of children to sexually explicit material in the decades since such tests were developed (Hagood, 1992). Drawings should not be used alone for the assessment of sexual abuse (Hagood, 1992), but only within the context of other effective counselling practices and integrated with a verbal description (Sadowski & Loesch, 1993), though this then runs the risk of obtaining an illusory correlation.

Signs as production problems Without understanding the production problems associated with each feature, it is difficult to assess their interpretative use (Freeman, 1975). Transparencies, size distortions, distance and placement can all be re-analysed in terms of production problems related to the process of drawing.

There are different sorts of transparency which occur in a HFD. One involves inclusion of internal details such as the stomach, which

supports the idea that the area within the boundary is the inside as well as the outer surface of the object. In the projective tradition, these transparencies were seen as indicative of schizophrenia/mania (Machover, 1949). Another transparency is when one part fails to occlude another as it should, for instance when the body outline is seen through the clothes. These therefore reflect the child's limited skills to cope with the task of representing three dimensions in two dimensions. The occurrence of transparencies in HFDs is usually low even when children are specifically asked to draw a figure with a specific transparency-inducing article of clothing (Mann & Lehman, 1976). The principle of each body part to its own space (Goodnow, 1977) naturally inhibits the production of transparencies and as children normally develop the HFD with clothes incorporated, those types of transparencies are rare. Profile figures more often produce transparencies where a lack of planning leads to the body contour showing through the arm that crosses it. As the drawings become more ambitious, figural overlap occurs more frequently.

Cultural variations were found for the occurrence of transparencies in scene drawings (Andersson, 1995). African children used more x-ray strategies but also used more advanced projection systems, questioning the assumption that transparencies are part of a lower stage of development. Stage type development may not be universally valid and some drawing strategies may be more culturally accepted than others. This supports Costall (1989) that intellectual realism is not a stage but a strategical choice.

The most obvious size distortion of the typical child's HFD is an oversized head. In the use of the HFD as a projective technique, the head size was traditionally seen as expressing "needs and responsiveness, intellectual strivings, and attempts to control the emotions" (Urban, 1963, p. 31), and Machover (1949) found oversized heads typical of children and youngsters with emotional and social maladjustment. When Koppitz (1968) evaluated this claim, however, she found that a large head occurred just as often on the HFDs of normal as disturbed children and occurred on a large percentage of all children's drawings.

The oversized head may be the result of its being the first feature drawn (Freeman, 1980; Allik & Laak, 1985). The lack of planning typical of young children's drawing means that they fill too much of the available space with the feature drawn first, leaving proportionately less room for the rest of the figure. If the child was asked to draw the head onto a pre-drawn figure, thc head was reduced in size (Selfe, 1983) as well as becoming stable in size over time (Allik & Laak, 1985). The presence of a neck in the pre-drawn figure may have altered the results in the Selfe study, but using figures with necks of different sizes or no necks at all gave similar findings (Thomas & Tsalimi, 1988). The smallest heads were produced on the figures with necks, the largest on the free drawings. The more realistic proportions, however, were achieved by drawing the torso first which is something children rarely do.

Freeman also suggested that the head may be drawn larger because it included more details than the body. However, when asked to draw a man showing his teeth, the head was not made any bigger (Henderson & Thomas, 1990). The head was made smaller though, when a back view was requested and the torso was increased in size to include jacket and buttons detail. Children attempt to draw new elements of a figure in proportion to the parts already drawn and the first item in a certain proportion to the space available on the page. They may have problems re-scaling the later items due to the stereotyped nature of their drawings (Allik & Laak, 1985). The child has a relatively stable concept of size in which any topic has to be drawn and this can be reliable over a two week interval.

The size of the HFD is widely used in the projective literature, usually as indicative of self-esteem or depression (e.g. Lehner & Gunderson, 1953). The size of the figure may be a useful clue in determining the level of mental functioning of children (Zuk, 1962). Children aged 6 or 7 drew smaller figures than those aged 12 to 14 years, taken as showing a relationship between mental age and figure size. This may simply be a relationship between chronological age and size, however, since evidence shows a tendency for HFDs to increase in size with age (Payne, 1990).

The importance of the topic may also affect the overall size of the human figure. Three quarters of children in Barbados where society and the family is usually dominated by the mother, drew females larger than males (Payne, 1990). This effect is also seen in experiments using children's drawings of Santa Claus at Christmas (e.g. Solley & Haigh, 1957; Craddick, 1961; Scchrest & Wallace, 1964). Typically, drawings of Santa Claus were seen as increasing in size at Christmas due to the importance of the topic for the child (Thomas & Silk, 1990). However, increased exposure to the figure over time may have meant that the children anticipated including more detail in their drawings nearer Christmas. Also, it was found that the size did not diminish after Christmas as expected (Wallach & Leggert, 1972). Children's drawings of their preferred presidential candidate did not increase in size as the election grew closer, as predicted (Truhon, 1990). Some candidate drawings were found to decrease in size after the election but Truhon considered that the candidates were not important enough to the children to show the size effect.

Using witches at Halloween (Fox & Thomas, 1990) it was found that drawings of threatening or nasty topics were drawn reduced in size only for those children in whom the topic elicited anxiety. Therefore the emotions associated with the topic and not just its importance or significance are important as well. Fox (1989, cited in Thomas & Silk, 1990) found that children's drawings of their parents were larger compared with ordinary men and women, perhaps the most concrete evidence that emotional significance does affect the size of the drawn figure.

The character labels used by children for HFDs that differ in size follow a pictorial convention with larger figures usually associated with a nasty characterisation and smaller figures with a nice characterisation. This has been supported by picture judgement tasks (Jolley, 1995). Where the child imagines personal contact with the figure, however, a smaller figure is chosen as nasty which is seen as a defence response. This effect using perception tasks has been found to be unreliable in children's own drawings, and over different testing sessions, children have been found not to use one principle consistently. It may be that the perception task makes the size

difference and task demands easy and obvious for the child whereas in the drawing task, children are not cued to use size as a possible distinguishing characteristic. Children copying an outline of a man, imagining him to be nice or nasty, were seen to employ this formula but this did not carry into other objects such as an apple (Thomas, Chaigne & Fox, 1989).

Success in the production task may depend on a more effective manipulation of the emotional characterisation (Black & Niven (1993, cited in Jolley, 1995). It is still not clear which principle the child will employ in any emotional state of mind and the level of fear experienced by the child towards the topic needs to be validated in order to make any firm conclusions.

Children will translate the personal significance of a figure into a spatial distance. Pet owners drew their pet closer than family figures (Kidd & Kidd, 1995), and children drew themselves closer to an in-group member than out-group (Holmes, 1995). Bombi and Pinto (1994) confirmed the prevalence of cohesion over distancing as methods of drawing self and friend. However, this effect depends on the control of points of reference available for locating elements on the page (Thomas & Gray, 1992). The edge of the paper is used as a cue sometimes, the pre-drawn figure is used as a cue at other times. Cultural variations in social scaling have also been found between African and Swedish children's drawings (Andersson, 1995; Aronsson & Andersson, 1996). Given the impact of other factors on placement, it should not be relied on too much as a sign of emotional attitude.

Summary The signs found in children's drawings used as a projective technique can also be assessed as production problems. Transparencies are often the result of the child's effort to portray action and movement. These transparencies may have clinical significance in some cases but it is important to note that they also occur for other reasons too such as a lack of planning. The size distortions that occur in children's HFDs can also be seen as a procedural problem and the result of bad planning. The oversized head may result from its emotional and intellectual significance for the child but it may equally be due to the lack of foresight used in the drawing procedure. The item

drawn first is often drawn too large to fit the rest of the figure on the same page with the same proportions. Research on the emotional significance of topics and the effect this has on the size of the drawing has found conflicting evidence. Some results point to an effect but this effect may be too unreliable to be taken seriously. Distance and placement has also shown possible effccts related to the significance of the relationship between elements in a picture, but the influence of other factors specific to the drawing process cautions against reliance on these features.

Emotional Indicators

Elizabeth Koppitz (1968) provided a different method for the interpretation of HFDs. She devised a list of emotional indicators (EI) which reflected the emotional maturity and health of the child as seen in the HFD. These were based originally on Machover, Hammer and Koppitz's own clinical experience, though given a more empirical basis (Thomas & Jolley, 1998). These items were also aimed specifically at children's drawings, rather than simply being extended from adult's to children's drawings. However, they show arbitrariness of choice and display prejudices about human behaviour and drawing, since indicators seen as signs of disturbance may also be legitimate artistic devices for the portrayal of response to feeling (Paine, 1992).

The indicators had to have clinical validity and be able to differentiate the children in a normal sample from those in a clinical sample. They must be unusual and occur infrequently, i.e. on 15% of the drawings or less at any given age, and must not be related to age or maturity. Items were grouped into those concerned with the quality of the HFD such as asymmetry and figure size, those regarding special features in the drawing such as genitals or teeth, and items omitted such as nose or feet. The emphasis in the use of these indicators is not on their individual status but the total number which groups of children achieve.

Normative data from nearly 2000 drawings from children of primary school age determined whether the indicators were rare and related to age or maturity. An original list of 38 items was cut down to 32 from the data from the normative test due to items such as 'hands

hidden' increasing with age and maturation and 'vacant eyes' occurring quite often in the drawings. Some items were given specific ages when they became significant such as poor integration valid at age 7 for boys, 6 for girls, due to their occurring at higher than acceptable levels before these ages.

These items were validated by comparing the EI scores of two groups of children matched for age and sex. The clinical group was attending a child guidance clinic for emotional/behavioural difficulties. Their WISC scores ranged from 90-148, with a mean of 110. The control group was rated as outstanding 'all round' by their teachers with "good social, emotional and academic adjustment" (p. 40), but their WISC scores were not reported. Only 30 of the 32 indicators that had fulfilled the other two criteria could differentiate between the two groups ('figure cut off by paper' and 'sun' were not valid). These became the final set of items used in any further analysis (see appendix 1 for full list).

Koppitz also investigated four different subgroups of child psychiatric patients - aggressive, shy/depressed, neurotic stealers and psychosomatic complaint sufferers - to discover the relationship between the EIs and behaviour symptoms. There was not found to be a one-to-one relationship between indicator and behaviour (something she herself had warned against looking for or expecting) but groups tended to draw certain patterns of indicators more often than others. For instance, shy children showed hands cut off, tiny figure, no nose and no mouth. This compared with aggressive children who tended to show asymmetry of limbs, teeth, long arms, big hands and genitals.

Koppitz (1984) also assessed the HFDs of older children and attempted to describe categories of emotional disturbance, such as 'impulsivity', insecurity', anxiety', and 'aggressive' which corresponded to clusters of EIs. The evidence these clusters are based on is less than clear and the experimental procedure employed earlier is not used in this later work. Studies trying to validate these clusters had only minimal success (Eno, Elliot & Woehlke, 1981; Kurdek & Darnell-Goetschel, 1987). The clinical sample used by Koppitz was compiled mainly from learning disabled populations and it is unclear whether the emotional disturbances of the child guidance clinic children of the 1968

studies were also involved in 1984. Golomb (1992) criticised Koppitz's work stating that "the empirical evidence upon which Koppitz claims to have based her clinical interpretation seems to crumble under scrutiny" (Golomb, 1992, p. 272). Altogether, this expansion of the theory to older children by Koppitz is unsatisfactory and reinforces earlier findings that the DAP test is unsuitable for children over 12 years of age (Harris, 1963).

Evaluation of the Emotional Indicators The diagnostic validity of using the overall indicator scores has been shown to be better than evaluating individual signs with reference to a clinical interpretation. Where it is reported, the inter-rater reliability for scoring the EIs is usually good, averaging around 0.8 or 0.9. The lowest reported is 0.54 (Yama, 1990) but several report figures much higher such as 0.87 (Bachara, Zaba & Rascin, 1975) or 92% agreement (Tharinger & Stark, 1990).

The Koppitz EIs have been shown to be successful in discriminating between children with learning and visual problems and children with no such problems (Bachara et al., 1975). Significantly more EIs were found in the clinical (mean 2.9) than the control group (mean 1.5). Signs showing feelings of inadequacy and a general sense of insecurity and helplessness occurred significantly more often in the drawings of the learning and visual problems group. Omission of feet and hands and attention to eyes (though this was not a Koppitz indicator) appeared more on the drawings from this group which was taken as showing lack of footing, direction and self-assurance, insecurity, inadequacy and inferiority. The validity of this interpretation depends on the extent to which the children in this study could be considered as having those problems suggested by their drawings.

Support has also been found comparing special education categories (learning disabled, educationally handicapped, behaviour disordered) for counselling-referred vs. non-referred samples (Eno et al., 1981). Few scorable drawings with more than four EIs were obtained which causes concern over their practical utility but is what gives them clinical significance. The referred group contained a higher number of indicators and were more likely to score certain indicators reflective of

those personality traits related to referrals (anxiety, poor self esteem, aggressiveness), a somewhat circular proof since they had already been referred. Clinicians failed to discriminate the drawings on the basis of inspection alone, a result in complete contrast to others (e.g. Dieffenbach, 1977). Factor analysis also did not support the Koppitz (1984) clustering of items, reinforcing the view that the indicators should be used as a global measure and not individually.

In an attempt to use the DAP test with a younger population, Goldman and Warren (1976) combined the Koppitz EIs and the Evanston Early Identification Scale (Landsman & Dillard, 1967) to construct a scale to identify children with emotional problems at kindergarten level. Koppitz items correlating with judgements of clinical risk (by classroom observation) included 'no body', 'no mouth', 'poor integration', 'tiny figure' and 'no nose'. Two factors were found which seemed important to the authors. High loading on the first factor were found for high risk and omissions of central body parts; peripheral body parts omissions had high factor loading on the second factor, both of which suggest a relationship between body integrity and severity of problem. The feature 'no body', however, is a normal part of the developmental sequence for some children who draw 'tadpole' figures, which casts doubt on the clinical significance of this item at the age studied in this research. Poor integration and no nose are also items which may not be unusual on such young children's drawings.

EIs have been successfully induced by a stressful experience though the increase was mediated by the psychological meaning of the event (Sturner, Rothbaum, Visintainer & Wolfer, 1980). Higher numbers of indicators were found in a drawing obtained after a blood test compared with a drawing obtained beforehand. However, this only occurred when the children were not prepared for the stressful blood test. Goodenough-Harris scores from the drawings did not change, so the GH scores as an index of cognitive functioning were not effective in distinguishing between the stress conditions. Previous negative findings for the indicators and state anxiety may have been due to the provoking event not being truly stressful (Engle & Suppes, 1970) or the second drawing coming too late after the event (Melamed & Siegel, 1975). Sturner et al. found little relationship between the indicators

and other anxiety indices which may have been the result of time differences in measurement or that the indicators tap different processes, but to validate the EIs as a measure of state anxiety, a relationship should be expected. Research on stress and anxiety using drawings has been criticised for failing to control for IQ and drawing quality (Sims, Dana & Bolton, 1983). The drawing may not reflect the internal psychological state, but instead anxiety may affect problem solving skills whereby overlearned responses are increased and primitive drawings obtained (Engle & Suppes, 1970). A positive relationship was found between the EIs on deaf children's drawings and emotional status measured on the Stress Response Scale by teachers (Johnson, 1989).

Predictive validity of the Koppitz indicators was shown in a study where a count of the EIs proved to be a highly significant predictor of later maladjustment over a period of 6 years (Currie, Holtzman & Swartz, 1974). A significant correlation (0.44) was found between adjustment and indicators. The number of indicators, however, was very low, even in the worse adjusted group (mean = 2). Five year old children classified as 'well-adjusted', 'adequately-adjusted' or 'possibly maladjusted' using the EIs (0/1/2+) were significantly differentiated using scores from the Goodman Child Learning Style (GCLS) and lock box tests (Glutting & Nester, 1986). The hit rate for the three groups overall was above chance but for the possibly maladjusted (2+ EIs) was only at chance level. This implies that the 'normal' children are adequately identified, but the problematic children are not. The EI scores were related to the motivational factors of the GCLS but not the cognitive factors of the lock box test.

Lingren (1971) attempted to replicate Koppitz's finding for shy and aggressive children matched in pairs for age and IQ. No differences were found between the groups (even using 0.1 as a p-value). This may have been because the HFD is not valid as a diagnostic tool but may also have been because the groups were not as severely disturbed as those of Koppitz. Lingren notes that it is possible to find statistically significant differences between groups in the literature, but their practical significance is limited. The Koppitz indicators have also failed to discriminate between a normal and disturbed sample rather than

between two disturbed groups (Fuller, Preuss & Hawkins, 1970). Only nine out of the 30 items were found more often on the clinical HFDs and using a cut-off of two EIs, only 58% of emotionally disturbed were identified compared to 82% of the normals. This study did not use total indicator scores however which are usually more successful. The groups were also not matched adequately since the controls were simply randomly picked from each age level. It is still notable, however, that certain features were isolated successfully

The indicators were also not successful when comparing children with and without behaviour disorders (Dieffenbach, 1977). Poor retest reliability (2-week interval, r=0.2) was found and only half the children were correctly identified by the Koppitz method, a result not significantly different from expectations by chance. Only the indicator 'no eyes' discriminated between the groups. Dieffenbach concluded that this research must seriously question the validity of the Koppitz EIs, though the matching procedure is not clear which presents a problem for adequately assessing the results. A special needs teacher asked to discriminate the drawings using her own implicit intuitive method of identification identified a significant 63% of the sample. This implies that there was some visible difference between the drawings which was not being identified by the indicators. The fact that the controls had a group administration of the DAP test instead of individual like the experimental group may have affected the data due to children copying from each other or discussing their drawings with each other.

Neither the presence of the examiner nor teacher rated skill in art affected the indicators in children's HFDs but no agreement was found between the Koppitz indicators and Children's Personality Questionnaire results (Pihl & Nimrod, 1976). No differences were found either, on the GH scores of hearing impaired and normally hearing children and the EIs did not perform as predicted in determining emotional disturbance (Cates, 1991). Both indicators of aggression, and judges, failed to discriminate aggressive and non aggressive children (Norford & Barakat, 1990). These children were outside the primary school age range where most positive results have been found, however, and the DAP test is considered most valid. Bereaved children,

were also found no more likely to include indicators of emotional disturbance in their HFDs (Forrest & Thomas, 1991).

The predictive validity of some of the Koppitz indicators was assessed using Vietnamese refugee foster care placements as an index of psychological adjustment, comparing drawings at one point in time with the number of placements 5 years later (Yama, 1990). Drawings were rated on four measures which were overall artistic quality rating, overall bizarreness rating, estimated adjustment of the client and 17 of the Koppitz EIs. All measures except EIs were found to be predictors of the number of foster placements 5 years later. The operational definition of psychological adjustment in terms of foster care placements is a problem for generalising from this study. Doubt may be cast on the indicators, given this result, but not necessarily the use of the DAP test since the bizarreness rating was found to be as effective as all the other measures put together. This implies that the relevant information is carried in the global features of the drawing. This also implies that the drawing carries long term characteristics of a person, in contrast to those experiments manipulating anxiety and other transient states, such as pain.

Drawings of 'self' and 'self in pain' were obtained from children suffering from sickle cell disease in an attempt to determine whether EI scores would change, reflecting the child's changed emotional state (Stefanatou & Bowler, 1997) The drawings were examined both for Goodenough-Harris scores and the Koppitz indicators. The expected difference in indicator scores over the two pictures was not found. The GH scores decreased from the 'self' to 'self in pain' drawing, which the authors believed showed something about the child's cognitive functioning during the pain crises. The 'in pain' drawing, however, was always the second one drawn and usually the figure was part of a scene, both of which can reduce the GH score.

The Koppitz EIs also failed to discriminate between impulsive and reflective children of differing SES and age (Soliscamara & Mata, 1985). Impulsivity indicators not part of the Koppitz list however did correlate but only for older children of low SES. It was suggested that a higher level of impulsivity needs to be present to find a relationship between cognitive impulsivity and HFDs. The failure of the indicators was

apparent, but as with the Yama study, something in the drawings was identified, for at least some of the children, which the indicators were not sensitive to.

The failure of the Koppitz system, but not necessarily the DAP test was also found by Tharinger and Stark (1990). They compared the Koppitz system and an integrative system (qualitative analysis using four characteristics) and also facial expressions, for their ability to discriminate between children with either mood disorder, anxiety disorder or both. Significant differences were not found for the Koppitz indicators but were found for the integrative system. Equally, the Koppitz indicators did not correlate with reported self-concept but the integrative system scores did. The depressed children drew fewer happy faces but did not draw more sad ones. There was some success for the indicators since only 23% of the controls had 3 or more indicators whereas 41% of the clinical samples did, but correlations suggested the Koppitz and integrative systems were each measuring something different. The Koppitz system may only discriminate on a broad between-group basis of clinical versus normal samples. The integrative system, however, only works when the groups' drawings are compared with each other, and cannot be used in isolation.

The Koppitz indicators have also been used to compare the HFDs of alleged sexual abuse victims (ASAV) with non-abused children (Hibbard & Hartman, 1990). Few clinically significant differences and no statistically significant differences between the groups EIs were found, however, questioning the validity of the indicators in this situation. Some of the indicators were found on more than 16% of both the ASAV and control drawings. The authors suggested that some situationally specific emotion regarding an impending physical examination affected the indicators produced in the drawings. However, the fact that some indicators occurred more often than the upper limit for significance of an indicator in normal children questions their validity in the clinical population.

Summary The studies seem to show that where differences were found in indicator scores, other problems associated with the methodology mean that conclusions can only be tentative. The studies use different

samples, some assessing within disturbance, other across normal and disturbed groups. Cognitive factors are implicated from the use of educational samples (Eno et al., 1980) and cognitive tests (Glutting & Nester, 1986), but it is unclear exactly what this relationship is. The issue of whether the indicators measure state or trait disturbances is also not resolved. Where the indicators failed to discriminate groups, other measures were more successful. One study seemed to show that intuitive identification was successful, (Dieffenbach, 1977) though another did not (Eno et al.), but these were for different sorts of samples so it is difficult to make any firm conclusions. Global judgements and ratings appear to have more success (Yama, 1990; Soliscamara & Mata, 1985; Tharinger & Stark, 1990), though it is unclear what these measure. They do imply, however, that there is some difference between the drawings which the indicators are simply failing to identify.

Draw-A-Person: Screening Procedure for Emotional Disturbance

A recent test similar to Koppitz involving collective use of features is the DAP:SPED (Draw-A-Person: Screening Procedure for Emotional Disturbance, Naglieri, McNeish & Bardos, 1991). Signs of emotional disturbance in the drawings of a man, woman and self are accumulated into a score that is standardised against normative data. These scores have been found to be significantly higher in a group of students with emotional disturbance attending a day treatment facility than a group of matched controls (Naglieri & Pfeiffer, 1992) as well as special education students with emotional disorders compared to regular education students (McNeish & Naglieri, 1993). Only the second study matched for IQ, however, and the considerable overlap between the ranges of scores in both studies questions their discriminative ability. The DAP:SPED also failed to discriminate between emotionally disturbed and undisturbed deaf children and misclassified them as to emotional functioning (Bricetti, 1994). The DAP:SPED is not valid for deaf children because there are differences between the drawings of deaf and hearing children.

The DAP:SPED is based on objective scoring rules, a global rating and a standardised scoring system and has shown good reliability and validity (Trevisan, 1996). Whether it shows an improvement on

previous DAP test systems is unclear and a direct comparison of the methods on the same sample would be necessary to clarify their comparative values. The features included in the DAP:SPED are extremely similar to the Machover and Koppitz items and thus may be cxpected to suffer from the same problems. New normative data were compiled which benefits the validity of the system, but only for American children and cultural variations in children's drawings have to be acknowledged (Wilson & Ligtvoet, 1992). This test has no independent supporting evidence and has yet to be published and gain popular appeal in the UK. The Koppitz system is among the most popular DAP interpretation system used (Cummings, 1986).

Global judgements

Another method for interpreting HFDs is based on a more global, holistic approach, using the gestalt of the drawing to discriminate between samples. The accuracy of various types of interpreter in discriminating drawings and the differing abilities of expert and novice 'judges' has been the focus of this approach, along with an attempt to determine what criteria are used to assign a drawing clinical status.

The use of trained, experienced psychologists as judges are found in many studies (Jones & Thomas, 1961). A problem exists between the belief in the use of skilled and trained psychologists and the need for the scoring to be objective enough for the lay person to use, thereby making the HFD test more accessible as a diagnostic tool. The assumption is that experts with accumulated knowledge of drawings can more successfully determine personality and behaviour from the HFD than other people. Studies have found, however, that formal training is not necessarily related to success in interpreting the HFD and experts are no more successful than amateurs. No differences in performance have been found between psychologists and non-psychologists (Hiler & Nesvig, 1965), clinicians, 1st and 3rd year trainees (Stricker, 1967), art therapists, mental health workers and lay people (Ulman & Levy, 1973) and teachers, administrators, students, secretaries and professionals (Arkell, 1976). No differences were found between art therapy students at the beginning and end of the year, suggesting this ability cannot be trained (Ulman & Levy, 1973).

Intuition, perhaps through years of unsystematic observation of figure drawings, rather than clinical training, has been suggested as playing a role (Arkell, 1976; Hiler & Nesvig, 1965).

The accuracy of the judges used to discriminate the HFDs of clinical and normal populations has varied in the literature. As already described, clinicians were unsuccessful for special education categories (Eno et al., 1981), though it is unclear whether the clinicians were discriminating the education categories, or those referred for counselling. Wanderer (1969) found that 20 clinicians peer-recommended as experts in the use of the DAP test, including Machover herself, were only able to identify learning disabled individuals' drawings from other categories of clinical groups and a normal group. They failed to diagnose correctly the other groups of neurotics, schizophrenics and homosexuals, even on a second trial. The use of a forced choice design artificially increasing the errors and the degree of overlap among the clinical categories, whilst assuming their mutual exclusivity, were just some of the criticisms of this study (Hammer, 1969). Chambers (1954, cited in Cummings, 1986), replicated the Wanderer design but using Rorschach protocols rather than the DAP test. Judges were able to successfully discriminate all the categories, which suggests a problem with the incremental validity of the HFD rather than with the design of the study as Hammer proposed. It must also be noted that the inclusion of a group of homosexuals as a clinical group, which had its own problems for Hammer, casts doubt on the relevance of the data in contemporary research (homosexuality was removed from the APA classifications of mental illnesses in 1973), especially with children.

Successful discriminations have been found using a special needs teacher who was found able to perform above chance levels (63%) for disturbed boy's drawings (Dieffenbach, 1977). Ninety percent of judges, including art therapists, were able to identify patients' artwork accurately above chance levels, though exact levels are not reported (Ulman & Levy, 1973). Amongst the numerous judges used, success levels were at around 80% for the discrimination of 10 normal and 10 emotionally maladjusted children's HFDs (Arkell, 1976). Hiler and Nesvig (1965) report accuracy rates of 65% for non psychologists, 64%

for psychologists. Stricker (1967) reports that the clinicians got 66% right, whereas the 1st and 3rd year students got 72% and 73% right. These figures for the students do not compare directly with those above, since they were not based on intuitive inspection alone.

The gestalt of an HFD is often evaluated and can lead to a sense of overall disturbance (Tharinger & Stark, 1990). Qualitative analysis was used to assess the psychological functioning of children suffering mood and/or anxiety disorder. This integrative system involved four characteristics: the inhumanness of the drawing, lack of agency, lack of well-being of the individual in the drawing (expressed in a facial expression of negative emotion), and the presence of hollow, vacant, stilted sense in the individual portrayed. This method was found to successfully discriminate between the clinical groups and correlated with reported self concept. Facial expressions were also examined, and significant differences between the groups were found for the number of drawings with a happy face. Depressed children drew fewer happy faces; they did not depict more sad faces.

The level of bizarreness has been found useful as a criteria for discriminating drawings (Yama, 1990; Hiler & Nesvig, 1965). Yama found that the bizarreness rating was as effective as overall artistic quality rating, estimated adjustment of the client and 17 of the Koppitz EIs put together. The ratings were all highly interrelated and their labels depend on the researchers' inclinations for interpretation, but its success implies the information is carried in the global features of the drawing. Hiler and Nesvig (1965) found that though bizarreness discriminated most successfully, it suffered from subjectivity, as some clinicians included anything distorted or out of proportion. Incompleteness (omission of major body parts), distortions (only at the extremes of the scale) and transparencies (obvious ones only) were other criteria which were successful. Happy, pleasant facial expressions were found more often on the normal children's drawings, but suffered from the problem of ambiguity in interpretation whereby a happy smile may also be seen as an unnatural grin. Some criteria described were not successful such as the conflict and anxiety indicators of line emphasis, erasures, size and pressure, and clothing and proportion. A question

arises over the relevance of this research for children, given that this was based on adolescents' drawings.

A prediction formula was devised from Hiler and Nesvig's successful criteria. Drawings score -1 for 'definitely bizarre' and 'major part omitted', and score +1 for 'happy face and 'nothing pathological' with the negative scores classified as patients. Both Hiler and Nesvig and Stricker (1967) found that students using the formula were better than clinicians, though Stricker only reports that the clinicians were reluctant to use the formula and did not explicitly determine those who did and those who did not.

A significant relationship between psychologists' ratings of adjustment and artists' ratings of artistic quality of a group of drawings revealed a possible confounding between psychological adjustment and artistic talent (Whitmyre, 1953). Nichols and Strumpfer (1962, cited in Mortensen, 1991) found that the factor 'quality of drawing' accounted for most of the variance in drawing scores in their study. However, low correlations were found between the artists' and psychologists' ratings in contrast to Whitmyre. The psychologists seemed affected most by the technical skill of the drawing, whereas the artists were concerned with its aesthetic appeal. Overall 'quality of drawing was therefore not necessarily seen as related to psychological adjustment.

"The formal accuracy of the drawn figure and degree to which the figure is differentiated with regard to detail and to individuality" (p. 55) was found to be the biggest single factor accounting for most of the variance in a factor analysis of 17 scales and 40 items derived from the literature (Adler, 1970). This factor was seen to reflect aspects of cognitive maturity or sophistication and develops in children with age and cognitive maturity, also relating to school performance. Many indicators of pathology were found to be a function of this cognitive immaturity. A second factor, similar to that of Nichols and Strumpfer and related to size and placement supports the view that size and placement may be a stable and independent variable in figure drawing. This contrasts with findings of Jolley (1995) and Hammer and Kaplan (1966) who found size and placement unreliable in children's drawings. Two other factors concerned a failure of the integrative process and failure of behavioural control. These smaller factors are independent of

each other but correlate with the first. Adler concluded that figure drawing is a one factor test, that being a cognitive one involving "the maturity or sophistication of the body image representation" (p. 56) and that drawings have other uses only to the extent that the other variables relate to cognitive maturity. This factor also appears to relate to the GH scale, since this also measures the detail and accuracy of the drawn figure. Adler warned against making clinical inferences from drawings unless the level of cognitive maturity is controlled or the drawing behaviour being analysed is independent of cognitive maturity.

Issues in the use of the DAP test

Despite the widespread use of the DAP test, it must be acknowledged that "test popularity is not an index of excellence" (Klein, 1986, p. 381). Research investigating the use of the HFD as a diagnostic tool found clinicians were concerned more with behavioural trends than specific traits (Arbit, Lakin & Mathis, 1959). The more specific the considerations, the less willing the clinician was to base his/her inferences on the HFD. The lack of incremental validity of the HFD test may mean that the additional information obtained from the DAP test is not enough to warrant the time and effort it takes to administer (Gresham, 1993). Defendants point out however that in the short time it takes to administer and score, there is no evidence to show that a simple behavioural method could perform as well (Naglieri, 1993). The issue of incremental validity depends on a test's sensitivity and specificity. When the base rate in a population for a clinical situation is low (e.g. 5%), even if the test has very good specificity and sensitivity (e.g. 95%), accuracy of diagnosis will be low (e.g. 50%) (Klein, 1986). Swensen (1957) criticised the use of percentage of agreement as a measure of reliability, since this is dependent on the base rate for the feature which is being investigated. Early studies (e.g. Bradshaw, 1952; Lehner & Gunderson, 1952, cited in Swensen, 1957) were criticised for not reporting the base rates and therefore rendering the percentage of agreement and their estimate of the reliability of the DAP invalid. Whatever its problems, however, Swensen also noted that clinicians routinely used the DAP test and feel it is valuable.

Swensen hypothesised that cases which did illustrate the theory remained in the clinician's mind and biased him/her in favour of the test. Wanderer (1969) agrees with Swensen that it is the occasional incidence of congruence between the diagnosis and drawing which biases the clinicians to believe in the instrument. It seems that pragmatic validation rather than research confirmation may account for its popularity and utility in clinical experience (Gillespie, 1994). Gresham (1993) explained the perception of, and the belief in, valid diagnoses from HFDs, in terms of an illusory correlation. This refers to "the relationship between test responses and symptoms/behaviour that are based on verbal associations rather than valid observations" (p. 183). Most of the evidence for this type of correlation comes from studies of other projective techniques such as the Rorschach and Incomplete Sentences Blank. Early work (e.g. Morris, 1949) as well as Machover's principles were formed using word association tasks and the HTP technique PDI relies on verbal associations. These, therefore, may suffer from this illusory correlation but its impact on later systems such as the EI scores and other global ratings is limited.

The effect of cognitive maturity

The issue of cognitive maturity in the interpretive use of children's drawings relates to the debate between delay and deviancy in the children's development. Differences that are seen between the drawings of children with emotional problems and those of normally adjusted children can be viewed as a deviancy in the developmental nature of the disturbed children, or a delay normally associated with their mental age. The extent to which the differences in their drawing behaviour is the result of their cognitive immaturity is important, but often overlooked.

The idea that disturbed children's drawings might be deviant comes from an assumption that realism is the endpoint of development and representation is like copying (Golomb, 1992). A standard, similar to the classical idea of ideal forms (Paine, 1992), achieved through the collection of normative data, is used to assess deviations. Belief in the influence of cognitive development on drawings was expressed early in drawing research (Oakley, 1935; Goodenough, 1926). Most early work

in drawings such as the descriptive stage models was involved in achieving the standard and charting the normal progression towards it. The DAM test formalised this into an objective scale. Goodenough saw the DAM test as a measure of intelligence and moderate correlations with IQ measures have been cited (Vane & Kessler, 1964; Abell et al., 1994; Abell et al., 1996; Aikman et al., 1992). The revised GH scale (Harris, 1963) altered the construct from a measurement of intelligence, to a measure of intellectual maturity or conceptual development. The GH scale has been used as measure of cognitive functioning in research (e.g. Sturner, et al. 1980; Stefanatou & Bowler, 1997).

However, this view ignores the problem solving, strategic nature of a child's drawing (Arnheim, 1974; Freeman, 1980; Golomb, 1992). Machover did acknowledge that drawings were produced out of a complex personality formed in the process of dealing with social and psychological problems, but she still thought some aspects of drawings were simply direct indicators of internal states, identifiable regardless of the production process (Freeman, 1975). Pathology indicators presented by Machover can be viewed as relevant only as they relate to cognitive immaturity measured by the formal accuracy and differentiation of the HFD (Adler, 1970). It is also necessary to know what production problems are associated with each feature, before obtaining a meaningful composite score based on a feature count, such as the Koppitz and GH scales (Freeman, 1975). Freeman thought that cognitive capabilities and executive functions such as planning and monitoring which can present problems in the drawing task for those with limited cognitive abilities need to be assessed. A delay in the GH scores of the drawings of clumsy children reflected planning difficulties in drawings, for example, which were more than just a visual perceptual deficit, and implies the involvement of cognitive factors in drawing beyond just visual perceptive influence (Barnett & Henderson, 1992). Research using computers to aid the drawing process has shown the limits which cognitive abilities place on the drawing process, and the danger of relying on drawings for assessment material (Olsen, 1992). This evidence suggests the interpretation of drawings requires acknowledgement of the underlying cognitive skills.

The vast majority of studies reviewed have not controlled for cognitive factors such as the influence of IQ or mental age. It is therefore very difficult to determine whether cognitive variables are affecting the drawings of clinical sample children and affecting the production of the features that are used to interpret those drawings. Where samples were pairwise matched for IQ (Lingren, 1971), no differences were found between the drawings of two clinical samples. Group matched samples controlling for IQ (Wysocki & Whitney, 1965) found some significant differences between the HFDs of children with and without physical disabilities, though this study suffers from other methodological problems. Maloney and Glasser (1982) concluded that some interpretive scales are able to discriminate clinical groups controlled for intelligence which suggests that the DAP is assessing deviant behaviour and maladjustment and not just cognitive immaturity. Golomb (1992), however, found that delay was characteristic of thematic drawings of emotionally disturbed children, compared to normally developing children. The few differences which were found disappeared by age 11, and the most marked differences were found in the use of colour and the child's comments while drawing, rather than any variables discussed by Machover or Koppitz. Golomb saw the negligible effect of IQ on graphic achievement as being congruent with Gardner's theory of multiple intelligences, suggesting that graphic intelligence is separate from spatial intelligence (Gardner, 1985).

The influence of cognitive maturity on children's drawings is an area within the use of HFDs as a projective technique that requires clarification. Due to poor experimental procedures and a lack of replication, it is hard to come to any firm conclusions and is difficult to determine whether the drawings of children depict their emotional or behavioural problems, independently of the effect of a cognitive delay.

Summary Evidence shows that expert judges are no better than novices in interpreting a drawing though many studies have shown evidence for greater accuracy than chance in discriminating the drawing of clinical samples for both types of judge. Various criteria used by these judges have been assessed, with global ratings such as

bizarreness showing the most promise. The issue of whether cognitive delay is responsible for the differences between the drawings of disturbed and normal children is unresolved from a review of the literature. Production problems associated with features depicted, and the influence of the GH scale and the quality and sophistication of the image which is drawn are issues which may be involved in the projective interpretation of the HFD. The vast majority of studies, however, fail to control for IQ or mental age and it is therefore difficult to arrive at any firm conclusions.

CHAPTER TWO

A COMPARISON OF THE HUMAN FIGURE DRAWINGS OF CHILDREN WITH EMOTIONAL/BEHAVIOURAL DIFFICULTIES AND THEIR CHRONOLOGICAL AND MENTAL AGE MATCHED CONTROLS

Introduction

It has been shown in chapter 1 that the DAP test is popular and used for a variety of assessments such as personality and emotional disturbance, sexual abuse and attitudes. The interpretation of individual signs has been extensively evaluated and some of the features that are given clinical status appear to be confounded with production problems. The Koppitz EIs were based on the features collected using the earlier sign interpretation, but were given a more empirical basis and better operational definitions and have also obtained more supportive evidence in the literature. However, there are as many studies supporting them as not and there is no clear evidence to believe they are a valid instrument to use. Study 1 in this chapter aims to replicate Koppitz (1968) study, using similar methodology and samples.

There is, however, the issue of mental age as a factor that may influence drawing, which has not been controlled in many studies, including Koppitz (1968). It is therefore necessary in study 1 to improve on the Koppitz study by controlling for MA to determine whether the differences that were found previously were simply due to a developmental delay on the part of the clinical sample.

The intuitive method of identification, discussed in chapter 1 as an alternative method for examining drawings, has been successful in the literature. Studies have shown that judges are sometimes able to see a difference between clinical and normal drawings, though experts do not usually perform better than novices. Study 2 in this chapter aims to evaluate this method and the issue of whether the judges should be experts or novices will also be examined.

Study 1

Using the Koppitz (1968) emotional indicators

INTRODUCTION

This study was designed to discover whether the Koppitz emotional indicators could differentiate between the HFDs of children with emotional/behavioural difficulties (EBD) and those of normally adjusted children matched not only for chronological age but also for mental age as well as gender and school. In the Koppitz (1968) study, the samples were matched for the whole group's range and distribution of age and sex, but the children were not matched for mental age or school. The clinical sample IQs ranged from 90 to 148, with the children described as 'at least normal' and many 'above average'. The control group IQs on the other hand were 'assumed' to be high average or superior but were not explicitly tested. By controlling these variables more tightly than Koppitz it should be possible to see whether the indicators occur more often on the drawings of the clinical sample. Previous studies have found mixed evidence. Support for the indicators has been found using samples with learning and visual problems (Bachara et al., 1975), and a mixture of learning disabled, educationally handicapped and behaviour disordered children (Eno et al., 1981). However, an early replication of Koppitz (Lingren, 1971) and using a behaviour disordered sample (Dieffenbach, 1977) failed to find support for the indicators.

Individual indicators appear in small numbers on single drawings and attempts to evaluate separate indicators have not been altogether successful (Daum, 1983; Eno et al., 1981). Total numbers of indicators are preferred to individual indicators therefore, in the analysis of results, since it improves their diagnostic power (Koppitz, 1968; Daum, 1983).

The inter-rater reliability of the Koppitz scoring system for emotional indicators was calculated using two other psychologists, ignorant of the sample from which any particular drawing came, as well as the investigator familiar with the sample drawings. This should determine how reliable the scoring system is across different raters as

well as assessing how strong the bias in the scoring may be due to the rater's knowledge regarding the samples.

METHOD

Participants

Clinical sample:

18 children aged 5 to 11 years (mean 8;2, SD 2;1) classified as having emotional/behavioural difficulties (EBD) were taken from case loads of the Pupil Support Service in York. The children were at stages 3, 4, or 5 of the 5-stage model of statutory assessment for special educational needs. Stage 3 involves parents, teacher and the Special Educational Needs co-ordinator together with outside support services helping the child. At Stage 4 the Local Education Authority considers the need for a statutory assessment and makes a multi-disciplinary assessment of the child. At Stage 5 a statement of special educational needs is considered in the light of the multi-disciplinary assessment of the child. It is not known how many were at each stage of the assessment process. There were 15 boys and 3 girls. See table 1 for numbers of children at each age.

Control sample (chronological age matched):

18 children aged 5 to 11 years (mean 8;3, SD 2;4) considered by teachers to be normally-adjusted. Each child was matched for gender and chronological age to a corresponding child in the clinical sample to within a month of the clinical sample child's birthday. These children each came from the same school and often the same class as the clinical sample children they were matched with. There were 15 boys and 3 girls. See table 1 for numbers of children at each age.

Control sample (mental age matched):

18 children aged 5 to 11 years (mean 8;0, SD 2;1) considered by teachers to be normally-adjusted. These children were matched for gender and mental age with the clinical sample, using scores on four subtests of the Wechsler Intelligence Scale for Children, third edition (WISC-III, Wechsler, 1992) or the Wechsler Preschool and Primary Scale of Intelligence, revised edition (WPPSI-R, Wechsler, 1990). These

children came from the same schools as the clinical children they were matched with. Seven of the children (6 boys, 1 girl) in this sample were taken from the CA matched control sample as they also matched a clinical child for mental age. There were 15 boys and 3 girls altogether. See table 1 for numbers of children at each age.

	Clinical		CA control		MA control	
Age at test	Boys	Girls	Boys	Girls	Boys	Girls
5 years	4	0	3	0	3	0
6 years	1	1	2	1	3	1
7 years	1	1	1	1	2	2
8 years	3	1	3	1	1	0
9 years	1	0	1	0	0	0
10 years	4	0	4	0	5	0
11 years	1	0	1	0	1	0
Total	15	3	15	3	15	3

Table 1. Numbers of boys and girls in each sample for each age.

Design

A between-subjects design was used to compare participants' performance on the Draw-A-Person test administered according to Koppitz's (1968) instructions and scored for emotional indicators according to Koppitz's criteria. Scores from a clinical sample of children with EBD were compared with two control samples of normally-adjusted children. The control samples were either matched with the clinical sample for chronological or for mental age.

Materials

A4 blank white paper, pencil and eraser. (For the younger children aged 5 and 6 years, oversized pencils were allowed.)

Four subtests (two verbal, two performance) of the WISC-III or WPPSI-R:

Verbal subtests- similarities
vocabulary

Performance subtests - block design
object assembly

Procedure

Children were tested individually, seated at a table of an appropriate height for them to work at, opposite the examiner. Given a sheet of paper, pencil and eraser, they were given the following verbal instructions: "*On this piece of paper I would like you to draw a whole person. It can be any sort of person you like as long as you make sure it is a whole person and not a stick figure or a cartoon figure.* " The children were allowed as much time as they wished to complete the drawing. Any queries were answered in a non-directive manner.

Children aged 5 years to 5 years 11 months were given four subtests of the WPPSI-R test. Children aged over 6 years were given four subtests of the WISC-III performed in the preferred order stated and following the instructions given in the manual.

RESULTS

Indicator scores for the HFDs in the clinical and control samples were obtained from three raters (MC, YG, KW), two of whom (YG & KW) scored the drawings after a training period, whilst ignorant of the sample to which they belonged. See appendix 1 for the scoring manual. The scores were then subjected to analysis using Kendall's Coefficient of Concordance (W, which converts to a Chi-Square statistic) to determine the extent of agreement among all three raters for the three sets of drawings. Coefficients were calculated for the clinical and control samples separately.

	Clinical	Control
X^2	10.36	7.6
p	0.006	0.02

Table 2. Chi-Square and associated p values for the measure of agreement among the three raters for the clinical and control HFDs' EI scores.

These results show very good agreement among the raters' scores. An agreed score arrived at through consensus decision was used in further analysis.

Children in the three samples were categorised according to how many emotional indicators they produced in their drawing. No child scored above 5 out of the total of 30 indicators. For a full breakdown of participants' indicators see appendix 2.

Due to the two control samples not being entirely independent of each other, two separate comparisons were made: clinical vs. CA matched controls and clinical vs. MA matched controls. The number of children producing HFDs with 0 or 1 indicators are compared with the number who produced 2 or more indicators (the procedure adopted by Koppitz) in each of the samples.

No. of EIs	Clinical	CA control
0-1	10	11
2+	8	7

Table 3 (above left). Numbers of children's HFDs showing either 0 or 1 indicators and 2 or more indicators for the clinical and CA control samples.

No. of EIs	Clinical	MA control
0-1	10	9
2+	8	9

Table 4 (above right). Numbers of children's HFDs showing either 0 or 1 indicators and 2 or more indicators for the clinical and MA control samples.

Chi-Square analysis (with Yates' correction) showed no differences between the numbers of HFDs in the clinical and CA control sample showing 0 or 1 indicators and numbers showing 2 or more indicators (X^2=0.00, df=1, ns). This was also the case for the clinical and MA control comparison (X^2=0.00, df=1, ns).

The clinical sample mean indicator score was 1.44, compared with the CA matched control mean score of 1.22 and the MA matched control sample mean of 1.39. T-tests showed no significant differences between the clinical and CA matched sample scores (t(34)=0.53, ns) or the clinical and MA matched sample scores (t(34)=0.13, ns).

Summary of results

The inter rater reliability results showed considerable agreement among the three different raters for all of the drawings. A consensus opinion of the three raters was used to compare the indicator scores of the

children in the three samples. Chi-Square analysis of the total numbers of indicators children produced in their drawings showed no differences between the clinical and two control samples. No group had any more children producing low numbers of indicators than any of the other groups. T-tests reinforced these results with nonsignificant differences found across the two comparisons.

DISCUSSION

The scores that Koppitz reported may have been influenced by her knowledge of which sample the child belonged to. The inter-rater reliability results of this present study, however, were found to be very high which reinforces the reliability of the scoring manual devised by Koppitz. From the list of 30 indicators, about two thirds are reasonably objective requiring a decision of presence/absence or measurement with a ruler, but one third of the indicators require a more subjective decision. The correlation found here among three scorers, two of whom had no knowledge of which sample the drawing belonged to, would seem to vindicate those potential problems of bias due to sample knowledge, though reflecting the scoring system's subjective nature with a correlation of less than 1.

The results of this study showed no significant differences between the emotional indicator scores of the clinical sample and those of the control samples. This casts doubt on the clinical validity of the Koppitz emotional indicators and their ability to discriminate a clinical from a normal sample, reinforcing the results of researchers such as Dieffenbach (1977), and going against the major work of Koppitz (1968). The design of the study was set up to replicate Koppitz's work by using the chronological age match control. Any differences Koppitz found were expected to occur using this sample as a comparison. However, no differences were found between the two samples' scores in this study.

The control for mental age, which Koppitz failed to adequately account for, would have reduced the differences between the two samples, if the odd features seen in the disturbed children's drawings were due to their delayed intellectual maturity. The results of this study appear to follow this theory since no differences did appear with

the MA match comparison. However, since no differences appeared in the CA match comparison either, there is a problem in resolving this issue.

Why these results occurred as they did may be due to a variety of factors. The most obvious would be as Dieffenbach concluded that the Koppitz emotional indicators simply do not have the ability to discriminate between the clinical and normal population and therefore have no clinical validity. Before coming to this conclusion, however, it is important to consider other factors.

The clinical sample used in this study may not have been emotionally disturbed to a severe enough degree to show a difference in their drawings. The children were all still in mainstream education which does signify that though they had problems these were not severe enough to warrant their being removed from school and their normally adjusted peers. In this respect, therefore, similarities with normal children should be expected. Though the children were considered to have emotional/behavioural difficulties in general and were mostly considered quite disruptive in the classroom, when alone with the examiner, the majority of the children responded as 'normal' children and did not tend to show signs of disturbance in their outward behaviour.

In the Koppitz study, the controls were considered as 'outstanding all round' showing good social, emotional and academic adjustment. The controls in this study were considered 'normally adjusted' which controlled the specific variable in question - that of emotional health - but may have meant that the two samples were not as extreme in their differences as the Koppitz ones. Dieffenbach (1977) also failed to find differences between his sample of behaviourally disordered children and controls defined as 'non-behaviourally disordered'. These samples were also different on what was supposed to be the crucial variable, but may not have been as disparate as the Koppitz samples. It may be useful to replicate this experiment with more severely disturbed children to determine if the DAP test lacked sensitivity, or validity.

The norms from which the emotional indicators were derived were collected by Koppitz in the late 1960s. The reason why the normally

adjusted children of this study showed just as many indicators as their disturbed counterparts may be due to the fact that the norms have changed and no longer apply in the 1990s. Trends in children's drawings change like fashion and therefore it is questionable whether those trends of the late 60s still apply today (Wilson & Ligtvoet, 1992). Evaluating this claim would involve collecting new normative data from which new EIs could be derived according to the Koppitz criteria of occurring rarely and not changing with age or maturation, on which new studies of validity could then be performed.

The question over the severity of disturbance in the clinical sample and the normative data will be assessed in later chapters (see chapters 3 and 4). Before this, it is necessary to examine the 'intuitive method of identification' using the drawings from study 1. Previous research has shown some success with this technique even where other measures have failed, and it is therefore important to determine whether the drawings collected for study 1 can be identified using this alternative method.

Study 2

Using the intuitive method of identification

INTRODUCTION

An alternative to the scoring techniques such as the Koppitz emotional indicators is found in the literature, employing people as judges to visually inspect the pictures and identify the clinical and control samples. Dieffenbach (1977) called this the 'intuitive method of identification'. Success is measured in the accuracy rates of the judges above chance levels. Positive evidence has been obtained using a variety of judges. Accuracy rates above chance have been found for psychologists and non psychologists (Hiler & Nesvig, 1965), clinicians and trainees (Stricker, 1967), art therapists, mental health professionals and lay people (Ulman & Levy, 1973) and school staff, students and professionals (Arkell, 1976). Judges' accuracy varies between 63% (Dieffenbach, 1977) and 80% (Arkell, 1976) and up to 90% of judges perform above chance (Ulman & Levy, 1973). This method

even works when no differences are found between the indicator scores of the drawings (Dieffenbach, 1977), though the use of only one judge limits the scope of this finding. Clinicians failed to discriminate referred and non referred samples of learning disabled, educationally handicapped and behaviour disordered samples (Eno et al., 1981), but the within-subject nature of the discrimination may have made this task more difficult. Peer rated experts in drawing interpretation, including Machover herself failed to successfully discriminate neurotics, schizophrenics and homosexuals from normal participants drawings though they did identify learning disabled participants drawings at above chance levels (Wanderer, 1969). There are problems with the clinical categories used by Wanderer and he also used adult drawings so the results are of limited relevance to children's HFDs.

The use of experts and trained professionals for interpretation is common, though expert judges are thought to perform at a comparable level to novices (Motta et al., 1993). Evidence from Hiler and Nesvig (1965), Stricker (1967), Ulman and Levy (1973) and Arkell (1976) suggests this may be true, since no differences were found between the performance of the expert and novice judges they employed, though the 'experts' were not the same in all cases. Changes in education and training for mental health professionals also questions the relevance of these data from 'experts' for contemporary research.

This study evaluated the intuitive method of identification using the drawings collected during study 1. The performance of people with experience of children's drawings (experts) is compared with the performance of people with no experience (novices) for their ability to discriminate the sample of disturbed children's drawings from the CA and MA matched control sample HFDs.

Different conditions were used to examine what factors would affect the identification of the clinical sample HFDs. Two different formats of the task were used. The group format required the judges to determine the clinical drawings from amongst many others in a randomly mixed pile. The pair format was expected to make the task easier by giving only two drawings and the knowledge that one was definitely from a clinical sample child. The judges were also required to perform the tasks both with and without the children's ages in order to

determine whether this would affect their accuracy. It was expected that accuracy would improve when the children's ages were available, especially if the drawings of the clinical sample children are delayed rather than deviant.

METHOD

Participants

Eight experts (people familiar with children's drawings) and 48 novices (people unfamiliar with children's drawings) took part. All the judges were over 18 years old. Six of the experts were academics who had undertaken research in children's drawings; one was a trained primary school teacher. The novices were selected from the undergraduate student population at the University of York.

Design

Eight tasks were used comprising two format conditions (group or pair) and two control conditions (CA or MA) as well as two age conditions (no ages or with ages). All the conditions were between-subject variables.

Materials

Three sets of drawings collected during study 1 were used, one from a clinical sample of children with EBD (n=18), one from the CA matched controls (n=18) and one from the MA matched controls (n=18). These were made into four tasks:

Group format- (1) randomly mixed group of clinical and CA matched control sample drawings (N=18 per sample)

(2) randomly mixed group of clinical and MA matched control sample drawings (N=18 per sample)

Paired format - (3) set of 18 pairs of drawings, one from the clinical sample paired with the appropriate CA matched control

(4) set of 18 pairs of drawings, one from the clinical sample paired with the appropriate MA matched control

A list of the children's ages coded as for the drawings, was also provided.

Procedure

Judges were asked to discriminate the clinical sample HFDs from the control sample drawings. They were not told how many drawings were from disturbed children. The following written instructions were given. The form of the instructions changed to allow for the different conditions used.

"The following are sets of drawings from British children. The children were all asked to draw a 'whole person'. Some of the children are considered to have emotional/behavioural difficulties, some of the children are considered normally adjusted. Your task is to discriminate between the children's drawings to identify the disturbed children's pictures.

(Group format) This task involves judging each drawing in turn, deciding whether the drawing came from a disturbed child or a normally adjusted one. You must record your decisions on the answer sheet supplied by putting a mark in the column to which you think the drawing belongs.

(Pair format) This task involves judging a pair of drawings, where one is a disturbed child's and one is from a normally adjusted child. You must record your decisions on the answer sheet supplied by choosing which of the pair is the disturbed child's drawing and writing it in the column provided."

The judges were informed which drawings were from boys and which from girls in all conditions. In the 'no ages' condition the judges were told that the children were all aged between 5 and 12 years. In the 'with ages' condition the ages of the children were supplied. No criteria were given for making the discrimination.

RESULTS

Each judge performed the task of discriminating the clinical sample children's drawings from the control sample in one of the following conditions:

	Format (2)	group and pair
	Control (2)	CA and MA
	Age (2)	no ages and with ages

Each judge obtained a total score correct out of the 36 HFDs in the 'group format' task, or out of the 18 pairs in the 'pair format' task. This

score consisted of a score for the number of correct clinical HFDs as a proportion of the number of drawings chosen as clinical, and a score for the number of correct control HFDs as a proportion of the number of drawings chosen as control. In the 'pair format' task, these last two scores were the same as the total scores.

Experts vs. Novices

Due to the small numbers available, the expert judges took part in the 'CA control, group format, with ages' and 'MA control, group format, with ages' conditions only. A test to assess the significance of the difference between two independent proportions was used to determine if the scores that the experts achieved were different from those of the novices in these conditions. The proportion of correct scores was compared across the two sets of judges using the sum values of the judges' correct responses.

Total scores: The number of correct scores as a proportion of the possible total number correct was compared for the four experts and six novices in the two conditions. In the 'group format' condition, the experts' total score was out of å possible 144 (4 x 36) correct responses, and the novices' total score was out of a possible 216 (6 x 36) correct responses.

Condition	Experts	Novices
CA/group/with ages	62/144 (43%)	113/216 (52.3%)
MA/group/with ages	85/144 (59%)	118/216 (54.6%)

Table 5. The total scores obtained by the 4 experts and 6 novices in each condition, as a proportion of the total possible correct score.

The experts' proportions of correct responses did not differ significantly from the novices' in the 'CA control, group format, with ages' ($z=-1.72$, ns) or the 'MA control, group format, with ages' ($z=0.82$, ns) conditions.

Clinical scores: The number of correct clinical drawings as a proportion of the number of drawings chosen as clinical by the judges was compared for the 4 experts and 6 novices in the two conditions. The

denominators in each of the proportions differed according to the total number of drawings that were chosen as clinical HFDs by the judges.

Condition	Experts	Novices
CA/group/with ages	37/79 (46.8%)	48/91 (52.7%)
MA/group/with ages	33/53 (62.3%)	44/82 (53.7%)

Table 6. The clinical scores obtained by the 4 experts and 6 novices in each condition, as a proportion of the total number of drawings chosen as clinical.

The experts' proportions of correct responses did not differ significantly from the novices' in the 'CA control, group format, with ages' ($z=-0.77$, ns) or 'MA control, group format, with ages' ($z=0.99$, ns) conditions.

Control scores: The number of correct control drawings as a proportion of the number of drawings chosen as control by the judges was compared for the 4 experts and 6 novices in the two conditions. The denominators in each of the proportions differed according to the total number of drawings that were chosen as control HFDs by the judges.

Condition	Experts	Novices
CA/group/with ages	25/65 (38.5%)	65/125 (52%)
MA/group/with ages	52/91 (57.1%)	74/134 (55.2%)

Table 7. The control scores obtained by the 4 experts and 6 novices in each condition, as a proportion of the total number of drawings chosen as control.

The experts' proportions of correct responses did not differ significantly from the novices' in either the 'CA control, group format, with ages' ($z=-1.77$, ns) or the 'MA control, group format, with ages' ($z=0.28$, ns) conditions.

Analysis of the different conditions

This analysis looked at the numbers of judges performing at and above chance levels in each of the conditions, using the novice data only, since there was an equal number of novice judges in all the conditions.

Using the percentages of correct HFDs identified allowed comparisons across score types by controlling for the differing denominators in the clinical and control scores, as well as the 'group' and 'pair format'.

Condition	Total %	Clinical %	Control %
CA/group/no ages	51	50	52
CA/pair/no ages	59	59	59
MA/group/no ages	53	55	52
MA/pair/no ages	57	57	57
CA/group/with ages	52	51	53
CA/pair/with ages	55	55	55
MA/group/with ages	55	54	56
MA/pair/with ages	62	62	62

Table 8. The average percentage score in each of the conditions.

None of the judges was able to perform better than chance in any of the conditions, using any of the three score types. The average score for the judges in all of the conditions ranged from 50% to 62%. The highest average percentage score of 62% in the 'MA control, pair format, with ages' condition converts to a raw score of 11 correct out of 18 pairs. This is not significantly different from chance expectation. General factorial ANOVAs on the percentage scores for the 6 judges in each condition showed a significant main effect of format for the total ($F(1, 40)=7.44$, $p<0.01$), clinical ($F(1, 40)=6.53$, $p<0.05$) and control ($F(1, 40)=6.34$, $p<0.05$) scores. The average scores in the 'pair format' were consistently higher than the 'group format' conditions, though it is important to note that neither of these scores was above chance level.

Individual HFD Analysis

The consistencies in the decisions that the judges made about the clinical status of each drawing were investigated, to determine whether the judges agreed in any of their decisions. The data consist of numbers of judges correctly and incorrectly identifying each of the 18 clinical, CA control and MA control HFDs. Binomial tests determined whether certain drawings were identified, correctly or incorrectly, by more judges than would be expected by chance (using the familywise error rate of $p<0.05/18=0.003$).

HFD	✔	✘	p<
C1	11	37	0.0003
C2	11	37	0.0003
C3	12	36	0.0009
C4	35	13	0.0024
C5	16	32	ns
C6	20	28	ns
C7	37	11	0.0003
C8	25	23	ns
C9	35	13	0.0024
C10	36	12	0.0009
C11	39	9	0.0000
C12	24	24	ns
C13	11	37	0.0003
C14	33	15	ns
C15	45	3	0.0000
C16	19	29	ns
C17	23	25	ns
C18	12	36	0.0009

Table 9. Numbers of judges correctly and incorrectly identifying the 18 clinical HFDs, with binomial test results.

Mann-Whitney tests comparing the number of judges who were correct with the number incorrect for each of the 18 clinical HFDs showed no significant difference between the number of judges correctly identifying the drawings and the number incorrectly identifying them ($z=-0.14$, ns).

The above table shows that 6 of the drawings (C4, 7, 9, 10, 11, 15) were identified correctly by significantly more judges than would be expected by chance. Equally, 5 of the drawings (C1, 2, 3, 13, 18) were identified incorrectly by significantly more judges than would be expected by chance. The remaining 7 drawings were identified correctly by no more judges than incorrectly.

HFD	✔	✘	p<
CA1	14	10	ns
CA2	8	16	ns
CA3	3	21	0.0003
CA4	13	11	ns
CA5	19	5	ns
CA6	9	15	ns
CA7	15	9	ns
CA8	23	1	0.0000
CA9	20	4	0.0015
CA10	21	3	0.0003
CA11	20	4	0.0015
CA12	19	5	ns
CA13	4	20	0.0015
CA14	17	7	ns
CA15	15	9	ns
CA16	8	16	ns
CA17	13	11	ns
CA18	6	18	ns

Table 10. Numbers of judges correctly and incorrectly identifying the 18 CA matched HFDs, with binomial test results.

Mann-Whitney tests comparing the number of judges who were correct with the number incorrect for each of the 18 CA matched HFDs showed no significant differences between the number of judges correctly and incorrectly identifying the drawings (z=-1.5, ns).

The above table shows that 4 of the drawings (CA8, 9, 10, 11) were identified correctly by significantly more judges than would be expected by chance (using the familywise error rate of $p<0.05/18=0.003$). Equally, 2 of the drawings (CA3, 13) were identified incorrectly by significantly more judges than would be expected by chance. The remaining 12 drawings were identified correctly by no more judges than incorrectly.

HFD	✔	✖	p
MA1	7	17	ns
MA2	16	8	ns
MA3	7	17	ns
MA4	19	5	ns
MA5	9	15	ns
MA6	11	13	ns
MA7	21	3	0.0003
MA8	13	11	ns
MA9	17	7	ns
MA10	16	8	ns
MA11	19	5	ns
MA12	20	4	0.0015
MA13	13	11	ns
MA14	15	9	ns
MA15	15	9	ns
MA16	16	8	ns
MA17	23	1	0.0000
MA18	6	18	ns

Table 11. Numbers of judges correctly and incorrectly identifying the 18 MA matched HFDs, with binomial test results.

Mann-Whitney tests on the above data comparing the number of judges who were correct with the number incorrect for each of the 18 MA matched HFDs showed a significant difference between the number of judges correctly and incorrectly identifying the drawings ($z=-2.6$, $p<0.01$)

The above table shows that 3 of the drawings (MA7, 12, 17) were identified correctly by significantly more judges than would be expected by chance. However, none of the drawings was identified incorrectly by significantly more judges than would be expected by chance. The remaining 15 drawings were identified correctly by no more judges than incorrectly.

DISCUSSION

The results of this investigation show that there were no differences in the performance of the expert and novice judges in discriminating the

clinical from the control samples at least in the 'group format, with ages' condition. This reinforces Motta et al. (1993) and the results of previous studies showing no difference between expert and novice judges (e.g. Hiler & Nesvig, 1965; Stricker, 1967; Ulman & Levy, 1973; Arkell, 1976).

However, the results showed that all the judges could not discriminate the HFDs above chance levels, with accuracy rates around 55%. This is in contrast to previous studies which showed accuracy rates of between 63% and 80%. There was a significant effect of format in the ANOVA result due to the fact that the average percentage scores in each of the 'pair format' conditions were consistently higher than in the 'group format'. The judges seemed to perform better in the 'pair format' where they have a comparison drawing for each clinical one. This was expected to make the task relatively easier, as the judges can directly compare the two pictures each time they make their decision, knowing that one of the HFDs is from a clinical sample child. It is important to note, however, that the improvement in success was not enough to take the judges' scores beyond chance levels.

The lack of effects of age and control are interesting. This shows that the judges were not affected differentially by whether the clinical sample drawing was paired with its CA or MA matched counterpart. This illustrates a similar pattern as was found with the emotional indicator scores in study 1, with no difference between the two control comparisons. The lack of an age effect shows that the judges' performance was not improved by having the children's ages made available to them. If the clinical sample children were drawing more immaturely than the controls, then it might have been expected that when given the ages of the children, the clinical drawings would stand out as of a similar age but younger looking in the 'CA control' condition, or of an older age but similar looking in the 'MA control' condition. Any immaturity in the drawings of the clinical sample children would be revealed by giving the age of the child. The judges were unaffected by the addition of the ages so this theory cannot be supported from these results. However, these results do not support the argument that the drawings of the clinical sample children are deviant as this would have allowed the judges to identify them, which was not the case.

The fact that the judges could not separate the two sets of drawings better than chance may be due to the lack of differences between the drawings, in terms of indicator scores as well as other measures such as spatial orientation on the page, items of clothing and types of eyes. This is similar to the work of Golomb (1992) who also found few differences between a clinical and normal sample on such variables, though using different drawing tasks. On all of these measures, the drawings are almost identical across the three samples, which makes it not surprising that the judges could not differentiate them.

It is therefore difficult to determine whether similarity between the drawings or lack of ability in the judges is responsible for their failure on the discrimination tasks. It is unclear whether the judges would be successful if the drawings were more different. This also means that the experts may be able to perform better than the novices, if the drawings were different.

Although overall the judges scored no better than chance, there were some drawings that a significant number of judges agreed upon, though this may have been correctly or incorrectly. The individual HFD analysis revealed consistencies in the judges' decisions for each of the drawings in the three samples. Almost half of the clinical sample HFDs were identified correctly by a significant number of judges. However, over one quarter of them were identified incorrectly by more judges than would be expected by chance. The judges were consistent on fewer of the control drawings since only one third of the CA matched sample HFDs and just over one quarter of the MA matched HFDs were identified correctly. The judges were consistently incorrect on far fewer of the control HFDs, however, since only 2 of the CA matched control pictures and none of the MA matched sample HFDs were identified incorrectly by a significant number of judges. It appears, therefore, that the judges showed more agreement when identifying the clinical HFDs but they were also more likely to mis-classify these drawings and identify them as normal. The judges were more liable to identify a clinical sample drawings as normal than to identify a control sample HFD as drawn by a disturbed child.

Summary of Chapter 2

Inter-rater reliability for use of the Koppitz scoring system was found to be significantly high. The main aim of study 1 to replicate the results of Koppitz (1968) was not achieved. This study showed no differences between the emotional indicator scores on the HFDs of emotionally disturbed and normal children matched for chronological and mental age. This may have been for three possible reasons. First, this could mean that the indicators are not valid. This conclusion may be a little too hasty to reach at this point, however, and there are other alternatives to examine. It may have been because the clinical sample were not severely disturbed enough for the sensitivity of the indicators. It is not clear how similar to the Koppitz clinical sample these children were and because they were still in mainstream school, they may have been too 'normal'. Also, the indicators may not have discriminated the samples because the normative data upon which they are based has changed and the features they include are no longer valid indicators to use. These issues will be pursued in subsequent chapters.

The issue of whether a mental age delay on the part of the clinical sample can explain the differences found between the drawings cannot be answered from this chapter. The expected difference using the CA control failed to appear and it is therefore unclear whether the MA control removed those differences.

The intuitive method was also unsuccessful as the judges were unable to discriminate the drawings above chance levels. Again, there is more than one possible reason for this. It could mean that this method is also invalid. However, it may have been that there simply were not any differences to be seen, and were any differences present, such as in terms of emotional indicators, the judges would have seen them. The severity of the clinical sample's disturbance may have also affected this method as it affected the indicators. The fact that the children were all in mainstream schools may have meant that the clinical sample were not disturbed enough for it to show in their drawings. Copying of schemas and symbols in the development of children's drawings may mean that the drawings of clinical children who are surrounded by the artwork of normal children will become more like their peers.

The issue of whether expert judges are superior to novices is unresolved by these results, because both judges failed the tasks for reasons outlined above. It is unclear whether experience is necessary, or may improve the ability to identify disturbed children's drawings. It is necessary to evaluate whether there would be a difference between the types of judges if differences between the drawings are present.

CHAPTER THREE

A COMPARISON OF THE HUMAN FIGURE DRAWINGS OF CHILDREN WITH EMOTIONAL/BEHAVIOURAL DIFFICULTIES AND THEIR CHRONOLOGICAL AND MENTAL AGE MATCHED CONTROLS, USING A MORE SEVERELY DISTURBED CLINICAL SAMPLE

Introduction

This chapter aims to determine whether the failure of the two methods examined in chapter 2 to discriminate the drawings of the clinical and CA and MA matched control HFDs was due to a problem with the severity of the clinical sample's disturbance. The clinical sample children were not disturbed to an extent which necessitated removal from mainstream school and therefore the differences between their drawings and normally adjusted children's were either not present at all or were not obvious enough for the indicators or judges to be sensitive to them.

The question of whether the clinical sample children were suffering a delay in their drawings due to mental age, and the issue of whether expert or novice judges perform differently using the intuitive method was not answered in chapter 2 and will be examined again in this chapter. The retest reliability of the emotional indicators over a three week interval will also be assessed.

Study 3

Using the Koppitz (1968) emotional indicators with a more severely disturbed clinical sample

INTRODUCTION

The results of study 1 found no differences between the emotional indicator scores of HFDs from children with EBD and either their CA or MA matched controls. The CA matched control comparison was expected to have found a difference, as the design was a replication of the Koppitz (1968) study which found significantly more emotional indicators on the drawings of children with emotional disturbance

compared with a group of CA matched controls. The MA match control comparison ought not to have shown a difference if the drawing was seen as a reflection of the child's mental age. If a difference had occurred across the MA comparison, then there might still be some deviant characteristics of a child's drawing not explained by mental age delay. However, study 1 was unable to resolve this issue since there was no difference over either comparison.

There may have been a problem with the type of clinical sample used in study 1, who were children with EBD, but who were still in mainstream schools. These children may not have shown a severe enough disturbance for the drawing scores to be different from the normally adjusted sample. By replicating study 1 on a sample of more severely disturbed EBD children, it should be possible to resolve the issue of whether the sample used in study 1 was responsible for the results or whether it is a problem inherent in the clinical validity of the indicators.

This next study aimed to determine whether the differences Koppitz found can be seen in a sample of EBD children from special educational settings rather than mainstream school. As in the first study, the clinical children were compared with CA and MA matched controls considered normally adjusted. Using children from a special education setting, the aim was to show whether the nonsignificant results found previously were a function of the sample used or the lack of clinical validity of the emotional indicators. The participants in this study were all boys, since they represented the majority of children attending special primary schools for EBD.

METHOD

Participants

Clinical sample:

Forty four boys aged from 7;0 to 11;8 (mean CA 9;10, SD 1;1; mean MA 8;5, SD 1;5). These children had all been 'statemented' for special educational needs associated with emotional/behavioural difficulties (EBD), though no clinical diagnoses were available for these children. They were all full time pupils of special schools for children with EBD. See table 2 for numbers of children at each age.

Control sample (chronological age matched):
Forty four boys aged from 6;11 to 11;9 (mean CA 9;10, SD 1;2; mean MA 10;1, SD 1;7), matched for CA to within one month of a corresponding clinical sample child's birthday. These children were taken from mainstream schools and were considered normally adjusted as well as outstanding 'all round' by their teachers. See table 2 for numbers of children at each age.

Control sample (mental age matched):
Forty four boys aged from 6;1 to 11;0 (mean CA 8;9, SD 1;6; mean MA 8;5, SD 1;5), matched for mental age to a corresponding child in the clinical sample, using scores from 4 subtests of the WISC-III. These children were taken from mainstream schools and were considered normally adjusted as well as outstanding 'all round' by their teachers. See table 2 for numbers of children at each age.

Age at test	Clinical	CA control	MA control
6 years	0	1	7
7 years	3	2	2
8 years	7	7	15
9 years	11	11	10
10 years	17	15	8
11 years	6	8	2
Total	44	44	44

Table 12. Numbers of boys at each age in each of the three samples.

Design

A between-subjects design was used to compare participants' performance on the DAP test administered according to the Koppitz (1968) instructions and scored for emotional indicators according to the Koppitz scoring manual. EI scores from the clinical sample were compared with both a CA matched control sample and a MA matched control sample. Retest reliability was assessed over a three week interval.

The materials and procedure sections are the same as in study 1.

RESULTS

For a full breakdown of participants' indicator scores see appendix 3.

Retest reliability

27 boys with EBD and 26 normally adjusted boys matched for CA were given the DAP test and were retested after a three week interval. Pearson's correlations for the Koppitz emotional indicator scores were 0.5 ($p<0.05$) for the disturbed and 0.2 (ns) for the normally adjusted boys.

As in study 1, children in the three samples were categorised according to how many emotional indicators they produced in their drawing. No child produced more than 5 indicators out of the possible 30. The data are shown in table 13.

Number of EIs	Clinical	CA control	MA control
0-1	18	32	26
2+	26	12	18

Table 13. Numbers of boy's HFDs showing either 0 or 1 indicators and 2 or more indicators for the clinical and CA and MA control samples.

A Chi-Square analysis revealed a significant difference among the three samples ($X^2=9.5$, df=2, $p<0.01$). A 2x2 Chi-Square analysis showed a significant difference between the clinical and CA control samples ($X^2=7.83$, df=1, $p<0.01$). Though more boys in the clinical group than the MA group scored 2 or more indicators this difference did not reach statistical significance ($X^2=2.23$, df=1, ns). There was no significant difference between the CA control and MA control samples ($X^2=0.21$, df=1, ns).

The clinical sample had a mean indicator score of 2.14, compared with the CA matched control mean score of 1.09 and the MA matched sample mean of 1.48. A one way ANOVA showed a significant difference among the three sample scores ($F(2,129)=7.63$, $p<0.001$). The clinical sample showed significantly more indicators than both the

CA and MA matched control samples (Tukey's comparisons, ps<0.05) which were not significantly different from one another.

Summary of results

Retest reliability coefficicnts showed that the scores from disturbed boy's drawings were more reliable than from normally adjusted boy's HFDs, though both coefficients were low. More indicators appeared on the drawings of the disturbed boys than both their CA and MA matched normally adjusted controls, though there was some discrepancy between the Chi-Square and ANOVA results. There were no differences between the numbers of indicators appearing on the two control samples' HFDs.

DISCUSSION

The results of this study showed significant differences between the emotional indicator scores of the clinical sample and the CA and MA matched samples but no differences between the scores of the two control samples.

These results contrast with those of study 1 that found no differences among the groups. The difference between these results is most likely due to the sample differences and the severity of the children's disturbance. The clinical sample in the first study were children with EBD but who were not disturbed to the extent that they had to be removed from mainstream education. The clinical sample in this present study however were all children whose difficulties were severe enough to warrant their removal from mainstream education and placement in a special school for children with EBD. The severity of their disturbance seems the most likely explanation for the differences in the results of the two studies.

The significant difference between the scores of the clinical and CA matched samples was expected since this would be a replication of the results Koppitz found in 1968. The boys with EBD drew significantly more emotional indicators than normally adjusted boys of the same chronological age. This result however may have been due to the fact that the clinical sample children were suffering a developmental or mental age delay that would explain the differences in indicator

scores. For this reason, the MA matched control sample was used. The significant difference between the clinical and MA controls (on the ANOVA) is therefore an intriguing result. It suggests that the differences between the indicator scores of the clinical and control samples are not necessarily due to a developmental delay on the part of the children with EBD but are due to some deviant or atypical characteristic of their drawings.

This result confirms the clinical validity of the Koppitz emotional indicators since it shows that they can discriminate between a clinical and normal sample. However, since the absolute number of indicators that is seen in the drawings of disturbed children is very low, the diagnostic use of the indicators must be limited. Using the indicator cut-off score of 2 to screen the children, 18 out of 44 (41%) clinical sample children would have been missed, and 30 out of 88 (34%) normal children would be wrongly identified. This suggests caution in using the indicators to interpret children's HFDs though the use of alternative methods with more successful hit rates may be justified.

Retest reliability for the indicator scores was satisfactory for the disturbed boy's drawings, but was poor for the normally adjusted boy's HFDs using Pearson correlations. The mean scores for the boys retested were similar at the two points in time and the low number of indicators and small range of scores on the normally-adjusted boy's drawings may have been responsible for the low reliability coefficient for these drawings. The coefficients were low, however, and are indicative of the general variability in children drawings, especially when specific features are considered. These results might also relate to the idea that disturbed children's drawings are more symbolic of their self-concept than those of normal children (DiLeo, 1973).

The Koppitz emotional indicators were devised almost 30 years ago and in the USA so it is possible that some of the indicators may not be valid for today's children in the UK. The difference found here between the indicator scores of the boys with EBD and the normally-adjusted controls may be a function of the old normative data upon which Koppitz based the indicators. As mentioned in the discussion to study 1, the lack of a difference in indicator scores found there may also have been affected by the outdated norms in children's drawings,

and it is therefore important to check the frequency of emotional indicators on a sample of present day children's human figure drawings.

The question over the normative data will be assessed in chapter 4. Before this, the intuitive method of identification is investigated again, using the drawings collected for study 3. In study 2, it was found that the judges could not discriminate the clinical and control drawings using this method, but there were no differences in EI scores between the drawings either. Study 3 has found differences between the drawings in terms of indicator scores and it is therefore necessary to assess whether the judges can also identify the HFDs.

Study 4

Using the intuitive method of identification with a more severely disturbed clinical sample

INTRODUCTION

Study 2 using judges and the 'intuitive method of identification' to separate the clinical sample drawings from the control sample drawings was unable to show a difference between the expert and novice judges and also showed that the judges did not perform better than chance in their discriminations. This result reflected the emotional indicator analysis results of study 1 that showed no difference among the indicator scores of the samples and led to the conclusion that there were no differences among the sets of drawings.

Study 3, using more severely disturbed children for the clinical sample, did find differences between the clinical sample HFD emotional indicator scores and control samples' scores. It is important, therefore, to determine whether the differences between the drawings will affect a judge's ability to separate the samples using the 'intuitive method of identification'. This should determine whether the failure of the judges in study 2 was due to a lack of differences between the drawings, or an inability in the judges to perform this task successfully.

Again, experts were compared to novices for their ability to distinguish between the clinical and both control samples, in the two

format and two age conditions. This was to determine whether the similar performance of the judges in study 2 was due to the nature of the drawings, or the judges.

METHOD

Participants

Eight experts (people familiar with children's drawings) and thirty two novices (people unfamiliar with children's drawings) acted as judges. All participants were aged over 18 years. Six of the experts were academics who had undertaken research in children's drawings (three of whom had also taken part in study 2); two were trained primary school teachers. The novices were selected from the undergraduate student population at the University of York.

Design

Eight tasks were used, comprising two format conditions (group or pair) and two control conditions (CA or MA) as well as two age conditions (no ages and with ages). The between-subject variables were the format (group/pair) and control (CA/MA) conditions. The within-subject variable was the age condition. Judges each performed one task, both without and with the boy's ages.

Materials

Three sets of drawings collected during study 3 were used, one from the clinical sample of boys with EBD (n=44), one from the CA matched controls (n=44) and one from the MA matched controls (n=44). These were made into 4 tasks (2 group, 2 paired):

Group format- (1) randomly mixed group of clinical and CA matched control sample drawings (n=44 per sample)

(2) randomly mixed group of clinical and MA matched control sample drawings (n=44 per sample)

Pair format- (3) set of 44 pairs of drawings, one from the clinical sample paired with the appropriate CA matched control

(4) set of 44 pairs of drawings, one from the clinical sample paired with the appropriate MA matched control

A list of the drawing codes with the relevant age of the child was provided when the judge was performing the task in the 'with ages' condition.

Procedure

Judges were asked to discriminate the clinical sample HFDs from the control sample drawings. The instructions given to the judges were the same as in study 2, except for the fact that all of the pictures were drawn by boys. No criteria were given for making the discriminations. Judges performed the tasks without and then with the relevant ages of the boys.

RESULTS

Each participant (8 experts and 32 novices) performed the task of discriminating the clinical sample children's drawings from the control sample in one of the following conditions:

between-subject	-	Format (2)	group and pair
	-	Control (2)	CA and MA
within-subject	-	Age (2)	no ages and with ages

Experts vs. Novices

A test to assess the significance of the difference between 2 independent proportions was used to determine if the scores that the experts achieved were different from those of the novices. The proportion of correct scores was compared across the two sets of judges using the sum values of the judges' correct responses.

Total scores: The number of correct scores as a proportion of the possible total number correct was compared for the 2 experts and 8 novices in each condition. In the 'group format' conditions, the experts' total score was out of a possible 176 (2 x 88) correct responses; the novices' total score was out of a possible 704 (8 x 88) correct responses. In the 'pair format' conditions, the experts' total score was out of a possible 88 (2 x 44) correct responses; the novices' total score was out of a possible 352 (8 x 44) correct responses.

Condition	Experts	Novices
CA/group/no ages	119/176 (67.6%)	412/704 (58.5%)
CA/pair/no ages	60/88 (68%)	263/352 (74.7%)
MA/group/no ages	109/176 (61.9%)	418/704 (59.4%)
MA/pair/no ages	66/88 (75%)	234/352 (66.5%)
CA/group/with ages	129/176 (73.2%)	482/704 (68.5%)
CA/pair/with ages	65/88 (73.9%)	288/352 (81.8%)
MA/group/with ages	127/176 (72.2%)	482/704 (68.5%)
MA/pair/with ages	69/88 (78.4%)	263/352 (74.7%)

Table 14. The total scores obtained by the 2 experts and 8 novices in each condition, as a proportion of the total possible correct score.

The experts' proportions of correct responses did not differ significantly from the novices' in any of the conditions except for the 'CA control, group format, no ages' condition ($z=2.2$, $p<0.05$) but this result is excluded when the familywise error rate of $p<0.05/8=0.006$ is used.

Clinical scores: The number of correct clinical HFDs identified as a proportion of the number chosen as clinical by the judges was compared for the 2 experts and 8 novices in each condition. The denominators in each of the proportions differed according to the total number of drawings that were chosen as clinical HFDs by the judges (except for the 'pair format' conditions which was fixed by the number of pairs used).

Condition	Experts	Novices
CA/group/no ages	52/73 (71.2%)	187/314 (59.6%)
CA/pair/no ages	60/88 (68.2%)	263/352 (74.7%)
MA/group/no ages	53/84 (63.1%)	153/243 (63%)
MA/pair/no ages	66/88 (75%)	234/352 (66.5%)
CA/group/with ages	56/71 (78.8%)	226/324 (69.8%)
CA/pair/with ages	65/88 (73.9%)	288/352 (81.8%)
MA/group/with ages	77/115 (67%)	208/285 (73%)
MA/pair/with ages	69/88 (78.4%)	263/352 (74.7%)

Table 15. The clinical scores obtained by the 2 experts and 8 novices in each condition, as a proportion of the total number of drawings identified as clinical by the judges.

The experts' proportions of correct responses did not differ significantly from the novices' in any of the conditions except for the 'CA control, group format, no ages' condition ($z=1.85$, $p<0.05$) but again, this result is excluded when the familywise error rate of $p<0.05/8=0.006$ is used.

Control scores: The number of correct control HFDs identified as a proportion of the number chosen as control by the judges was compared for the 2 experts and 8 novices in each condition. The denominators in each of the proportions differed as with the clinical scores.

Condition	Experts	Novices
CA/group/no ages	67/115 (58.3%)	225/390 (57.7%)
CA/pair/no ages	60/88 (68.2%)	263/352 (74.7%)
MA/group/no ages	57/92 (62%)	264/463 (57%)
MA/pair/no ages	66/88 (75%)	234/352 (66.5%)
CA/group/with ages	73/105 (69.5%)	246/380 (64.7%)
CA/pair/with ages	65/88 (73.9%)	288/352 (81.8%)
MA/group/with ages	50/61 (82%)	274/419 (65.4%)
MA/pair/with ages	69/88 (78.4%)	263/352 (74.7%)

Table 16. The control scores obtained by the 2 experts and 8 novices in each condition, as a proportion of the total number of drawings identified as control by the judges.

The experts' proportions of correct responses did not differ significantly from the novices' in any of the conditions except for the 'MA control, group format, with ages' condition where the experts' proportion of correct scores (50/61) differed significantly from the novices' (274/419) ($z=2.58$, $p<0.01$).

Given that in the majority of cases, the experts' scores did not differ from the novices', and there were equal numbers of participants in each condition, further analysis has collapsed the data across the two sets of participants.

Analysis of the different conditions

The analyses then looked at the numbers of judges performing at and above chance levels in each of the conditions.

Condition	Total score	Clinical/x	Control/x
CA/group/no ages	2	1	0
CA/pair/no ages	5	5	5
MA/group/no ages	3	2	0
MA/pair/no ages	3	3	3
CA/group/with ages	7	3	3
CA/pair/with ages	10	10	10
MA/group/with ages	8	3	2
MA/pair/with ages	8	8	8

Table 17. Numbers of judges out of 10 performing significantly better than chance (using the family-wise error rate of $p<0.05/10=0.005$) for each condition, using the three different scores.

The above table shows that more of the judges performed better than chance in the 'with ages' conditions, and that more judges performed better than chance for the clinical drawings than the control drawings, in the 'group format' conditions. Between 20% and 100% of the judges performed better than chance overall.

The scores were then converted to percentages in order to allow comparisons easily across score types by controlling for the differing denominators in the clinical and control scores, as well as the group and pair tasks.

Condition	Total %	Clinical %	Control %
CA/group/no ages	60	62	59
CA/pair/no ages	73	73	73
MA/group/no ages	60	65	58
MA/pair/no ages	68	68	68
CA/group/with ages	69	72	67
CA/pair/with ages	80	80	80
MA/group/with ages	70	72	63
MA/pair/with ages	76	76	76

Table 18. The average percentage score in each of the conditions.

The above table shows a great similarity among the three score types in the 'group format' conditions. The average clinical scores were consistently higher than the average control scores. Also, the 'pair format' scores were usually higher than the 'group format' scores, and the 'with ages' condition scores were usually higher than the 'no ages' condition scores.

Repeated measures ANOVAs were performed on the percentage scores for the 10 judges in each condition. The between-subject variables showed a main effect of format for the total scores ($F(1,36)=36$, $p<0.001$), clinical proportion scores ($F(1,36)=12.85$, $p<0.001$) and control proportion scores($F(1,36)=34.33$, $p<0.001$). The within-subject variable showed a main effect of age for the total scores ($F(1,36)=68.2$, $p<0.001$), clinical proportion scores ($F(1,36)=55.98$, $p<0.001$) and control proportion scores ($F(1,36)=38.9$, $p<0.001$).

Individual HFD Analysis

As in study 2, the numbers of judges correctly and incorrectly identifying each of the HFDs was calculated in order to determine whether there were any consistencies in their decisions. Since there was no difference between the experts' and novices' performance, the data were collapsed across the two different types of judge.

Clinical HFDs

Since the repeated measures ANOVAs reported above showed no effect of control sample, the data for the clinical HFDs were collapsed across this factor. The data then consisted of numbers of judges correctly and incorrectly identifying each of the 44 clinical HFDs in the 'group format' and 'pair format', 'no ages' and 'with ages' conditions.

Binomial tests comparing the total number of correct judgements with the total number of incorrect judgements for the 44 clinical HFDs showed significantly more correct judgements than incorrect judgements being made in the 'pair format, no ages', 'group format, with ages' and 'pair format, with ages' conditions ($ps<0.001$) but not in the 'group format, no ages' condition.

Mann-Whitney tests comparing the numbers of correct judges with the numbers of incorrect judges for each of the 44 HFDs showed

significantly more judges were correct than incorrect for the 'pair format, no ages' ($z=-6.63$, $p<0.001$), 'group format, with ages' ($z=-4.57$, $p<0.001$) and 'pair format, with ages' ($z=-7.22$, $p<0.001$) conditions but not the 'group format, no ages' condition ($z=-0.51$, ns).

Where 17 or more out of the 20 judges in each condition identified a clinical HFD either correctly or incorrectly, this is significantly above the number to be expected by chance using the familywise error rate of $p<0.05/44=0.001$. Only 20% of the drawings were identified by more judges than chance in the 'group format, no ages' condition. 61% of the HFDs were identified by a significant number of judges in the 'pair format, with ages' condition.

Four of the drawings (clinical HFD (C) numbers 7, 8, 18 and 29) were correctly picked by significantly more judges than chance in all the conditions; figures 1-4 show scaled versions of these drawings. C7 (Fig. 1) was drawn by an 8 year old boy. The drawing scores 3 indicators (big figure, teeth, hands cut off). C8 (Fig. 2) was drawn by an 8 year old boy. The drawing scores 1 indicator (hands cut off). C18 (Fig. 3) was drawn by a 9 year old boy. The drawing scores 5 indicators (poor integration, tiny head, no eyes, no nose, no mouth). C29 (Fig. 4) was drawn by a 10 year old boy. The drawing scores 2 indicators (shading body/limbs, shading hands).

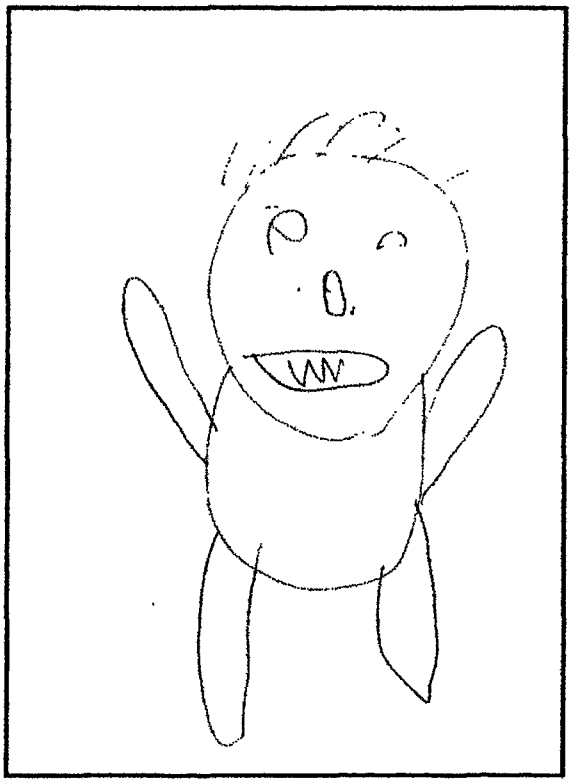

Figure 1. HFD C7, age 8;9

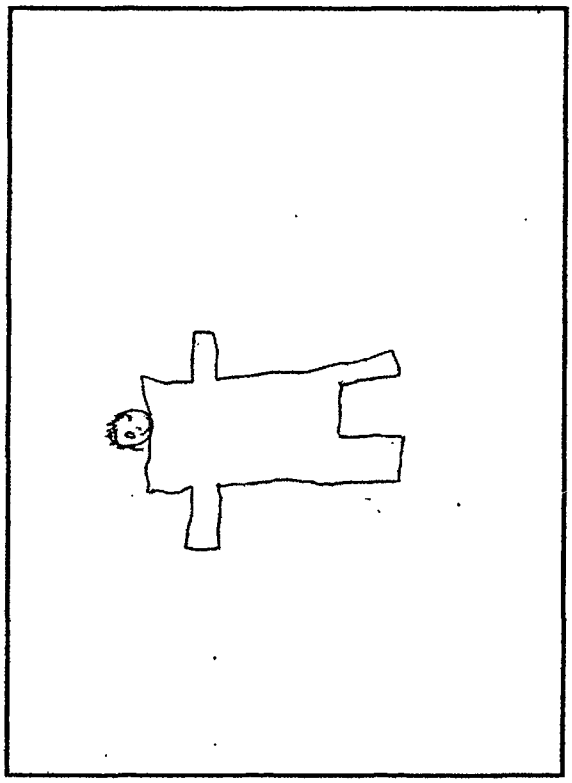

Figure 2. HFD C8, age 8;10

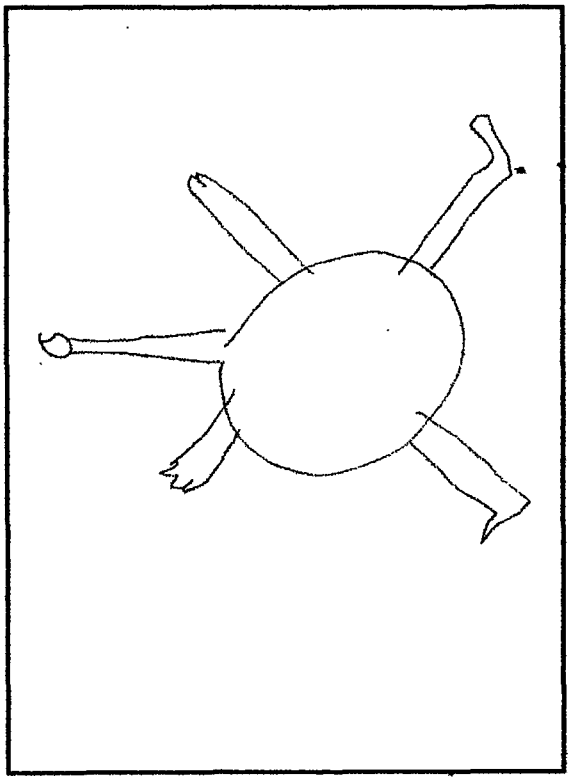

Figure 3. HFD C18, age 9;7

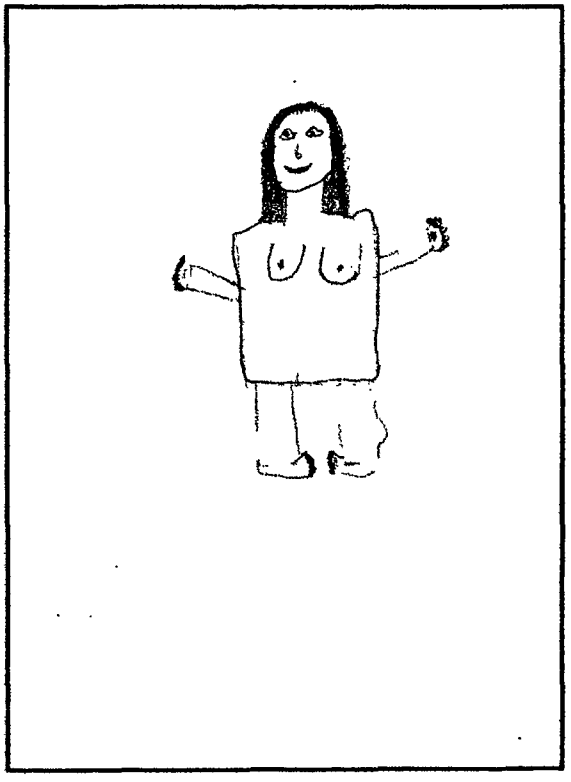

Figure 4. HFD C29, age 10;5

CA matched HFDs

The numbers of judges correctly and incorrectly identifying each of the 44 CA matched control HFDs in the 'group format' and 'pair format', 'no ages' and 'with ages' conditions were also calculated.

Binomial tests comparing the total number of correct judgements with the total number of incorrect judgements for the 44 CA matched control HFDs in each of these conditions showed significantly more correct judgements than incorrect judgements being made in all the conditions ($p<0.001$).

Mann-Whitney tests comparing the numbers of correct judges with the numbers of incorrect judges for each of the 44 HFDs showed significantly more judges were correct than incorrect for the 'group format, no ages' ($z=-4.42$, $p<0.001$), 'pair format, no ages' ($z=-6.91$, $p<0.001$), 'group format, with ages' ($z=-6.37$, $p<0.001$) and 'pair format, with ages' ($z=-7.53$, $p<0.001$) conditions.

Where 10 out of the 10 judges in each condition identified the CA matched HFD either correctly or incorrectly, this is significantly above the number to be expected by chance using the familywise error rate of $p<0.05/44=0.001$. Two of these control sample HFDs (CA matched HFD (CA) numbers 2 and 23) were identified correctly by a significant number of judges in all the conditions. See figures 5 and 6 for scaled versions of these drawings. CA2 (Fig. 5) was drawn by a 7 year old boy. The drawing scores no indicators. CA23 (Fig. 6) was drawn by a 10 year old boy. The drawing scores 1 indicator (short arms).

MA matched HFDs

The numbers of judges correctly and incorrectly identifying each of the 44 MA matched control HFDs in the 'group format' and 'pair format', 'no ages' and 'with ages' conditions were also calculated. Binomial tests on the total number of correct judgements compared with the total number of incorrect judgements for the 44 MA matched control HFDs in each of these conditions showed significantly more correct judgements than incorrect judgements being made in all the conditions ($p<0.001$).

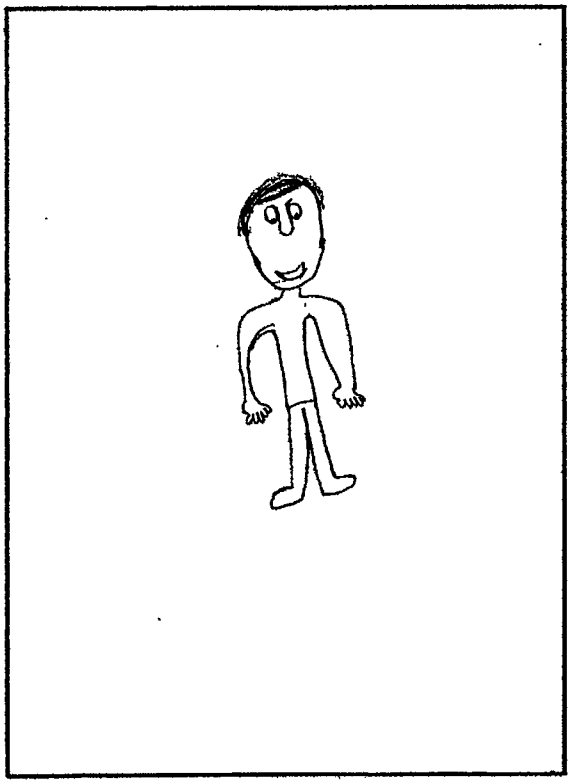

Figure 5. HFD CA2, age 7;5

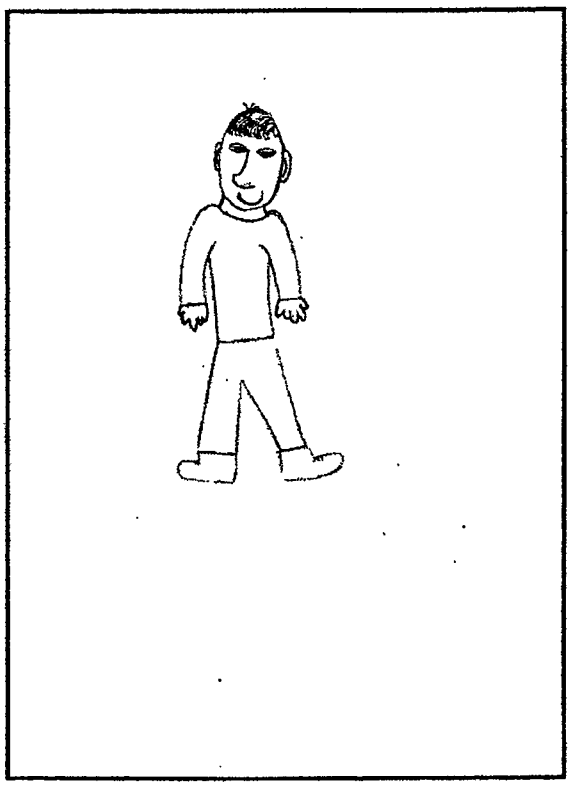

Figure 6. HFD CA23, age 10;3

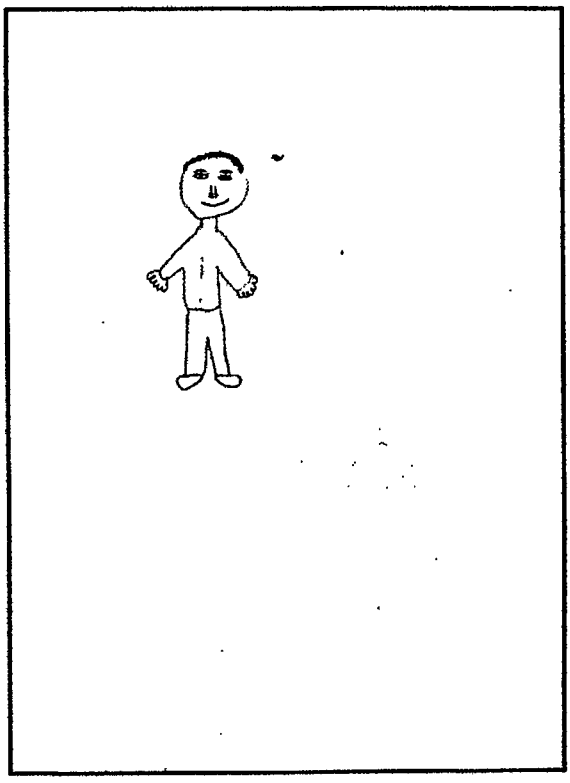

Figure 7. HFD MA33, age 9;2

Mann-Whitney tests comparing the numbers of correct judges with the numbers of incorrect judges for each of the 44 HFDs showed significantly more judges were correct than incorrect for the 'group format, no ages' ($z=-6.55$, $p<0.001$), 'pair format, no ages' ($z=-4.99$, $p<0.001$), 'group format, with ages' ($z=-6.28$ $p<0.001$) and 'pair format, with ages' ($z=-6.12$, $p<0.001$) conditions.

The number of judges required for a significant binomial test in the MA matched sample were the same as the CA matched sample. One drawing (MA matched HFD (MA) number 33) was identified by a significant number of judges in all the conditions. See figure 7 for a scaled version of this drawing. MA33 (Fig. 7) was drawn by a 9 year old boy. The drawing scores 1 indicator (short arms).

Summary of results

There was no significant difference found between the experts' performance on these tasks and the novices performance. Judges were able to perform the tasks successfully and discriminated the clinical sample HFDs from the control drawings significantly above chance levels. Their performance was affected differentially by the different format and age conditions, with the 'pair format' and 'with ages' conditions achieving most success, but was not affected by which control (CA or MA) was used.

The individual HFD analysis showed no differences between the experts and the novices. Significantly more judges identified the three sample drawings correctly in some of the conditions, with the 'pair format' and 'with ages' conditions more successful. Some HFDs were consistently picked correctly or incorrectly by significantly more judges than would be expected by chance.

DISCUSSION

The results of this study found that using the intuitive method, the judges were able to discriminate the clinical sample HFDs from the control sample drawings above chance level, though their performance was affected by the various conditions which were used. This means that there is a visible difference between the drawings which the judges were able to detect. The failure of the judges in study 2, therefore,

must have been due to a lack of differences between the drawings, rather than a lack of ability on the part of the judges. It is less clear what the judges were using to make their discriminations, though it may be related to the emotional indicator scores since the judges are only successful where differences in EI scores have been found.

Identification of the clinical sample drawings was not differentially affected by whether the drawing was grouped or paired with its CA or MA matched control. This result reinforces the Koppitz indicator analysis pattern of results where the two control samples had comparatively the same numbers of indicators, and were not different from each other in that respect. The clinical sample drawings looked equally different from their MA matched counterparts as from their CA matched counterparts. The visual appearance of the clinical sample HFDs, therefore, is not simply the same as those drawn by children of the same mental age.

The judges improved in the 'pair format' conditions and the 'with ages' conditions. In the 'group format, no ages conditions', at most only one quarter of the judges were able to discriminate the clinical HFDs from the controls above chance levels and the judges agreed correctly on only 20% of the HFDs. This rose to 90% of the judges in the 'pair format, with ages' conditions able to discriminate the drawings above chance, and 61% of the drawings being consistently and correctly identified as clinical. It is possible that the improvement in the 'with ages' condition was simply due to practice as the judges repeated the task. However, it was impossible to counterbalance the order of presentation of the tasks, as the judges could have recalled the children's ages between the conditions. It can only be assumed that the practice effect did not affect the results to a significant extent. This result has implications for the practical usefulness of interpreting an HFD since success increases as a function of the conditions in which the drawing is presented. Without the age of the child and mixed in with lots of other drawings, identification of a clinical child's picture is difficult, if not impossible, whereas when paired with a normal child's drawing and given the ages of the children, it is relatively easy to identify the clinical HFD.

Many of the drawings were correctly chosen consistently by more judges than would have been expected by chance. Those four clinical pictures which were consistently identified correctly as clinical in all conditions were very different from each other. The control drawings which were consistently identified as from normally adjusted children in all the conditions were all typical canonical representations of the human figure.

As with study 2, the results of this investigation have shown no differences between the experts and novices ability to discriminate the clinical sample HFDs from their CA and MA matched control drawings. This has been shown in 'pair format' tasks and 'without ages' too, conditions which the expert judges did not complete in study 2. The experts who were able to discriminate the drawings successfully were not acting in any special way, since the novices performed in a similar fashion. The lack of a difference between the experts and novices supports the previous studies which showed no differences between experts and novices (e.g. Hiler & Nesvig, 1965; Stricker, 1967; Ulman & Levy, 1973; Arkell, 1976) This also supports the opinions of Motta et al. (1993) but is far from reinforcing their conclusion from this that the DAP test is invalid, since the judges (experts and novices) were *both* able to discriminate the clinical sample drawings successfully.

Summary of Chapter 3

This chapter has found that the Koppitz indicators occur more often on the drawings of severely disturbed boys when compared with both CA and MA matched boy's drawings. This suggests that the Koppitz indicators are valid for discriminating clinical from normal samples. The difference compared to CA controls was expected whether the clinical children's drawings were delayed or deviant. The fact that the differences existed when MA was controlled suggests that there is something deviant in the clinical children's drawings and the differences are not just due to a developmental delay on the part of the clinical children. This also has implications for the influence of cognitive factors on children's drawings and particularly the DAP test. Cognitive ability as measured by the WISC can be seen to have a

negligible effect since the same results are found when this is controlled, as with the CA matched control.

The indicators occur in very small numbers, however, limiting their practical significance. Though a statistically significant difference may have been found, it is questionable whether the difference is clinically significant. Using a cut-off score of 2 indicators, since Koppitz thought one was not significant alone, many clinical sample children would not be identified using the indicators. The average number occurring in the known clinical group was only around 2 which is very low. This has implications for the clinical use of the DAP test and the practical use of the Koppitz indicators.

Using the intuitive method, it was found that the judges were able to discriminate the severely disturbed HFDs above chance levels. Judges were successful when the clinical drawings were compared both with the CA and MA matched HFDs. The judges were more successful in the pair conditions and with the boy's ages available. This means that the failure of the judges in chapter 2 was not because the method was invalid, but because there were no differences between the drawings to be seen. In this present chapter, where a difference exists in terms of the indicator scores, the judges were also successful, suggesting that the drawings of the disturbed and normal children are different.

The difference is easier to see when the boy's ages are taken into account which suggests that the drawings are delayed because without the boy's ages, the drawings may be mistaken for younger, normally-adjusted children's and the discriminations become closer to chance. The improvement in the 'pair format' task is because the judges have a direct comparison to use, knowing one of the drawings to be normal. These results have implications for the clinical use of the DAP test since different contexts can create different opinions.

Experts did not differ from novices in terms of the proportions of drawings they could correctly identify. This means that experience with children's drawings is not necessary for this task and to discriminate the HFDs of normal and disturbed boys. Evidence (e.g. Hiler & Nesvig, 1965; Ulman & Levy, 1973; Arkell, 1976) has shown that experts do not differ from novices but some believe (e.g. Wanderer, 1969; Motta et al.,

1993) that the intuitive method is not valid, since expert judges have been found unable to discriminate HFDs. The experts in this chapter may have differed from those cited in the literature, but the method has been found to be valid, since both expert and novice judges can discriminate the clinical HFDs.

The success of the judges in the tasks indicates that there is a visual difference between the drawings that helped them to identify the samples. This difference is seen between clinical and both the CA and MA matched drawings. This difference may be linked to the emotional indicators, since the judges' success only appears where there are significant differences on the indicators as well. Some drawings were also found which were identified both incorrectly and correctly by a significant number of the judges. This suggests that the judges were all identifying similar aspects of these pictures. What the visual difference between the disturbed and normal drawings involves and whether it is tied in with the emotional indicators is an issue which needs to be resolved.

This chapter has examined only one of the three criteria that Koppitz used to validate the emotional indicators - that the features must discriminate between the drawings of clinical and normal samples. The other two criteria of occurring rarely in normal drawings and not increasing with age must also be evaluated in order to form any firm conclusions about the validity of the indicators.

CHAPTER FOUR

ASSESSMENT OF THE NORMATIVE DATA

Introduction

At the end of chapter 2, it was noted that the indicators may have failed to discriminate the clinical and control samples due to a change in the normative data upon which they are based. Though differences have been found subsequently, in chapter 3, using the more severely disturbed sample which may have been the only reason why study 1 failed, it is still important to investigate the issue of the normative data. It may be that both the lack of severity of the sample and the normative data combined to create the nonsignificant differences seen in chapter 2.

Also, at the end of chapter 3, it was noted that only one of Koppitz's criteria has been examined. In order to make any firm conclusions regarding the validity of the indicators, all three of the criteria should be investigated. It has already been found in chapter 3 that the indicators can discriminate between clinical and normal samples, so it is important to determine whether the features are still rare and do not increase with age in normal children's drawings. These two criteria are examined in this chapter.

The issue of cultural influences on children's drawings explores the idea that the symbols and schemas may change over time and between places, with copying as a probable method of transmission. Several researchers (e.g. Goodenough, 1926; Harris, 1963; Mortensen, 1984) have traced the changes in children's drawings of the human figure. These three sets of data have considered the HFD and the features included or excluded by children at different points in time, but only with reference to the developmental items such as those found on the GH scale. The DAP:SPED (Naglieri et al., 1991) includes recent norms for features which are similar to the Koppitz indicators, but these were also American norms and there has been no British consideration of these specific features. There may be differences between the norms that Koppitz collected in USA in 1968 and what British children draw in the 1990s in terms of the significant features

for assessing emotional disturbance. This chapter aims to evaluate the normative data upon which the indicators are based, to determine any changes which may have taken place and any implications which such changes might have on the results presented so far in this thesis.

Study 5a

Collection of the normative data

INTRODUCTION

The collection of large numbers of drawings in an effort to establish age related norms began in the relatively early stages of research into children's drawings when the HFD was considered to be of use as a measure of intelligence (Goodenough, 1926). Subsequent work re-standardised the Goodenough data (Harris, 1963), reviewing and updating the occurrence of features on children's HFDs.

Koppitz (1968) collected 1856 HFDs from children in schools in America, using group administration of the Draw-A-Person test. These drawings were then analysed for the percentages of developmental items occurring for the two genders at each age level. The developmental items included features such as head, arms, legs, facial features and clothing. These items were defined as increasing in frequency of occurrence, not being affected by the drawing medium or instructions given nor by the effect of learning or maturation. This analysis led to a measurement of mental maturity using those items considered 'Expected' which occurred on at least 85% of children's HFDs and items considered 'Exceptional' which occurred on 15% or less of normal children's HFDs.

The drawings that were collected were also used in the development of the emotional indicators. The drawings were scored for various items and the percentages of their occurrence at each age level for each gender was calculated. The items which were used however in this analysis were those which were considered to have clinical significance by those clinicians who used the HFD as a projective technique. The original 38 items which Koppitz selected were based on Machover's (1949) and Hammer's (1958) work and Koppitz's own

clinical experience. They fell into three categories: those which were related to the quality of the HFD, those special features not usually found on HFDs and, lastly, omissions of items usually expected on children's HFDs at given age levels. These last items came from the list of 'Expected' items from Koppitz's Developmental items.

In order to determine which items would be most useful as emotional indicators, Koppitz produced three criteria that any feature within a child's drawing must satisfy before it could become an emotional indicator. The criteria were:

1. It must have clinical validity, i.e. it must be able to differentiate between HFDs of children with and without emotional problems.
2. It must be unusual and occur infrequently on the HFDs of normal children who are not psychiatric patients, i.e. the sign must be present on 15% or less of the HFDs of children at a given age level.
3. It must not be related to age and maturation, i.e. its frequency of occurrence on HFDs must not increase solely on the basis of the children's increase in age.

Koppitz scored all 1856 drawings for the original 38 emotional indicators, working out the percentages of each item's occurrence for each gender at each age level. Preliminary results removed some items from the list due to their occurring so often as not to be considered rare (e.g. vacant eyes, big head) or increasing in frequency of occurrence as the children grew older (e.g. broken lines, hidden hands). Some items were given ages at which they became significantly rare, having been common at previous age levels. This preliminary analysis left 32 items that passed the second two criteria and underwent the validation study to determine which fulfilled the first criterion.

The validation study which Koppitz then carried out reduced the indicators to 30 since 'figure cut off by paper' and 'sun' did not discriminate between the clinical and normal samples (for a full list of the emotional indicators, see appendix 1). These emotional indicators were the ones used in all subsequent investigations.

Normative data such as those collected by Koppitz reflect the trends in children's drawings that are due to cultural influences as well as maturational processes. The drawings of younger children are thought to be affected more by maturational processes whereas the

older child's drawing may be affected more by cultural influences such as teaching methods and the media (Mortensen, 1984). Changes in drawing trends over time have been noted (e.g. Wilson & Ligtvoet, 1992).

Harris (1963) re-evaluated the normative data of Goodenough (1926) and found some differences in the two standardisation samples' drawings which he labelled 'secular trends', using these trends to hypothesise the influence of culture on children's drawings. Harris found that the core features of the human figure had not changed since the Goodenough data, though the children in his study showed improvements in body and limb detail such as hair, controlling transparencies and the drawing of hands. Harris explained the changes which had occurred in terms of the improved health education in schools where the body and its development and comfort had come to reflect a greater 'body acceptability' than had been evident in Goodenough's era (Harris, 1963). Changes in the art education of children in the 1950s, from stylised copying and stereotyped models to an emphasis on freedom of expression, as well as changes in child development theory and parenting in the 1950s where the child is given greater freedom than in the 1920s, were also factors which Harris suggested may have helped explain the changes that occurred in the drawings.

Mortensen (1984) collected and described the HFDs of 10 children of each gender at each age level between the ages of 5 and 13 years. Each child drew a man, woman and self drawing. She compared her results with those of Harris (1963) and Goodenough (1926) and found that there was a "surprising overall similarity between the results" (Mortensen, 1991, p. 144) though more of her results were closer to those of Harris than Goodenough. In general, the results were very similar for most drawing features, with almost complete coincidence among all three studies for presence of head, eyes, nose, mouth, legs, trunk, clothing and profile. Major differences were found for depiction of pupils and eye proportions, hair, outline of head and trunk and facial features. For these variables, the Mortensen data showed higher percentages of occurrence across the age range. The earlier data of Goodenough and Harris showed higher percentages for

presence, proportions and position of ears, correct number of fingers and proportion of feet. The ear data could be explained by the fact that the Mortensen data included more female drawings than the Goodenough and Harris data which were based only on drawings of a man. Male figures are more commonly drawn with ears than the female figures (Mortensen, 1984). Though her data were not as representative as the Goodenough or Harris results, Mortensen took the similarities among the three studies as a sign of the generality of children's drawing development in Western culture.

As well as the norms changing over time, trends in children's drawings are different according to local and regional influences. Certain aspects of children's drawings are conventional in the sense that they reflect the current fashion of a specific culture. Cross cultural studies have shown that where drawing and representational art are not regularly practised activities, the results of the Draw-A-Man test show primitive HFDs (Cox & Bayraktar, 1989). Where drawing is practised, however, the resultant HFDs do not necessarily look like the Western ideal (Paget, 1932; Reuning & Wortley, 1973, cited in Cox, 1993).

How drawing trends are established and transferred between populations was noted by Wilson and Wilson (1985). They commented on a trend of drawing profile figures with both arms extending forwards from the back of the figure, occurring in the drawings of children in Los Angeles in the 1920s. This convention had been noted previously in the drawings of Italian children by Ricci in the 1880s which suggested that the style may have travelled with the Italian immigrants to America then been passed on to the other children. This research would seem to suggest that the mode of transportation of the various conventions in children's drawings is varied and unusual, and this also means that the drawings of any given group may not be representative of a population in general.

The present investigation aimed to re-evaluate the 30 Koppitz emotional indicators in terms of the normative data upon which they are based. Previous norms have been based on American (Goodenough, 1926; Harris, 1963) and Scandinavian drawings (Mortensen, 1984) but no norms appear to exist for British culture. The similarity between the

Mortensen and Harris data may reflect the generality of drawing development for those features the two researchers considered, but in the context of this thesis, the indicators involved are not necessarily part of the normal development of drawing. The normative data collected by Koppitz is relevant to these features, but these data are almost 30 years old and from a sample of American children.

A sample of children from the north of England was used to determine any changes that there may be between the 1968 American Koppitz normal sample and a current sample of British normally adjusted children. The sample used in this present study was not stratified and fully representative of any general population of children and does not give age-related norms. These data may be useful, however, in creating a background against which the results of study 1 and 3 can be presented. The differences that may occur between the normative data of Koppitz and the present investigation may help to explain why the normally adjusted children in the first study drew as many indicators in their HFDs as the clinical population. It will also establish whether the differences that were found in study 3 remain with updated indicators.

METHOD

Participants

1598 children from mainstream primary schools in the north of England took part. There were 818 boys and 780 girls. Exact numbers of each gender at each age level are given in the following table:

Age (years)	5	6	7	8	9	10	11
Boys	220	109	136	94	113	90	56
Girls	223	96	124	110	105	69	53

Table 19. Numbers of children's HFDs for each gender at each age level between 5 and 11 years inclusive.

Design

Group administration of the DAP test was given for all children. Drawings were scored for all EIs without using the age restrictions given by Koppitz (1968). Inter rater reliability was assessed using one

other person, trained in the Koppitz scoring system who also scored 10% of the drawings.

Materials and Procedure

The drawings were obtained using the same materials and procedure as in studies 1 and 3. Drawings were scored for all emotional indicators in the Koppitz scoring manual, regardless of age restrictions.

RESULTS

Inter-rater reliability The results showed considerable agreement between the two raters for the 160 drawings scored by both. Pearson's correlation between the scores given by the two raters was 0.86. One danger with using Pearson's Product Moment Correlation for inter-rater reliability is that one rater might have consistently higher or lower scores that would not be reflected in the correlation coefficient. An alternative is to use intra-class correlations, but inspection of the data indicated that this was not a problem.

The percentages of children producing the Koppitz emotional indicators in their HFDs at each age level was calculated in order to determine if any changes had occurred since the Koppitz normative data were collected. In order for an indicator to be valid, it must adhere to the two criteria stated by Koppitz. First, it must be rare and occur on 15% or less of children's drawings. Second, it must be unrelated to age and maturation and not increase with age. For the raw percentage data, see appendix 4.

For the Koppitz normative data drawings, most of the items occur rarely and do not increase in occurrence with age. Some items are found often on the younger children's drawings and are not significant at these ages such as poor integration on a 5-year-old's drawing or shading of body or limbs on a 7-year-old boy's drawing. The Koppitz data for girls are very similar to those of the boys except for the ages at which certain items become significant. For example, shading of body or limbs is significant for girls at a younger age (7 years) and no arms and no nose are also significant at a younger age that the boys' drawings.

The data from the present study show that most of the indicators still appear rarely on the normative population HFDs. A comparison of the Koppitz data with the results of this study revealed some items unchanged, some changed in the ages at which they become significant, and some items no longer fulfilling the criteria at all. The results are shown in the following table.

	Koppitz		New data	
Emotional Indicators	Boys	Girls	Boys	Girls
Poor integration of parts	7	6	6	5
Shading of face	5	5	No longer valid	
Shading of body/limbs	8	7	No longer valid	
Shading of hands/neck	7	7	5	5
Gross asymmetry of limbs	5	5	5	5
Slanting figures	5	5	5	5
Tiny figures	5	5	7	8
Big figure	8	8	No longer valid	
Transparencies	5	5	5	5
Tiny head	5	5	5	5
Crossed eyes	5	5	5	5
Teeth	5	5	Invalid	5
Short arms	5	5	No longer valid	
Long arms	5	5	5	5
Arms clinging to body	5	5	5	5
Big hands	5	5	5	5
Hands cut off	5	5	6	7
Legs pressed together	5	5	No longer valid	
Genitals	5	5	5	5
Monster/grotesque figure	5	5	5	5
Three or more figures drawn	5	5	5	5
Clouds	5	5	5	5
No eyes	5	5	5	5
No nose	6	5	8	6
No mouth	5	5	5	5
No body	5	5	5	5
No arms	6	5	6	5
No legs	5	5	5	5
No feet	9	7	6	7
No neck	10	9	Invalid	11

Table 20. Minimum ages for scoring the thirty emotional indicators using Koppitz's (1968) norms and those of a present-day UK sample.

Seventeen of the original thirty items had not changed at all, but these tend to be the indicators that are the most unusual and do not occur very often even on disturbed children's drawings. It must be noted that 2 items (teeth and no feet) changed only for boys.

Seven of the items had changed age restrictions only and were still the same as the Koppitz data after the specific age. Three of the changes brought the ages at which the indicators become significant forward (poor integration, shading hands/neck and no feet), four of the changes made the age of significance later (tiny figure, hands cut off, no nose and no neck).

Seven items did not conform to the original Koppitz criteria or were found to have unusual distributions over the age range making their practical usefulness doubtful. These included the following:

1. Shading of face. This item was not consistently below 15% in boys. It increased in occurrence at 8 and 9 then decreased over age 10 and 11 but stayed quite high (13%). In girls this item consistently increased over the ages showing a relation to age or maturation.
2. Shading of body/limbs. This item was only below 15% at age 5 years for boys then showed a general increase in occurrence related to age or maturation. In girls, this item was above 15% at 6, 9 and 10 years and below 15% in between showing an alternating pattern of great irregularity.
3. Big figure. In both genders, this item showed an increase in occurrence related to age or maturation.
4. Teeth. For boys only, this indicator showed a steady increase in occurrence related to age or maturation.
5. Short arms. This item was consistently above 15% on girl's drawings and was only below 15% on 5 year old boys drawings.
6. Legs together. On boys' drawings, though rare, this item showed a steady increase in occurrence. In girls this item showed a disturbed distribution with an unusual increase at age 9 though was below 15% at other ages.
7. No neck. Though decreasing with age, this item was above 15% at all ages for boy's drawings.

The shading data collected may be different from those of Koppitz due to ambiguity in the criteria for scoring these items. The scoring in

this study may have been over-inclusive of types of shading not intended by Koppitz to be an emotional indicator. Also, the poor integration item is very subjectively scored which makes it difficult to know whether any differences found were due to real changes in the occurrence of the indicator, or a function of the scoring method. These two items were amongst those which were problematic for the inter-rater reliability of this study, but the high correlation coefficient between the two raters overall shows that considerable agreement is found for the indicators overall.

The differences between boys and girls in terms of the ages at which the indicators became significant according to Koppitz were not the same in this study for any of the indicators apart from 'no arms'. Poor integration and shading of hands/neck were now significant at the same age for both genders rather than at a later age for the boys than the girls as Koppitz had found. The age difference for no nose had widened as the boys were not expected to produce a nose now until two years later than Koppitz found but the girls were expected to produce one only a year later than Koppitz had found. Feet were now expected earlier on the boys' drawings than Koppitz had found but the girls' age of significance had stayed the same. No neck had also switched as the boys no longer were expected to produce a neck by 11 years whereas the girls were expected to at that age.

Summary of results

The results show that the normative data upon which Koppitz based her emotional indicators have changed. Various indicators can no longer be considered valid and some had to have the ages at which they may be considered valid changed. From the results, a new, revised list of indicators can be compiled.

List of revised indicators

1. Poor integration of parts (boys 6; girls 5)
2. Shading of hands and/or neck (boys & girls 5)
3. Gross asymmetry of limbs
4. Slanting figures
5. Tiny figure (boys 7; girls 8)

6. Transparencies
7. Tiny head
8. Crossed eyes
9. Teeth (no longer valid for boys)
10. Long arms
11. Arms clinging to body
12. Big hands
13. Hands cut off (boys 6; girls 7)
14. Genitals
15. Monster or grotesque figure
16. Three or more figures spontaneously drawn
17. Clouds
18. No eyes
19. No nose (boys 8; girls 6)
20. No mouth
21. No body
22. No arms
23. No legs
24. No feet (boys 6; girls 7)
25. No neck (no longer valid for boys; girls 11)

DISCUSSION

The results of this investigation show that the normative data upon which Koppitz based her emotional indicators has changed. For just over half of the indicators, no differences were found between the Koppitz data and the results of this study. For a considerable number, however, changes in the ages at which the indicators became rare enough to attain clinical significance altered. For a few indicators, their occurrence was now found to be related to age and maturation, showing an increase with age, or was found on a consistently high number of children's drawings so that it was unable to be considered clinically significant.

The indicators which had not altered in their occurrence on children's drawings since Koppitz collected her normative data were the extremely unusual and bizarre features such as slanting figure, tiny head, crossed eyes, genitals, monster/grotesque figures or three or

more figures drawn. Nowadays these features occur extremely rarely on any child's drawing, especially one such as 'tiny head' since children typically draw oversized heads on their figures. These features are also not major features of the conventional HFD, so it is unsurprising that their occurrence remains unchanged.

The changes in ages for some of the features showed some interesting findings. It is no longer unusual for a 5 year old's drawing to show a tiny figure without hands, though Koppitz considered both these items significant and indicative of disturbance. However, poor integration and shading of hands/neck on a 5 year old girl's drawing should according to the present data be considered significant whereas Koppitz did not think so. The data for drawing feet and neck have changed: in particular, the boys are now expected to produce feet on their HFD much younger than in Koppitz's day, but only girls aged 11 are expected to draw a neck instead of at age 9 and 10 years for girls and boys respectively as Koppitz thought. These changes mean that drawings scored using the Koppitz criteria with the respective age restrictions, may result in some indicators, which should be considered significant, going unnoticed. Conversely, some features which would be found on most normal children's drawings at a particular age would be considered clinically significant.

In the present study, indicators such as teeth and big figure were found to occur too often to be considered significant for many ages and a trend of increasing with age meant that the items were more likely to be due to increasing effects of culture and learning on the drawings rather than the children's disturbance. This casts doubt on the supposed link between indicators such as teeth and big figure and aggression. Even Koppitz found these items occurring more often in general than the more bizarre items such as genitals, reflecting the fact that a substantial contingent of normally adjusted children have always drawn their human figures in that way, a fact which casts doubt on the clinical significance of such details.

Short arms in particular were also found very often on the children's drawings and rather than being considered of clinical significance may be more to do with the size distortions commonly found in HFDs like oversized heads. The short arms may be due to the

children's lack of awareness of exactly how long arms are supposed to be when drawn hanging straight at a person's side as they are conventionally drawn by children.

The changes that were found can be seen as reflecting changes in the way children draw the human figure over time. Though these changes are only relevant to the emotional indicators they have implications for developmental aspects too, since they show a more general tendency for children's HFDs to fluctuate according to influences such as the media and schooling. The trends noted here may also be considered local trends specific to the north of England, representing the specific form which the Koppitz EIs take in this region. More importantly, the results found here suggest the potential problems associated with using scoring systems developed from age related norms that are not current to the research.

Study 5b

Re-analysis of study 1 using the revised indicators

Using the new age restrictions and omitting the nonsignificant indicators, it was possible to review the results of study 1 based on the new normative data. The results of this analysis are shown in the following tables. No drawing showed more than 4 of the indicators. For a full breakdown of the indicators scored see appendix 5.

No. of EIs	Clinical	CA control
0-1	16	16
2+	2	2

Table 21. (above left) Numbers of children's HFDs showing either 0 or 1 revised indicators and 2 or more revised indicators for the clinical and CA matched control samples.

No. of EIs	Clinical	MA control
0-1	16	14
2+	2	4

Table 22. (above right) Numbers of children's HFDs showing either 0 or 1 revised indicators and 2 or more revised indicators for the clinical and MA matched control samples.

These tables show again, as in the original comparison, similar distributions of scores across the three samples. Fisher Exact analysis revealed no significant differences occurring between numbers of HFDs in the clinical and CA control sample showing 0 or 1 indicators and numbers showing 2 or more indicators (p=0.7). This was also the case for the clinical and MA control comparison (p=0.3).

The clinical sample had a mean revised indicator score of 0.6 compared to the CA matched sample mean of 0.56 and the MA matched sample mean of 0.67. T-tests confirmed there were no significant differences between the clinical and CA matched control sample scores (t(34)=0.19, ns) or between the clinical and MA matched control samples (t(34)=0.18, ns)

This re-analysis still showed nonsignificant results as with the Koppitz original indicators. Therefore even using the revised indicators, there were no more of them occurring on the clinical sample children's HFDs than on the CA and MA matched control HFDs.

Study 5c

Re-analysis of study 3 using the revised indicators

The HFDs in study 3 were also scored for the revised indicators. Children in the three samples were categorised according to how many revised indicators they produced in their drawing. For a full breakdown of the indicators scored see appendix 5. No child produced more than 5 indicators. The data are shown in table 23.

Number of EIs	Clinical	CA control	MA control
0-1	29	42	40
2+	15	3	4

Table 23. Numbers of children's HFDs showing either 0 or 1 revised indicators and 2 or more revised indicators for the clinical and CA and MA control samples.

A 3x2 Chi-Square analysis revealed that there were significant differences among the samples (X^2=17.2, df=2, p<0.001). When

separate comparisons are made, 2x2 Chi-Square analysis showed a significant difference between the clinical and CA control samples (X^2=10.5, df=1, p<0.01), and also between the clinical and MA control samples (X^2=6.7, df=1, p<0.02), but not between the CA control and MA control samples (X^2=0.18, df=1, ns).

The clinical sample HFDs had a mean revised indicator score of 1.27 compared with the CA matched control sample mean of 0.32 and the MA matched control sample mean of 0.57. A one way ANOVA showed a significant difference among the three groups (F(2,129)=13.49, p<0.001). The clinical sample scored higher than the two control samples (Tukey's comparisons, ps<0.05) which were not significantly different from each other.

Using the revised indicators, significantly more of the indicators appeared on the drawings of the disturbed children than their normally adjusted counterparts, matched for CA and MA. There was no difference between the number of indicators appearing on the two control samples' drawings.

Retest reliability of the revised indicators Using the 27 disturbed children's drawings and 26 normally adjusted children's HFDs who were retested after 3 weeks (in study 3), it is also possible to examine the retest reliability of the revised indicators. A Pearson's correlation coefficient of 0.5 (p<0.01) for the disturbed children's HFDs and 0.6 (p<0.01) for the normally adjusted children's HFDs shows that the revised indicators are reliable over a three week interval.

DISCUSSION

The results of study 1 remained nonsignificant with the application of the revised indicators. This implies that even when the indicators are revised and based on relevant normative data for the samples used, they are still not useful or valid for discriminating this clinical from the normal samples. The reason why the clinical and control samples in study 1 showed no differences in the occurrence of the indicators in their HFDs was therefore not due to the use of American normative data from 30 years ago. It is more likely that the samples used in study

1 were too similar and that the severity of the clinical sample's disturbance was not extreme enough to be identified by the indicators, whether they are the original Koppitz ones or new revised ones.

The results of the revised indicator analysis for the study 3 data showed the same significant difference (on the ANOVA) between the clinical and control sample scores as in study 3. The Chi-Square result now achieves significance. The fact that the same pattern of results was found using the revised indicators as the original Koppitz indicators reinforces the conclusion from study 3, that the indicators are clinically valid and can discriminate between a clinical and normal sample. The indicators were revised to account for secular changes across time and location in order to make them more appropriate for the population upon which they were being used. The differences that were found in the normative data study, between the Koppitz 1968 norms and the UK 1996 norms, showed a change in the way that the normally adjusted populations of children draw the human figure. However, those changes still did not account for the differences in the way that the children with more severe EBD drew the human figure and the emotional indicators they included, compared to normally adjusted matched controls.

The clinical usefulness of the indicators, however, is impaired by the revision of the normative data since the list of valid indicators is reduced and the average number found even in severely disturbed drawings is very low. Koppitz admitted that the appearance of only one indicator in a drawing should not be considered clinically significant, yet the average number which occurred in the clinical sample in chapter 3 is close to one when using the revised indicators. Implications for the clinical use of the indicators are serious, since using the original Koppitz indicators, some features would be considered clinically significant when they actually occur on a large proportion of normal children's drawings or may be merely a part of the normal course of development.

The relationship between some items and age or maturation is only partly accounted for by Koppitz. Items that increase in occurrence with age were considered invalid as they are a part of the normal course of development. Items that decreased in occurrence were considered

significant once they fell below 15%. However, it may also be possible that these items which decrease in occurrence with age should also be considered as a normal part of development and not indicative of a deviancy as Koppitz thought. These items which were common on young children's drawings may be a sign of immaturity or delay at later ages and questions the idea that these features show a deviant nature in the child's drawing. If these are also removed from the indicator list this leaves only 17 valid items.

The results also place a question over the validity of recent data cited in studies which used the Koppitz indicators without rectifying them for alterations in normal children's drawings. Many authors have acknowledged that an appreciation of normal development is necessary before a full understanding of the abnormal or unusual can be gained. The studies which used the Koppitz emotional indicators may have used items which should no longer have been considered valid, due to their occurring on high numbers of normal children's drawings, or having a relationship with age or maturation.

New indicators This chapter has only evaluated the original 30 Koppitz emotional indicators. It is possible that items not previously considered may be valid for discriminating disturbed and normal children's drawings, though it is difficult to establish what these items might involve. Koppitz derived her list of items from those of Machover and from her own clinical experience. The Machover items have been extensively evaluated, as outlined in chapter 1, with most research failing to validate them. By assessing what the judges perceive as the differences between the drawings, it may be possible to resolve the issue of possible items but it would remain to be determined whether any new items not considered before would be valid.

Summary of chapter 4

This chapter has shown that the normative data upon which the indicators are based has changed. This reflects the cultural variations that can be seen in children's drawings and the changes which occur over time. In terms of the two criteria of occurring rarely and not increasing with age it can be seen that some indicators are no longer

valid, whilst some have only changed in the age at which they become rare enough to be considered clinically significant.

The revised list of indicators is smaller than the original, but the pattern of results remained the same when the earlier data were re-analysed. There continued to be no differences among the samples from chapter 2, and the differences remained among the samples in chapter 3. This means that though some indicators were no longer valid, differences still existed where they had occurred previously.

Since the pattern of results remains the same using the revised indicators, it seems more likely that the nonsignificant differences in chapter 2 were due to the relatively mild level of disturbance in the clinical sample. Using the revised list for the drawings of chapter 3 also did not change the overall pattern of results. It is therefore tempting to conclude that revising the normative data has had no effect on the emotional indicators, but there are important implications for these results nonetheless.

The main implication is the reduction in the number of valid indicators, which limits the practical significance of the indicators more so than they were already limited by low occurrences of the original features. Though the changes may seem subtle for some of the indicators, they make a difference when using the HFD in a clinical setting. Questions are also raised about the relevance of recent research that has placed interpretations on drawing features, without considering the normal course of children's drawing development, or the changes in children's drawings over time and between cultures. The issue of whether items increasing in occurrence with age are the only features related to maturation has also been raised, with the proposal that items decreasing in occurrence with age might also be related to maturation too.

In order to evaluate the Koppitz method, however, the three criteria will be adhered to for the purposes of this thesis in order to give comparable results with Koppitz. Future analysis will use both the Koppitz original and revised lists of indicators to allow for comparable results and also to see what differences are found when the normative data differences are taken into account.

CHAPTER FIVE

A COMPARISON OF THE HUMAN FIGURE DRAWINGS OF CHILDREN WITH EMOTIONAL/BEHAVIOURAL DIFFICULTIES AND THEIR GOODENOUGH-HARRIS MATCHED CONTROLS. USING THE INTUITIVE METHOD OF IDENTIFICATION WITH ART THERAPISTS AS EXPERTS

Introduction

Chapters 2 and 3 both considered the issue of whether the differences that are seen between clinical and normal samples are due to a deviant nature or a delay in the mental age of the clinical sample. These chapters considered mental age as measured by WISC subtests and found that it did not explain the differences between the drawings in terms of emotional indicators and the intuitive method. The Goodenough-Harris (GH) scale is a measure which also purports to measure intellectual maturity but is much closer in form to drawings, as the dependent measure being tested, and its involvement will be assessed in this chapter.

The GH scale is a cognitive measure of intellectual maturity that uses the quality of the drawing and the technical drawing ability of the children in terms of feature inclusion and proportions of the HFD. GH scale scores have modest correlations with standard IQ scores. Many (12 original, 10 revised) of the emotional indicators are confounded with the GH scale. The confounded items, however, must be removed from the indicator lists to create a valid comparison between the clinical and GH matched control drawings since the GH scaled scores require a full count through all the possible items.

The intuitive method is examined again in this chapter, using drawings matched for GH scores. This method has been successful in the previous chapters, only where indicator differences occur between the drawings. It is important to determine if matching the drawings for GH scores will remove the visual differences between them. This should give a clearer indication of what the judges are using to make their decisions.

This chapter introduces a different set of expert judges for the intuitive method. Previous experts were a mix of academics and teachers who had familiarity with children's drawings. Art therapists are used in this chapter to investigate whether their specific clinical training in artwork for diagnosis and therapy would help them with these tasks. Previous research has used art therapists as expert judges, usually finding that they perform no better than other types of interpreter. This chapter aims to determine whether these new judges can improve on the performance of the experts and novices already used. Though in chapter 3, all the judges were better than chance, it is important to determine if these new judges are more sensitive to the visual differences between the drawings, or whether there is a ceiling to the judges' performance which has already been reached. If this is the case, it may be that only some of the drawings contain the apparent differences that the judges are sensitive to, and it is not possible to identify the rest.

Study 6

Using the emotional indicators with human figure drawings matched for Goodenough-Harris scores

INTRODUCTION

The results of studies 3 and 4 suggest that there is a difference between the clinical sample children's human figure drawings and normally adjusted controls, which is not explained by mental age delay. However, this is mental age as measured by WISC items. Goodenough (1926) demonstrated that drawing for children "had a more cognitive than aesthetic meaning" (Harris 1963, p. 20). Therefore it is possible that the clinical sample children's drawings reflect a cognitive delay which was not identified by the WISC test items.

Harris (1963) thought that the notion of intelligence as it is measured by drawing, should be replaced with intellectual maturity or as he preferred to call it 'conceptual maturity', moving away from previously popular ideas of unitary intelligence. Harris believed, from evidence in the literature, that the child's drawing of an object is an

index of his/her conception of that object. This allowed for the child's concept of the human figure to be used as an index or sample of their concepts in general. The DAM test was therefore seen as testing primarily the ability to form concepts. The concept of a person as a concrete object undergoes elaborate differentiation with age and the child's drawing of the object revealed discriminations s/he has made about the object as a member of a class. The ability to specify relevant and significant features of concepts in drawing was seen as increasing with maturity.

The GH scoring system works on the basis that as a child matures his/her drawing becomes more differentiated and specific. Goodenough and Harris thought that the inclusion of more features and improved proportions in the human figure drawing as the child matures reflects his/her concept formation. Whether this is exactly the case is uncertain but it is possible that the scores are an indicator of the child's representations and a guide to his/her cognitive maturity. Equally, since this measure comes directly from the child's drawing it may be useful as a guide to the child's drawing ability and may be a useful measure to control for when comparing aspects of children's HFDs such as EI scores.

The present study was designed to see whether the indicator score differences would remain if the clinical sample drawings were matched for GH scores with normally adjusted controls. If the indicators are sensitive to disturbance which is independent of the child's mental age and drawing level, then the differences should remain but if the clinical sample children's higher indicator scores are just indications of poorer drawing ability, then the emotional indicator score differences previously found should disappear. The normative data upon which the GH scale is based is older than that of the Koppitz EIs, which may be a problem in using the scale, since it may under, or over-estimate drawing performance. However, any problem will affect both samples equally and therefore the position of both in relation to each other should remain.

Some items occur on both the GH scale and the indicator list and had to be removed to avoid confounding in the matching procedure. The GH scale credits items that increase in occurrence with age,

whereas the indicators include items omitted. For example, points are gained on the GH scale for including hands, whereas hands cut off is credited on the indicator list. The GH scale also credits points for good proportions whereas poor proportions such as short or long arm length is credited on the indicator list. Since the GH raw score was then scaled for age, removing items was not possible as it would decrease the scores artificially and make the scaled score inappropriate. Therefore the confounded items had to be removed from the emotional indicator list. The items removed were: tiny head, short arms, long arms, hands cut off, no eyes, no nose, no mouth, no body, no arms, no legs, no feet and no neck, leaving 18 in total.

METHOD

Participants

All participants were boys from the north of England, aged 6 to 11 years.

Clinical sample:
as in study 3.

GH matched control sample:
Forty four boys aged 6 to 11 years (mean 9;2, SD 1;4) matched for Goodenough-Harris scaled score to a corresponding child in the clinical sample. These children were from mainstream primary schools in the north of England.

Age at test	Clinical	GH control
6 years	0	2
7 years	3	1
8 years	7	12
9 years	11	11
10 years	17	10
11 years	6	8
Total	44	44

Table 24. Number of boys at each age in the two samples.

Design

A between-subjects design was used to compare participants' scores for emotional indicators from their HFDs. Scores used were from the Koppitz and revised lists without the 12 items that were confounded with items on the Goodenough-Harris scale.

Materials

Pencils, erasers and plain A4 paper were provided for each child.

Procedure

Drawings were used which were obtained from the clinical sample as described in study 3. The control sample drawings were selected from those obtained during group administration of the DAP test for study 5.

RESULTS

Correlation of GH and EI scores

The GH and indicator scores correlate significantly (-0.35 Koppitz; -0.49 revised) with the confounding items left in. These correlations become nonsignificant when the items are removed from the indicator list (0.04 Koppitz; -0.1 revised).

The confounded items from the Goodenough-Harris scale were removed from the original Koppitz indicator list, and the revised indicator list. The clinical and GH matched control sample HFDs were both scored for the remaining indicators.

Koppitz Indicators

The children's drawings were categorised according to how many original Koppitz indicators were in their drawing. No child produced more than 5 indicators out of the possible 18. For a full breakdown of participants' indicators see appendix 6. The summary data are shown in table 25.

No. of EIs	Clinical	GH control
0-1	27	28
2+	17	16

Table 25. Numbers of boy's HFDs showing either 0 or 1 Koppitz indicators and 2 or more indicators for the clinical and GH matched control sample.

Chi-Square results showed no significant difference between the clinical and GH matched samples (X^2=0.00, df=1, ns). The mean score for the clinical sample was 1.3 and the mean for the control sample was 1.2. A t-test showed no significant difference between the clinical and GH matched sample scores (t(86)=0.43, ns).

Revised Indicators

The children's drawings were categorised according to how many revised indicators were in their drawing. No child produced more than 2 indicators out of the possible 13. For a full breakdown of participants' indicators see appendix 6. The summary data are shown in table 26.

No. of EIs	Clinical	GH control
0-1	37	42
2+	7	2

Table 26. Numbers of boy's HFDs showing either 0 or 1 revised indicators and 2 or more indicators for the clinical and GH matched control sample.

Chi-Square results showed no significant difference between the clinical and GH matched samples (X^2=1.98, df=1, ns).

The mean score for the clinical sample was 0.77 and the mean for the control sample was 0.39. A t-test showed a significant difference between the clinical and GH matched sample scores (t(86)=2.79, $p<0.01$).

The difference between the Chi-Square and t-test results is due to the fact that there was significant difference between the numbers

producing no indicators (clinical 17, GH 29) and those producing one or more indicators (clinical 27, GH 15) which was identified by the t-test but not by the Chi-Square using 2 indicators as the cut-off point. Using a non-parametric test such as a Mann-Whitney U test, the significant difference remains ($z=-2.7$, $p<0.01$).

Summary of results

The results showed no significant differences between the clinical and GH matched control sample scores on the Koppitz indicators. With the revised indicators, there were no differences when the numbers scoring 0 and 1 were compared with 2 or more, but there were significantly more indicators scored on the clinical sample HFDs when the scores themselves were compared. This was due to significantly more control drawings showing no indicators at all compared to the clinical HFDs.

DISCUSSION

The negative correlation between the GH and indicator scores means that the indicator scores increase as the GH scores decrease, so the worse performance on the drawing ability measure corresponds to more indicators being present. The correlations are reduced and nonsignificant when the confounding items are removed. This implies that the indicators are just an indirect measure of cognitive maturity, similar to the GH scale.

The results of this study showed no difference in Koppitz emotional indicator scores between the clinical and control sample drawings. The only difference to be found was when the revised indicators were used and a comparison made between the number of drawings with no indicators at all, and those drawings with at least one indicator. Here, there were significantly more control HFDs with no indicators at all. However, though this result achieves statistical significance, it is of limited practical significance.

It is important to acknowledge that the working number of indicators has been reduced both by the normative study and the GH scale confounded items. This questions the validity of those that are left. The average number of indicators in the clinical sample has fallen

to around one indicator per drawing which is insignificant by Koppitz's own admission.

The confounding items from the GH scale reduced the number of Koppitz indicators from 30 to 18 and the revised indicators down to 13. This shows that a large part of the original indicators were simply a result of delay in development, since they are measured on the GH scale and can therefore be seen as involved in drawing/concept development. The drawing delay may be symptomatic of the child's disturbance which delays much of their cognitive and emotional development, and therefore may still be useful for determining disturbed children. However, equivalent delay is found in normally adjusted children, such as the control sample in this study.

The only difference to remain between the boys in this study was their emotional health/status, given that the GH scores were matched, and the indicator scores largely disappeared. This implies that the high scores previously found in the clinical sample pictures was mainly due to their lack of drawing skill, as measured by the GH scale. The developmental delay in this skill was not identified by the WISC items since when these scores were controlled, the indicator differences remained. This lack in skill can be found in normally adjusted children too, so it is a mistake to say that the lack of certain items expected at a certain age means that the child has emotional health problems. The drawing may be immature for many reasons other than the emotional status of the child. Those indicators showing delay which were confounded with GH scale items may have been the ones accounting for the differences previously found.

There may be a problem that the GH matched control drawings were taken from the sample of children used to produce the revised list. This would therefore inevitably lead to these drawings containing fewer revised indicators since the same sample is being used to validate the items as was used to construct them. However, since a similar pattern is found as with the Koppitz indicators which do not suffer this problem, the impact of this problem is limited.

Following this discovery that the indicator differences no longer appear when the GH scale is accounted for, it is necessary to determine if the visual differences remain. When the clinical sample HFDs were

matched for CA and MA, novice and expert judges could both determine successfully the clinical from control drawings. It is important therefore to examine whether the judges can still identify the clinical sample HFDs from their GH scale matched controls.

Study 7

Using the intuitive method of identification with human figure drawings matched for Goodenough-Harris scores

INTRODUCTION

It has been found that judges can discriminate the clinical sample drawing from CA and MA matched controls where the drawings have differences in indicator scores (i.e. study 4 not study 2). Study 2 was unable to determine whether the failure of the judges was due to a lack of any difference between the drawings, or an inability on the part of the judges to perform the task. Study 4 went some way towards rectifying this, showing that where the indicator scores of the drawings were different, the judges were able to identify the clinical HFDs and successfully complete the task.

When the same set of study 4 clinical HFDs were compared with GH matched controls, the Koppitz indicator differences disappeared. This present study attempts to determine whether judges can also discriminate the clinical HFDs from their GH matched controls. This will examine whether the judges' success depends on the appearance of indicator differences in the drawings. The results of study 4 have shown the judges capable of discriminating the clinical sample HFDs, therefore a failure on this present task would reflect the lack of differences between the drawings and not the judges' inability to perform the task.

Novice judges were used since it has been shown previously that there were no differences between them and experts. The different format and age conditions were also used again to determine whether these would still affect the judges performance.

METHOD

Participants

Twenty novice judges were used from the undergraduate student population at the University of York. All were over 18 years.

Design

Four tasks were used, consisting of two format conditions (group and pair, between-subject variable) and two age conditions (no ages and with ages, within-subject variable). Judges performed one task, both with and without the children's ages.

Materials

Two sets of drawings from study 6 were used, one from the clinical sample and one from the GH matched control sample. These were made into 2 tasks:

Group format- (1) randomly mixed group of clinical and GH matched control sample drawings (n=44 per sample)

Pair format- (2) set of 44 pairs of drawings, one from the clinical sample paired with the appropriate GH matched control

A list of the drawing codes with the relevant age of the boys was provided when the judge was performing the task 'with ages'.

Procedure

Judges were asked to discriminate the clinical sample HFDs from the control sample drawings. Instructions were the same as in study 4 for each of the relevant conditions. No criteria were given for making the discriminations. Judges performed the tasks without and then with the ages of the children.

RESULTS

Each participant (n=20) performed the task of discriminating the clinical sample children's drawings from the control sample in one of the following conditions:

between-subject - format (2) group and pair

within-subject - age (2) no ages and with ages

The scoring was the same as for study 4 with a total score as well as clinical and control scores expressed as proportions.

Analysis of the different conditions

The numbers of judges performing at and above chance levels (using the familywise error rate of $p<0.05/10=0.005$) in each of the conditions for each of the score types was analysed. Only one judge was found to be able to perform better than chance ('group format, with ages' condition, total score).

The scores were converted to percentages in order to control for the differing denominators in the clinical and control scores, as well as the group and pair tasks. The average percentage scores are shown in table 27.

Condition	Total	Clinical	Control
group/no ages	49	47	50
pair/no ages	49	49	49
group/with ages	58	59	58
pair/with ages	59	59	59

Table 27. The average percentage score in each of the conditions, for each of the score types.

The highest average of 59% converts to a score of 26 out of 44 in the pair task which is not significantly above chance.

Repeated measures ANOVAs showed a significant main effect of age for the total ($F(1,18)=37.34$, $p<0.001$), clinical ($F(1,18)=46.1$, $p<0.001$) and control ($F(1,18)=28.83$, $p<0.001$) percentage scores. There was no main effect of format and no interaction. The main effect of age is due to the fact that the average scores were higher when the ages were available, though it is important to note that the scores were still not above chance levels.

Individual HFD Analysis

The number of judges correctly and incorrectly identifying each of the 44 clinical and GH matched control HFDs was calculated for the 'no ages' and 'with ages' conditions. The data are collapsed across the

'group' and 'pair format', since the ANOVA showed no effect of format. Binomial tests were used to determine whether significantly more judges than would be expected by chance agreed in their decisions. The familywise error rate of $p<0.05/44=0.001$ was used.

Clinical HFDs

The number of judges correctly and incorrectly identifying the clinical HFDs is shown in table 28, with corresponding binomial test results. Significantly more incorrect judgements were made for the 'no ages' condition and significantly more correct judgements for the 'with ages' condition. Mann-Whitney tests comparing the numbers of correct judges with the number of incorrect judges for each of the 44 HFDs showed significantly more judges incorrect than correct for the 'no ages' condition ($z=2.64$, $p<0.01$) but no difference for the 'with ages' condition ($z=1.79$, ns). As can be seen in table 28, a significant number of judges correctly identified 3 HFDs in the 'no ages' condition and 8 in the 'with ages' condition. 9 HFDs were incorrectly identified in the 'no ages' condition and none was incorrectly identified in the 'with ages' condition.

GH matched control HFDs

The number of judges correctly and incorrectly identifying the GH matched HFDs is shown in table 29, with corresponding binomial test results. Significantly more correct judgements were made in the 'no ages condition' and the 'with ages' condition. Mann-Whitney tests showed no significant difference for the 'no ages' condition ($z=1.69$, ns), but significantly more judges correct than incorrect for the 'with ages' condition ($z=3.87$, $p<0.001$). As can be seen in table 29, a significant number of judges correctly identified 10 HFDs in the 'no ages' condition and 10 in the 'with ages' condition. 5 HFDs were incorrectly identified in the 'no ages' condition and 3 were incorrectly identified in the 'with ages' condition.

	No ages			With ages		
HFD	✔	✘	p<	✔	✘	p<
C1	9	11	ns	6	14	ns
C2	17	3	0.001	16	4	ns
C3	15	5	ns	9	11	ns
C4	6	14	ns	6	14	ns
C5	4	16	ns	8	12	ns
C6	13	7	ns	8	12	ns
C7	15	5	ns	18	2	0.001
C8	10	10	ns	14	6	ns
C9	6	14	ns	14	6	ns
C10	10	10	ns	13	7	ns
C11	11	9	ns	11	9	ns
C12	17	3	0.001	20	0	0.001
C13	5	15	ns	10	10	ns
C14	3	17	0.001	10	10	ns
C15	6	14	ns	5	15	ns
C16	4	16	ns	4	16	ns
C17	5	15	ns	12	8	ns
C18	14	6	ns	19	1	0.001
C19	11	9	ns	11	9	ns
C20	2	18	0.001	11	9	ns
C21	3	17	0.001	10	10	ns
C22	11	9	ns	17	3	0.001
C23	1	19	0.001	12	8	ns
C24	7	13	ns	5	15	ns
C25	3	17	0.001	6	14	ns
C26	4	16	ns	8	12	ns
C27	4	16	ns	4	16	ns
C28	6	14	ns	19	1	0.001
C29	19	1	0.001	19	1	0.001
C30	14	6	ns	7	13	ns
C31	7	13	ns	16	4	ns
C32	12	8	ns	9	11	ns
C33	1	19	0.001	20	0	0.001
C34	8	12	ns	7	13	ns
C35	7	13	ns	11	9	ns
C36	16	4	ns	17	3	0.001
C37	16	4	ns	15	5	ns
C38	3	17	0.001	13	7	ns
C39	15	5	ns	13	7	ns
C40	7	13	ns	5	15	ns
C41	14	6	ns	16	4	ns
C42	8	12	ns	6	14	ns
C43	3	17	0.001	5	15	ns
C44	1	19	0.001	5	15	ns
sum	373	507	0.001	490	390	0.001

Table 28. Numbers of judges correctly and incorrectly identifying the clinical HFDs in the two age conditions.

	No ages			With ages		
HFD	✔	✘	p<	✔	✘	p<
GH1	15	5	ns	6	14	ns
GH2	16	4	ns	18	2	0.001
GH3	11	9	ns	14	6	ns
GH4	12	8	ns	11	9	ns
GH5	7	13	ns	4	16	ns
GH6	16	4	ns	15	5	ns
GH7	19	1	0.001	16	4	ns
GH8	8	12	ns	11	9	ns
GH9	4	16	ns	10	10	ns
GH10	17	3	0.001	15	5	ns
GH11	16	4	ns	16	4	ns
GH12	19	1	0.001	11	9	ns
GH13	5	15	ns	11	9	ns
GH14	3	17	0.001	0	20	0.001
GH15	5	15	ns	8	12	ns
GH16	10	10	ns	13	7	ns
GH17	8	12	ns	13	7	ns
GH18	7	13	ns	14	6	ns
GH19	17	3	0.001	17	3	0.001
GH20	6	14	ns	9	11	ns
GH21	10	10	ns	8	12	ns
GH22	16	4	ns	20	0	0.001
GH23	6	14	ns	7	13	ns
GH24	7	13	ns	10	10	ns
GH25	2	18	0.001	3	17	0.001
GH26	7	13	ns	8	12	ns
GH27	8	12	ns	11	9	ns
GH28	14	6	ns	16	4	ns
GH29	20	0	0.001	20	0	0.001
GH30	11	9	ns	14	6	ns
GH31	14	6	ns	19	1	0.001
GH32	17	3	0.001	18	2	0.001
GH33	0	20	0.001	13	7	ns
GH34	11	9	ns	10	10	ns
GH35	10	10	ns	11	9	ns
GH36	18	2	0.001	19	1	0.001
GH37	20	0	0.001	20	0	0.001
GH38	3	17	0.001	7	13	ns
GH39	19	1	0.001	20	0	0.001
GH40	13	7	ns	9	11	ns
GH41	20	0	0.001	20	0	0.001
GH42	14	6	ns	16	4	ns
GH43	6	14	ns	11	9	ns
GH44	0	20	0.001	1	19	0.001
sum	487	393	0.002	543	337	0.000

Table 29. Numbers of judges correctly and incorrectly identifying the GH matched HFDs in the two age conditions.

DISCUSSION

The results show that the judges could not successfully differentiate the clinical from control HFDs. Only one judge out of 20 performed better then chance overall and the highest average score was not better then chance. The judges did improve in the 'with ages' condition but not enough to make a difference to the significance of their results. Where the judges were consistent in their decisions, above that which would be expected by chance, they were often incorrect. Without the children's ages, most of these incidences (9 out of 12), for the clinical HFDs, occurred in the wrong direction, with the judges identifying the drawings as controls. Only three pictures were correctly identified by a significant number of judges. With the children's ages added, this changed to all the consistent judgements (8 out of 8) being made in the right direction, with two of the same drawings involved. The judges seemed to have more success with the control drawings with two thirds of the consistent judgements (10 out of 15) for the control HFDs in the right direction, even without the children's ages. When the ages were given, similar numbers were obtained (10 out of 13), with seven for the same pictures. About one quarter of the HFDs in the control sample were correctly picked, but only a small minority of the clinical ones were. Overall, even the consistent judgements were not enough to make the judges' scores better than chance.

The results imply that the success of the judges depends on the drawings' differing in two respects - the indicator scores and GH scores. When these two scores are similar between the drawings, it becomes impossible to discriminate the HFDs on the basis of visual inspection. These two measures are confounded and it is difficult to determine whether one or the other is ultimately responsible for the visual differences between the drawings. It is possible that the judges were using the indicators as a guide to sample membership since when indicator differences were not present, the judges' success in the task disappeared also. However, the indicators are obscure items which lay people may not know about. The GH scale is equally obscure but uses features which people are more commonly aware of such as the inclusion of body parts and proportions between them. It may be more likely that the judges are more sensitive to these features of drawing

development than features such as the Koppitz indicators as symbols of disturbance. This implies that the differences that they were identifying previously were due to differences in GH scores and, in particular, the low GH scores of the clinical sample. This means that the judges are sensitive to a cognitive delay in the drawings, rather than emotional disturbance, which casts doubt on the validity of the DAP test.

Studies 2 and 4 comparing expert and novice judges' use of the intuitive method were limited since the judges who took part were not trained in the use of artwork in a clinical setting. The experts were people who had familiarity with children's drawings only, and may not have directly used them in a clinical sense. It is possible therefore that the use of judges from a more clinical setting, who have had more applied training, such as art therapists, might be useful.

Study 8

Art therapists using the intuitive method of identification

INTRODUCTION

Studies 2 and 4 in this thesis have used experts and found that they were no better than novices in discriminating the drawings of the clinical sample from those of the controls. The experts used were academics and teachers who had experience with children's drawings. The performance of these experts may have been restricted due to the limited relevance of their experience. This present study attempts to address this issue by using art therapists as experts. Research has previously found that the amount of clinical experience judges had with drawings did not correlate with success in discriminatory tasks (Ulman & Levy, 1973). Art therapists did not perform more successfully than the other mental health professionals and novices used, though 90% of the judges were above chance levels. Following a review of the literature on projective drawing techniques, including the DAP test, Neale and Rosal (1993) concluded that art therapists cannot diagnose from artwork because "there is not enough information about drawing techniques for children" (p. 47).

The present study aims to determine whether the training and clinical experience using artwork for diagnosis and therapy will help art therapists succeed at these tasks and if they can improve on the other judges who only had familiarity with children's drawings. Though Ulman and Levy found that their art therapists did not perform better than novice judges, they used artwork from adults rather than children and changes in training may alter the performance of art therapists today. A question also surrounds the ceiling of the judges' performance, since no judge has yet got all the drawings right and they average around 70% correct overall. It is important to know if this can be improved on if a different set of judges is used. This study employs all the tasks completed by the other experts in chapter 3. The art therapists also attempted the tasks from the previous study in this chapter, using the GH matched control drawings, as the novices were unable to succeed in this condition.

METHOD

Participants

Eleven art therapists took part in this study. Nine were recent graduates from postgraduate courses in the South of England. Two were art therapists from the York area. All were practising art therapists.

Design and Procedure

Tasks taken from studies 4 and 6 were used to compare the performance of the art therapists, with the experts and novices already involved in chapters 3 and 5. The art therapists followed the instructions used in studies 4 and 6. They were required to discriminate the drawings in piles or pairs to identify the clinical HFDs, performing the tasks both with and without the boy's ages available.

Materials

Four sets of 44 drawings collected during study 4 and 6 were used, one from the clinical sample of boys with EBD, one from the CA matched controls, one from the MA matched controls and one from the GH matched controls. These were made into six tasks:

Group format- (1) randomly mixed pile of clinical and CA matched control sample HFDs

(2) randomly mixed pile of clinical and MA matched control sample HFDs

(3) randomly mixed pile of clinical and GH matched control sample HFDs

Pair format- (4) 44 pairs of drawings, each pair containing one clinical sample and one CA matched HFD

(5) 44 pairs of drawings, each pair containing one clinical sample and one MA matched HFD

(6) 44 pairs of drawings, each pair containing one clinical sample and one GH matched HFD

A list of the boy's ages was provided for the relevant conditions.

RESULTS

The results are shown in terms of the total number of correct discriminations made for each task by the art therapists. The mean from the two art therapists in each task (only one for 'MA control, group format' condition) is shown. Table 30 shows the proportion of total correct responses, percentages of correct responses and significance levels for the different conditions.

Condition	Total	%	p<
CA/group/no ages	52/88	59	ns
CA/group/with ages	61/88	69	0.001
CA/pair/no ages	27/44	61	ns
CA/pair/with ages	35/44	80	0.001
MA/group/no ages	50/88	57	ns
MA/group/with ages	62/88	70	0.001
MA/pair/no ages	25/44	57	ns
MA/pair/with ages	28/44	64	ns
GH/group/no ages	43/88	48	ns
GH/group/with ages	45/88	51	ns
GH/pair/no ages	22/44	50	ns
GH/pair/with ages	27/44	61	ns

Table 30. Mean total number of correct discriminations with associated binomial test results for the 6 different tasks.

These results show that the art therapists could significantly discriminate the clinical HFDs from the 'CA control, group and pair format' tasks and 'MA control, group format' task in the 'with ages' condition. They were unable to discriminate the drawings above chance in the 'GH control' conditions. The art therapists were unable to discriminate the drawings in the 'MA control, pair format' tasks and without the boy's ages. Due to their being a maximum of only two art therapists in each condition, it is difficult to further analyse their scores.

The proportion of drawings that were correctly discriminated by the judges are used for comparison. The total number of correct judgements made as a proportion of the total number of possible correct judgements is shown for the eight novices (10 in GH matched conditions), two experts and two (or one) art therapists who completed each task.

	Novices		Experts		Art Therapists	
Condition	n/x*	%	n/x	%	n/x	%
CA/group/no ages	412/704	59	119/176	68	104/176	59
CA/group/with ages	482/704	69	129/176	73	122/176	69
CA/pair/no ages	263/352	75	60/88	68	54/88	61
CA/pair/with ages	288/352	82	65/88	74	69/88	78
MA/group/no ages	418/704	59	109/176	62	50/88	57
MA/group/with ages	482/704	69	127/176	72	62/88	71
MA/pair/no ages	234/352	67	66/88	75	49/88	56
MA/pair/with ages	263/352	75	69/88	78	56/88	64
GH/group/no ages	430/880	49	-	-	86/176	49
GH/group/with ages	511/880	58	-	-	90/176	51
GH/pair/no ages	215/440	49	-	-	43/88	49
GH/pair/with ages	261/440	59	-	-	54/88	61

Table 31. Proportion of total correct judgements in each condition for the three types of judges.

*n= number of correct judgements made by the judges; x= total number of correct judgements possible.

'-' indicates where the experts did not take part in those tasks.

The results of the art therapists were compared to those of the 2 experts and 8 or 10 novices using a test to assess the significance of the difference between two independent proportions. The following table shows the results for the test of proportions between the art therapists and the experts and novices, for each of the conditions.

Condition	Nov vs. AT	Exp vs. AT
CA/group/no ages	ns	p<0.05(exp)*
CA/group/with ages	ns	ns
CA/pair/no ages	p<0.01(nov)	ns
CA/pair/with ages	ns	ns
MA/group/no ages	ns	ns
MA/group/with ages	ns	ns
MA/pair/no ages	p<0.05(nov)	p<0.01(exp)
MA/pair/with ages	p<0.05(nov)	p<0.05(exp)
GH/group/no ages	ns	n/a
GH/group/with ages	p<0.05(nov)	n/a
GH/pair/no ages	ns	n/a
GH/pair/with ages	ns	n/a

Table 32. Results for the test of proportions between the judges for the 12 different conditions.

*Letters in brackets indicate which judge type was superior in performance (nov=novices; exp=experts; AT=art therapists).

The use of the familywise error rate ($p<0.05/20=0.002$) discounts the significant results in the table, which in any case tended to be in the wrong direction to what would be expected. The results showed that the art therapists performed no better than the experts or novices already evaluated. It was expected that they would perform better on the tasks than the novices and experts due to their specific training, but where significant results were found (without the familywise rate), these were in the wrong direction.

DISCUSSION

The results show that the art therapists performed no differently from the other experts and novice judges previously used (and may even have performed worse). The art therapists correctly discriminated the

same proportion of drawings as the other judges, in each of the conditions.

This resolves the issue that the other experts did not have the right experience to make them more successful than novices. The results suggest that experience is neither necessary nor advantageous in order to discriminate the disturbed boy's HFDs from CA or MA matched controls. It is also not sufficient in order to discriminate the clinical HFDs from GH matched controls.

This reinforces the results of other researchers such as Ulman and Levy (1973) who also found that art therapists were not better than other mental health professionals and novices. It seems that experience with children's drawings, whether academic or clinical, does not improve the ability to discriminate clinical from normal drawings. However, this does not necessarily cast doubt on the validity of this method itself as people such as Motta et al. (1993) would prefer, since in some conditions even novice judges are able to perform above chance levels.

There does seem to be a ceiling on the performance of anyone on this task. Some judges came nearer to this upper limit than others, but in general, the judges identify 75% with the boy's ages available, at least for the drawings used here. This has implications for the usefulness of the DAP test, since it is only possible to identify some of the drawings, with any accuracy. Since the art therapists performed at the same level as the novices, this means that their training is not helpful in identifying those HFDs which the other judges could not, as may have been expected. However, due to the low number of judges in each condition it is not possible to determine whether significantly more art therapists than chance identified certain drawings. It can only be assumed that they isolated the same HFDs as the other judges in order to achieve the same proportion of correct discriminations overall.

The GH scale is again implicated in these results. As with the previous expert and novice judges, the art therapists were unable to discriminate the clinical HFDs from the GH matched control HFDs. The art therapists, as the other judges before them, may be simply picking up on the low GH scores of the clinical sample and are sensitive

only to a cognitive delay in the drawings, casting doubt on the validity of the DAP test as a measure of emotional disturbance. What the judges are actually using to identify the drawings still needs to be clarified. The differences may be carried in specific features such as the emotional indicators or GH scale items or may be more apparent in the global impression of the drawing. This should be investigated in order to clarify the differences between the drawings and what the judges are indeed using to make their discriminations.

Summary of Chapter 5

This chapter has found that the Koppitz indicators do not occur more often on the drawings of severely disturbed boys when compared with GH matched drawings. The lack of differences when GH scores are controlled suggests that the previous differences seen in terms of indicator scores were due to differences in GH scores. The clinical sample children were typically delayed in their drawing development as their GH scaled scores were lower than average for their ages. The confounding of indicator items with the GH scale reinforces the conclusion that the indicator differences seen previously were due to differences in GH scale scores. This conclusion is important because it casts doubt on the view that the DAP test is clinically useful for assessing emotional disturbance. Any differences which may be seen between clinical and control samples in terms of emotional indicators may be due to cognitive delay specific to drawing and related abilities on the part of the clinical children.

Using the intuitive method judges were also not able to discriminate the severely disturbed HFDs from the GH matched HFDs. This reinforces the pattern seen in previous chapters whereby the judges are only successful when indicator differences are found. This result suggests that in the intuitive method the judges may be sensitive to visual differences that are similar to indicator items. However, since the indicator items are also confounded with the GH scale, it may be that the judges are using GH scale type items to make their successful discriminations.

The art therapist judges were no more successful on the tasks than either the other experts or novices. This means that experience of

children's drawings and training in art therapy is not necessary in order to discriminate a clinical from a control drawing above chance levels, when the clinical sample child is severely disturbed. The issue of expert vs. novice judges has been debated in the literature and the results seen in this and previous chapters suggest that novice judges who have no previous experience with children's drawings are as successful in identifying a disturbed child's HFD as an expert. The question of what the judges use to make their decision is less clear.

In each of the studies using the intuitive method some drawings were found which were identified consistently by more of the judges than would be expected by chance. This suggests that the judges were all identifying the same things in these particular pictures, though it is not clear what that is. Those few drawings may explain the success of the judges overall, but this is unlikely, since a significant number of judges identify certain individual drawings even when the overall success of the judges is at chance levels, such as in chapter 2 and this present chapter. Also, the judges are sometimes consistently incorrect, rather than correct, in the drawings they identify as belonging to a disturbed child. There is obviously something in these drawings which makes the majority, if not all of the judges, think they belong to disturbed children. This factor may be closely linked to the emotional indicators or it may be contained in some other factor such as a more global impression of the picture.

CHAPTER SIX

BIZARRENESS AND THE EMOTIONAL INDICATORS. DIFFERENCES BETWEEN DISTURBED AND NORMAL HUMAN FIGURE DRAWINGS

Introduction

Bizarreness is a factor which has been implicated in the clinical use of drawings in several studies discussed in chapter 1 (e.g. Hiler & Nesvig, 1965; Yama, 1990). It appears to account for much variance in drawings and is sometimes more successful for discriminating them than are the emotional indicators. This chapter therefore aims to investigate whether the drawings from the severely disturbed children of chapter 3 are more bizarre than the CA, MA and GH matched control children's drawings. If bizarreness is related to the indicators, then the pattern of results should be similar, with differences between the clinical and CA and MA matched control samples but not GH matched controls. The pattern of results will help to clarify how the sets of drawings differ from each other in terms of a global rating. This could be useful for determining the clinical validity of drawings.

It is also important to discover how the emotional indicators are linked to a global impression of bizarreness in the drawings of disturbed children, in order to fully understand the role that the indicators play and assess their validity. The presence of the indicators may create the impression of bizarreness, which may explain the decisions of the judges using the intuitive method. If, however, the indicators are not related to the level of bizarreness in a drawing, then it will be important to clarify what factors are involved in the drawings and explain what the judges were using to discriminate the clinical from control HFDs.

Study 9

Ratings of bizarreness

INTRODUCTION

It has been found in previous research (Hiler & Nesvig, 1965), that a general impression of bizarreness was used by judges as a criterion for

discriminating normal and clinical populations. Yama (1990) obtained separate ratings for artistic quality and bizarreness and, though they were both related and predicted overall adjustment, it was shown that ratings of drawing bizarreness were more successful in identifying a clinical drawing than any other measure, including the Koppitz indicators which did not predict later adjustment at all.

It has been shown in this thesis that judges can successfully determine the clinical from the CA and MA control HFDs, using the 'intuitive method of identification' (Dieffenbach, 1977). The judges, however, failed to successfully differentiate the clinical from GH matched drawings. The differences (or, in the GH matched pictures, the lack of differences) in indicator scores could have been seen as the reason for the judges' success (or lack of it). The indicator scores of those HFDs chosen as clinical, however, when examined, were not different from the scores of those identified as control drawings, so it is doubtful whether the judges were using the indicators as a means of determining the clinical from control HFDs.

The bizarreness of the pictures may have been a factor which the judges were using as a guide to sample membership in the previous judges studies. If this were the case, then the bizarreness ratings of the pictures would be expected to be significantly different in the clinical and CA and MA samples, but not in the clinical and GH samples. This could then account for the previous decisions of the judges.

The following study obtained ratings of bizarreness for all the HFDs previously collected in study 3 and study 6, with the aim of answering the above question. This study should also be able to see if the drawings usually chosen as clinical (as well as the actual clinical ones) are given higher ratings of bizarreness than those which were chosen as controls (and are actually controls). It can also be determined whether the drawings given higher ratings of bizarreness also have higher indicator scores. This should show whether the indicators are more, or less, useful than these ratings.

METHOD

Participants

Nine undergraduate students at the University of York acted as judges in this study.

Design

Six tasks were used: three control matches (CA/MA/GH) and two age conditions (no ages and with ages). The between-subject variable was the control matches. The within-subject variable was the age condition. Three judges each performed one task, both without and with the boy's ages. The drawings were randomised for each judge. A six point rating scale from 'not bizarre' to 'completely bizarre' was used.

Materials

Four sets of drawings collected previously were used, one from the clinical sample of boys with EBD from study 3 (n=44), one from the CA matched control sample (n=44) and one from the MA matched control sample from study 3 (n=44) and one from the GH matched control sample from study 6 (n=44). These were made into three mixed sets of 88 HFDs:

Clinical and CA	(1) randomly mixed group of clinical and CA matched control sample drawings
Clinical and MA	(2) randomly mixed group of clinical and MA matched control sample drawings
Clinical and GH	(3) randomly mixed group of clinical and GH matched control sample drawings

Sets of these drawings were also produced with the boy's ages added for the 'with ages' condition.

Procedure

Judges were asked to rate each drawing on a six point scale from not bizarre to completely bizarre, recording their answers on a separate sheet. The following written instructions were given:

"*'Bizarre': strange in appearance or effect; eccentric; grotesque.*

(Oxford English Dictionary)

The following is a set of 88 human figure drawings collected from boys aged between 6 and 12 years of age. The children were asked to "draw a whole person". Your task, taking each drawing in turn, is to decide how 'bizarre' you consider the drawing. This is done on a scale of increasing bizarreness from one to six, whereby 'one' means the drawing is not bizarre at all, and 'six' means the drawing is completely bizarre.

---> increasing bizarreness --->

1	2	3	4	5	6
not					completely
bizarre					bizarre

Each drawing has a code in the bottom corner. Please mark your rating response in the column next to the drawing code on your answer sheet."

Judges first performed the task without the boy's ages, then repeated the task with drawings which had the boy's ages added.

RESULTS

Each judge gave each drawing a rating from 1-6. Ratings were analysed separately for the three control conditions and two age conditions. Kendall's coefficient of concordance was used to determine the agreement among the three judges for each of the sets of drawings (clinical vs. CA matched; clinical vs. MA matched; clinical vs. GH matched) in both the ages conditions. This converts into a Chi-Square statistic which is shown in table 33.

	Clinical vs. CA control		Clinical vs. MA control		Clinical vs. GII control	
	no ages	with ages	no ages	with ages	no ages	with ages
X^2	47.2	52.8	29.6	47.3	78.5	96.4
p<	0.001	0.001	0.001	0.001	0.001	0.001

Table 33. Chi-Square and associated p values for the measure of agreement among the three judges in each condition.

The above table shows that there was very good agreement among the three judges for the bizarreness ratings in all the conditions. This allows confident use of the mean rating in further analyses.

NB. For examples of drawings which achieved high ratings of bizarreness, see figures 20 to 33, presented in study 11.

Clinical vs. CA matched

	No ages	With ages
Clinical	2.92	3.3
CA control	1.89	1.67

Table 34. Mean ratings for the clinical HFDs when mixed with the CA control samples, and the CA control HFDs, in the 'no ages' and 'with ages' conditions.

A repeated measures ANOVA showed a main effect of sample ($F(1, 86)=55.25$, $p<0.001$), no main effect of age ($F(1, 86)=0.77$, ns) but a significant interaction ($F(1, 86)=11.68$, $p<0.001$). Post hoc tests (Tukey's HSD) showed that the clinical sample HFDs were rated significantly higher than the CA matched HFDs in the 'no ages' ($p<0.001$) and 'with ages' ($p<0.001$) conditions. The clinical sample HFDs were rated significantly higher in the 'with ages' condition compared with the 'no ages' condition ($p<0.05$), but the CA matched HFDs rating were not significantly different across the age conditions.

Clinical vs. MA matched

	No ages	With ages
Clinical	2.92	3.02
MA control	2.33	2.21

Table 35. Mean ratings for the clinical HFDs when mixed with the MA control samples, and the MA control HFDs, in the 'no ages' and 'with ages' conditions.

The clinical HFDs were rated as significantly more bizarre than the MA matched control sample ($F(1, 86)=13.83$, $p<0.001$). There was no main effect of age ($F(1, 86)=0.22$, ns) and the interaction which occurred with

the CA control sample just failed to reach significance here ($F(1, 86)=3.88$, $p<0.052$).

Clinical vs. GH matched

	No ages	With ages
Clinical	2.38	3.05
GH control	2.52	2.77

Table 36. Mean ratings for the clinical HFDs when mixed with the GH control samples, and the GH control HFDs, in the 'no ages' and 'with ages' conditions.

The ANOVA showed that the clinical ratings were no different from the GH matched control ratings ($F(1, 86)=0.09$, ns) though there was a main effect of age ($F(1, 86)=34.68$, $p<0.001$) and a significant interaction ($F(1, 86)=7.16$, $p<0.01$). Post hoc analysis showed that the clinical sample HFDs were rated significantly higher in the 'with ages' compared with the 'no ages' condition (Tukey's HSD, $p<0.001$).

Correlations were calculated between the mean ratings for each drawing in each condition and the respective EI scores (Koppitz and revised). To avoid falsely inflating the correlations, 12 confounding GH scale items were excluded from the indicator scores for the correlations involving the indicators and the GH scale (clinical HFDs in the GH matched control condition, GH matched control drawings). The 24 correlations calculated ranged from 0.007 to 0.53 with no particular pattern; 11 were significant but the results were inconsistent.

Correlations were also calculated between the mean ratings for the drawings in each condition and the chronological ages of the boys. Correlations ranged from -0.59 to 0.04 with 7 out of 12 statistically significant but there was no consistent pattern.

Comparison with study 4 and study 7

Studies 4 and 7 found that some drawings were correctly and incorrectly identified by more judges than would have been expected by chance. The EI scores of these HFDs did not differ. The bizarreness ratings of these drawings were examined. The mean ratings of those

drawings that were identified either correctly or incorrectly by significantly more judges than were expected by chance (not using the familywise error rate) were compared with those that were not identified by a significant number.

		Sig. correct	Nonsig.	Sig. incorrect
Clinical (with CA)	No ages	4.14 (12)	2.65 (23)	2 (9)
	With ages	4.22 (20)	2.61 (18)	2.28 (6)
Clinical (with MA)	No ages	3.97 (11)	2.72 (24)	1.67 (9)
	With ages	3.63 (21)	2.67 (19)	1.42 (4)
Clinical (with GH)	No ages	2.71 (8)	2.44 (21)	2.11 (15)
	With ages	3.92 (12)	2.95 (25)	1.9 (7)
CA control	No ages	1.43 (14)	1.96 (25)	2.8 (5)
	With ages	1.24 (21)	2 (21)	2.67 (2)
MA control	No ages	1.86 (14)	2.52 (29)	3.67 (1)
	With ages	1.86 (21)	2.49 (22)	3.67 (1)
GH control	No ages	1.76 (15)	2.56 (21)	3.88 (8)
	With ages	1.83 (16)	3.15 (24)	4.25 (4)

Table 37. Mean ratings of the drawings identified correctly, incorrectly, and by a nonsignificant number in previous judges' studies ((n) = number of drawings).

The above table shows that the mean bizarreness ratings of those drawings previously chosen correctly as clinical were consistently higher than the drawings chosen incorrectly. The reverse was true for the control HFDs where the mean ratings of those drawings identified correctly were consistently lower than those identified incorrectly.

Summary of results

Very good agreement was found among the three judges for the bizarreness ratings in all the conditions, allowing confident use of the mean rating in further analyses.

The overall ratings for the drawings were quite low, considering that the scale had six points, but significant effects were found on repeated measures ANOVAs. The clinical sample HFDs were rated as more bizarre than the CA matched control sample and when the ages were given, the clinical ratings increased but the CA control ratings

remained the same. The clinical HFDs were also rated as more bizarre than the MA matched control sample though the difference between the ages conditions which occurred with the CA controls just failed to reach significance here. The clinical ratings, however, were no different from the GH matched control ratings though the clinical HFD ratings again increased significantly with the addition of ages; the GH control HFDs remained the same.

Correlations between mean ratings and emotional indicator scores of the clinical and control HFDs were inconsistent though more of the clinical drawings than controls correlated significantly. Correlations between the HFDs' bizarreness ratings and chronological ages of the boys were also inconsistent, though more of the ratings of the boy's drawings were related to the boy's ages in the clinical sample than the controls.

The mean bizarreness ratings of those drawings previously identified correctly as clinical were found to be consistently higher than the drawings either chosen incorrectly or in no particular direction. For the control HFDs, those correctly identified were found to have consistently lower bizarreness ratings than the rest.

DISCUSSION

The results of this present study showed that the bizarreness ratings of the clinical HFDs were significantly higher than the CA and MA matched control HFDs but not significantly different from the GH matched control sample. The judges considered the clinical pictures as more bizarre than the controls, except for those controls that had the same GH scale scores.

The addition of ages changed the bizarreness ratings for some drawings. The clinical HFDs' ratings tended to increase whereas the control samples' remained the same. This suggests that the clinical drawings could possibly be seen as the work of a younger child, although still considered somewhat bizarre, and it is not until the judges see the real ages of the children that they significantly increase the level of bizarreness. This implies that the bizarreness of the clinical children's drawings is linked to a developmental delay in their drawing

ability. If the drawings were simply deviant, then the judges should have thought them equally bizarre both with and without the ages.

It is also interesting to note that the bizarreness ratings correlated significantly and negatively with the GH scores of the drawings. This means that the drawings that had higher scaled GH scores were seen as less bizarre. This reinforces the idea that the drawings are somehow delayed and not deviant.

There was no clear relationship between the bizarreness ratings and the emotional indicator scores of the drawings. More of the clinical HFDs were related and more of the controls were not, but there was no consistent pattern of results. This could be seen as meaning that it is not the indicators in the drawings that are making them bizarre. However, since some of the drawings' ratings did have significant correlations with the indicator scores this conclusion must be tentative. There may be some reason why only some of the drawings showed this relationship. The clinical HFDs had a higher incidence of indicators overall and this may have affected the correlations, making them higher than for the control samples' HFDs. The controls had a much lower incidence of indicators and show a floor effect, with many drawings simply having no indicators at all. The correlations therefore may have been suppressed somewhat by this. The indicators were quite a crude measure, from 0 to 5 at most, with most clinical children scoring 1 or 2. Although the judges did not use the whole of the bizarreness rating scale of 1 to 6, the range of scores was wider than for the indicator scores. This may have adversely affected any correlation and altered the perception of the relationship between the two variables. It is therefore difficult to determine conclusively from these results, whether the ratings of bizarreness were related to the indicator scores or not.

The relationship between the ratings of bizarreness and the chronological ages of the children might have been expected to disappear once the ages were given, but in fact it did not. It might have been expected that the clinical drawings would be considered bizarre at all ages and the control HFDs would not be considered bizarre at any age; therefore there would be no relationship with age. However, there was an irregular pattern of correlations which is difficult to interpret

and may again have been affected by the difference in the scales for the two variables.

The differences in bizarreness ratings between clinical and controls would have been the predicted outcome if the judges were using the level of bizarreness of the drawings in studies 4 and 7 to make their decision about whether a drawing was from a clinical child or normal child. The clinical drawings which had been identified correctly previously (using the 'group format' only) were found to have higher bizarreness ratings than those that had not been identified correctly by a significant number, and those that had been identified incorrectly by a significant number of judges. The reverse was found for the control pictures. This result was found even for the GH controls where the judges had previously been unable to discriminate the clinical from GH controls. Those few clinical drawings that had been identified correctly in the 'GH control' condition were found to have higher bizarreness ratings and those identified incorrectly had lower ratings.

The bizarreness ratings appear to have been more successful than the emotional indicators in explaining the previous performance of the judges. When the scores of the correctly and incorrectly chosen pictures were compared with regard to EI scores, no consistent pattern was found. The judges who were successful previously seem to have been using a global impression of the drawing to make their decisions rather than a feature analysis such as the emotional indicators.

The significant differences between the clinical and control HFDs in bizarreness ratings found here follows the same pattern as the analysis of the emotional indicators. The clinical HFDs were significantly different from the CA and MA control samples but not the GH control sample. This means that there is a difference between the drawings of the clinical sample and CA/MA control samples which disappears when the GH scaled scores are controlled. This difference is related to both the indicator scores and the bizarreness of the drawings as well as the GH scale items.

The question remains whether the difference is connected to a developmental delay or deviancy. The evidence from the involvement of the GH scale suggests a delay since, when drawing development as

measured by this scale is accounted for, indicator score differences disappear, differences in bizarreness ratings disappear and the judges are not successful in discriminating the samples. It is also important to note that the working number of indicators was much decreased when the confounded items from the GH scale were removed. This has further exaggerated the normative data study results that had reduced the 30 Koppitz indicators to the list of 23 revised indicators, valid for boys. It is also important to note that the delay in drawing measured by the GH scale is not necessarily caused by EBD since the GH control sample used were normally adjusted children.

There is a suggestion that the indicators and bizarreness could be related, and it is not completely clear what it is which separates the clinical from control HFDs and which factor the judges were using to make the decision of sample membership. It is necessary to separate these two and determine whether a drawing without the indicators is still considered either clinical or bizarre: if it were, then the indicators could be seen as unnecessary and invalid; if it were not, then the two factors may be inextricably linked.

Study 10

Manipulating the human figure drawings

INTRODUCTION

The pattern of results for the bizarreness ratings has shown the same pattern of results as the indicators. They show differences between clinical and CA and MA matched control samples but not GH matched controls. This implies that the level of bizarreness of a drawing is related to the number of indicators it contains and these factors may both determine whether or not the drawing is given clinical status.

However, the clinical status of the HFDs recorded using the intuitive method seems to be more related to the bizarreness of the drawings than to the indicators. The HFDs consistently identified as clinical had higher bizarreness ratings, but did not have more indicators. Therefore, it appears that the judges using the intuitive method were assessing the bizarreness of the drawings rather than the

features such as the indicators. It is unclear, however, how the bizarreness rating of a HFD relates to the indicators and it is this which is addressed in this study.

A subsample of drawings is used in this study. These drawings are those extremes of the distribution which are consistently seen as clinical, using the intuitive method, and are rated highly for bizarreness. Drawings from study 6 fulfilling these criteria are chosen first, since the GH scale is implicated in the validity of the indicators. Drawings which fulfil these criteria in this condition are assumed to contain whatever allows the judges to differentiate the pictures. Removing the indicators from these pictures will determine whether the indicators are affecting the level of bizarreness of those pictures, and thus whether it is assigned clinical status or not.

If the indicators are valid for assessing the disturbance of a child through his/her drawing, then without them, the drawings should no longer be considered bizarre. If a drawing remains at the same level of bizarreness without the indicators present, then it can be assumed that the indicators are not useful for assessing emotional disturbance.

METHOD

Participants

Five participants acted as judges in this study. The participants were students at the University of York and were aged at least 18 years. They were all females.

Design

Pilot data suggested that removing varying numbers of indicators from the drawings had no effect on the ratings of bizarreness, so all the indicators were removed for this study. Two versions of each picture were prepared with either the Koppitz or the revised indicators removed separately. This allowed for comparison with other studies which used the Koppitz indicators, whilst also acknowledging the differences in the normative data shown earlier in the thesis.

The drawings used were those which had been judged to be drawn by disturbed children and had higher than average bizarreness

ratings. Drawings were selected from the clinical and GH matched control samples first.

Three forms of each drawing were obtained - an original, one version with the Koppitz indicators removed and one version with the revised indicators removed. A within-subject design was used whereby all the judges saw all forms of each drawing.

Materials

Thirty six drawings were compiled. Fourteen originals (9 clinical, 1 CA match, 1 MA match and 3 GH matches), 14 with the Koppitz indicators removed and 8 with the revised indicators removed. For example, items omitted (e.g. 'hands cut off', 'no nose', 'no neck') were added to the drawings; proportions were corrected for 'tiny figure', 'big figure', 'tiny head', long arms' and 'short arms'; items such as 'shading body/limbs', 'teeth' and 'clouds' were erased; 'poor integration' and 'asymmetry' were also corrected. For three drawings, removal of the Koppitz indicators automatically removed the revised indicators too. Three drawings did not contain any revised indicators to remove.

Procedure

The judges were asked to rate the randomly mixed pile of 36 drawings for bizarreness, using the scale of 1 to 6 described in the previous study. The ages of the boys were provided for the judges. The following written instructions were given.

"The following is a set of 36 drawings of the human figure from boys between the ages of 7 and 11 years old. Your task is to rate these drawings for bizarreness on a scale of 1 to 6 whereby 1 means the drawing is not bizarre at all and 6 means the drawing is completely bizarre.

1	2	3	4	5	6
not bizarre					completely bizarre
<------decreasing			increasing------>		

Some of the drawings may look very similar to one another but they are all different. You should treat each drawing separately and once a drawing has been rated, do not go back and alter your decision."

RESULTS

The results of this experiment are ratings of bizarreness on a scale of 1 to 6 for each of 36 drawings from five judges. The ratings of the original HFDs are compared with the ratings of the HFDs without the Koppitz and revised indicators separately. Results are shown in table 38.

Inter-rater reliability for the ratings of bizarreness was assessed using Kendall's Coefficient of Concordance which showed good agreement across the five judges ($W=0.09$, $X^2=14.3$, $p<0.01$). This allowed for reliable use of the mean rating from the five judges.

For scaled versions of four examples of these pictures in their three different forms, see figures 8-19. HFD C2 is shown in its original form (fig. 8); with the asymmetry, shading of face and poor integration removed (fig. 9); and with asymmetry and poor integration removed (fig. 10). HFD C22 is shown in its original form (fig. 11); with shading of face and body, clouds, arms clinging and big figure removed (fig. 12); and with arms clinging and clouds removed (fig. 13). HFD C31 is shown in its original form (fig. 14); with poor integration, arms clinging and big figure removed and neck added (fig. 15); and with arms clinging and poor integration removed (fig. 16). HFD GH44 is shown in its original form (fig. 17); with crossed eyes and short arms removed and hands and neck added (fig 18); and with crossed eyes removed and hands added (fig. 19).

	Mean Bizarreness Rating		
HFD	Original	No Koppitz EIs	No revised EIs
C2	2.8	2.8	2.8
C7	3.4	3.6	3.2
C12	4.4	4	4*
C18	4.6	4.2	4.2*
C22	3.4	3	3.6
C28	3.8	3.8	4
C29	4.4	4.8	4.4
C31	3.4	3.6	2.6
C33	4	4.2	4.2*
CA28	3.6	2.8	-
MA13	3.6	3.6	3.6
GH14	5.2	4.8	-
GH25	4.6	4	-
GH44	4.8	5.2	5
Mean	4	3.89	3.78

Table 38. Mean ratings of bizarreness for the 36 drawings.

'*' indicates where the removal of the Koppitz indicators was the same as removing the revised indicators and therefore only one alteration to the drawing was necessary and the rating is simply repeated in this column.

'-' indicates where there were no revised indicators present in the drawing, therefore this version was omitted and no rating possible.

It can be seen in the above table that the ratings of bizarreness do not alter substantially when the Koppitz or revised indicators are removed from the drawings. The mean rating for the original version of the drawing is the same as the rating when the Koppitz or revised indicators are removed. The rating of the judges was not altered across the presentation of the three different versions of the HFD. This result was confirmed by t-tests showing no differences between the original and 'no Koppitz' versions ($t(26)=0.43$, ns) or between the original and 'no revised' versions ($t(23)=0.78$, ns).

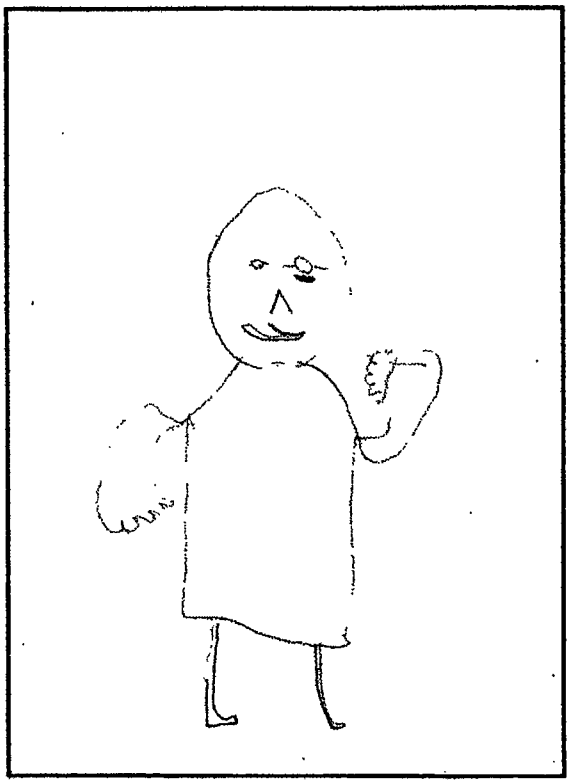

Figure 8. HFD C2, age 9;2, original

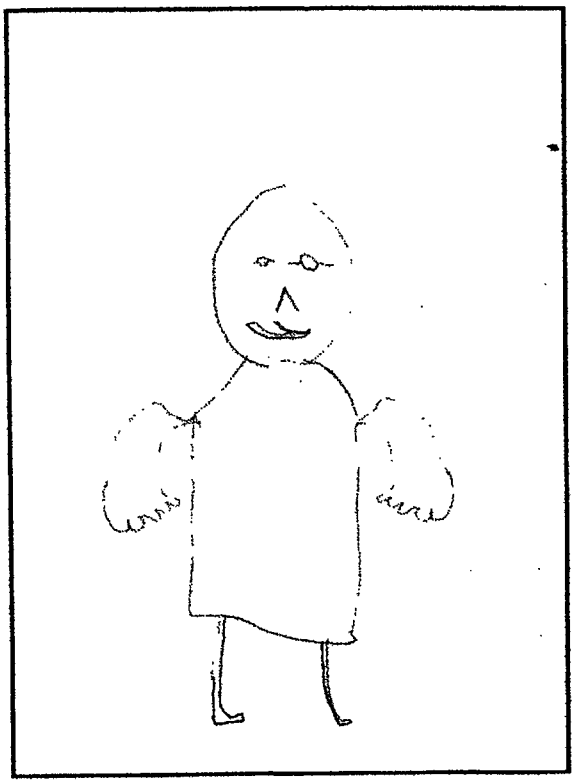

Figure 9. C2, no Koppitz EIs

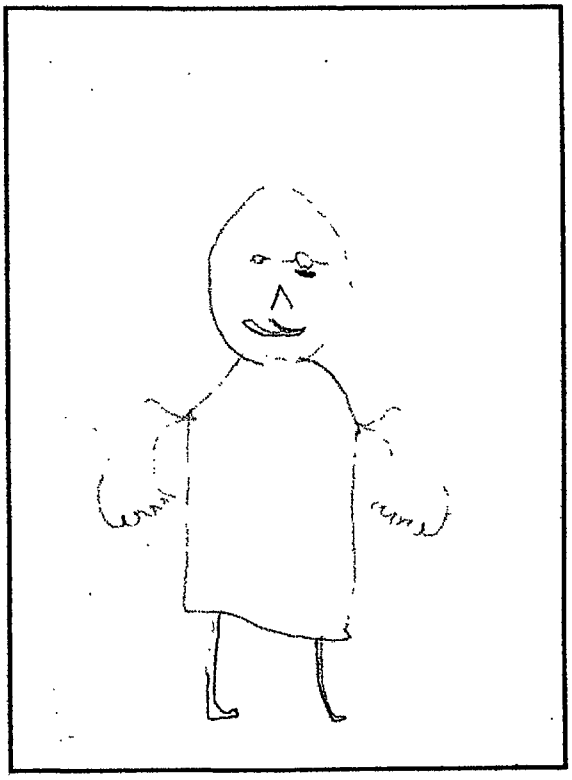

Figure 10. C2, no revised EIs

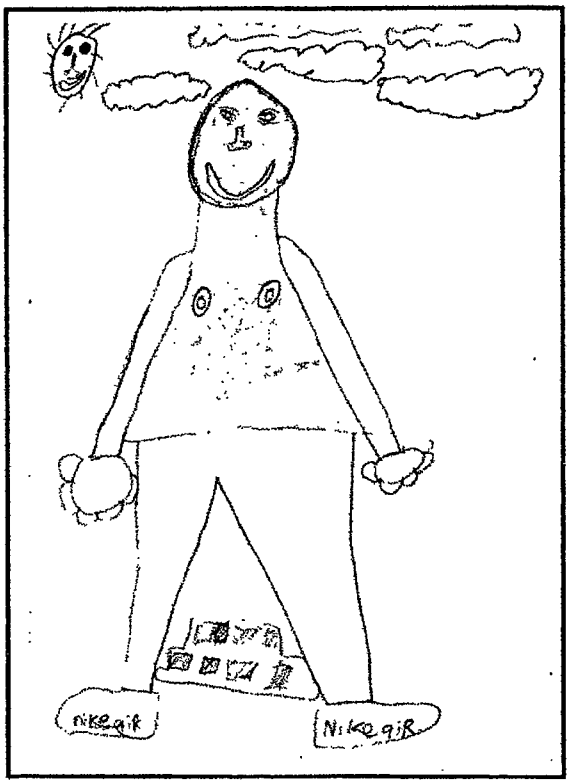

Figure 11. HFD C22, age 10;2, original

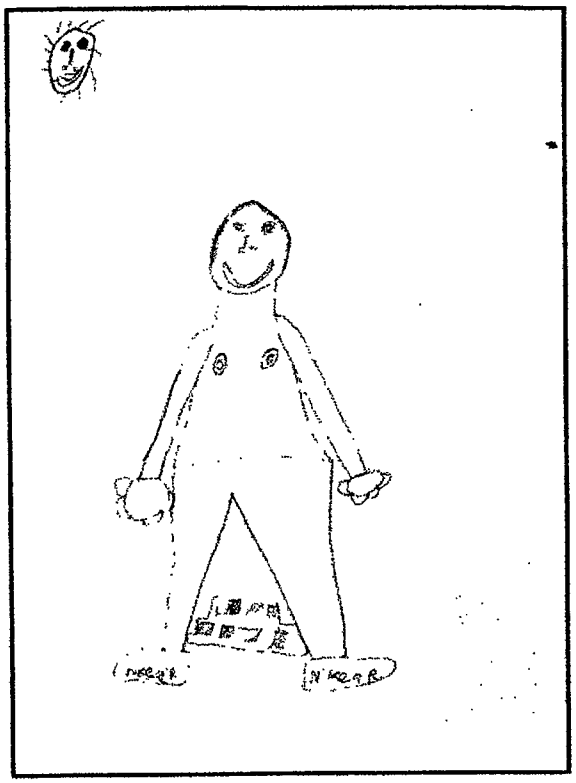

Figure 12. C22, no Koppitz EIs

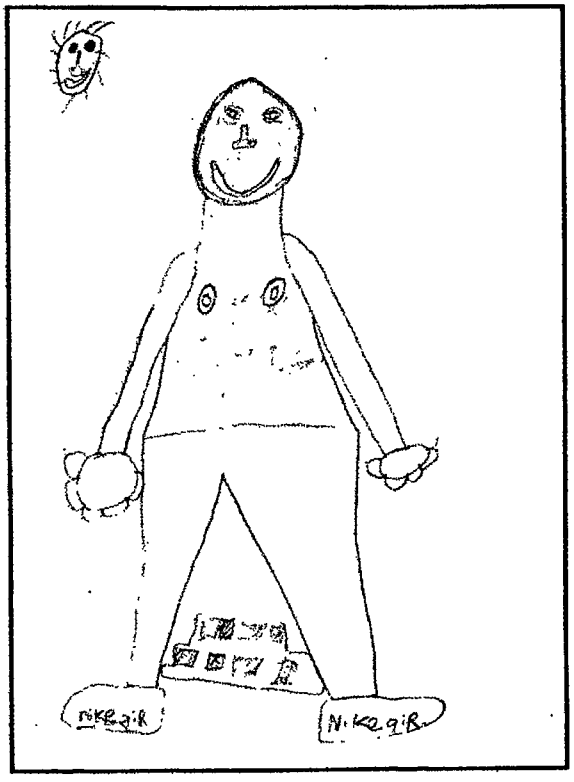

Figure 13. C22, no revised EIs

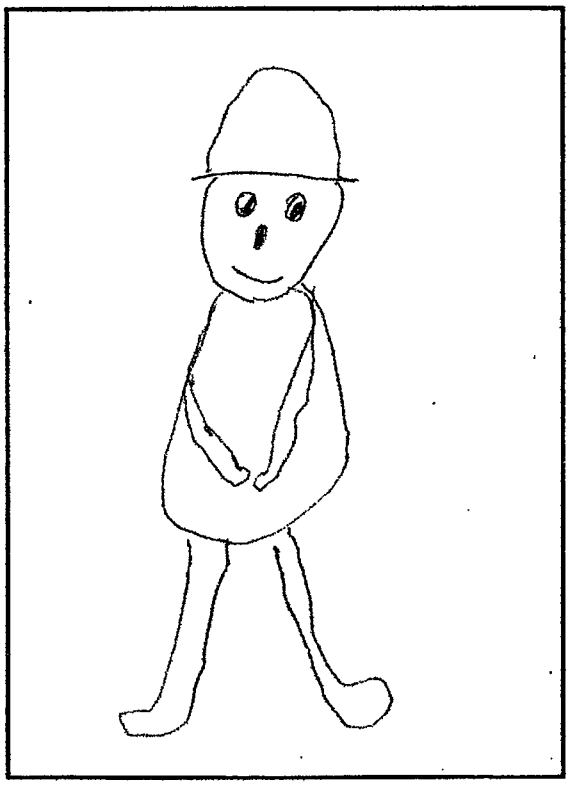

Figure 14. HFD C31, age 10;6, original

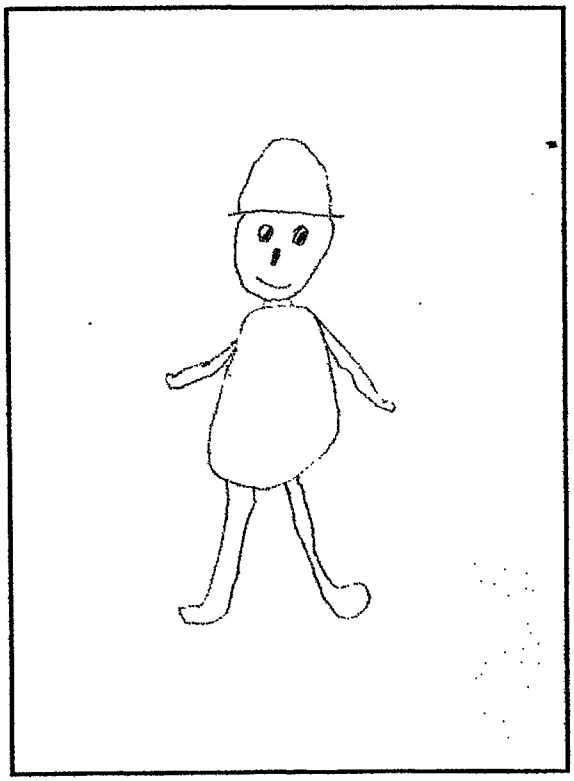

Figure 15. C31, no Koppitz EIs

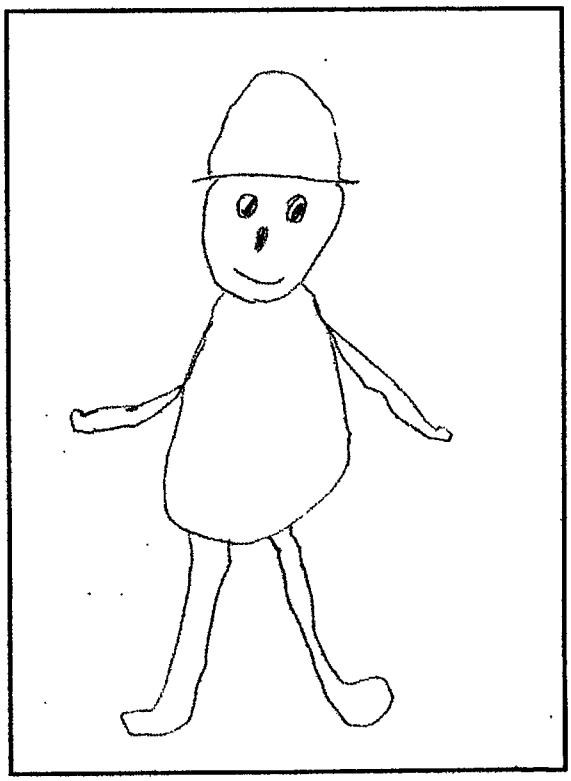

Figure 16. C31, no revised EIs

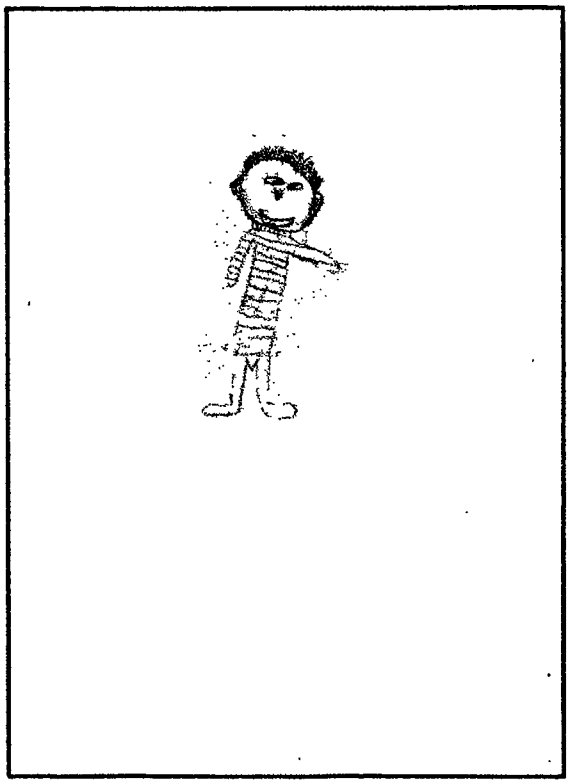

Figure 17. HFD GH44, age 11, original

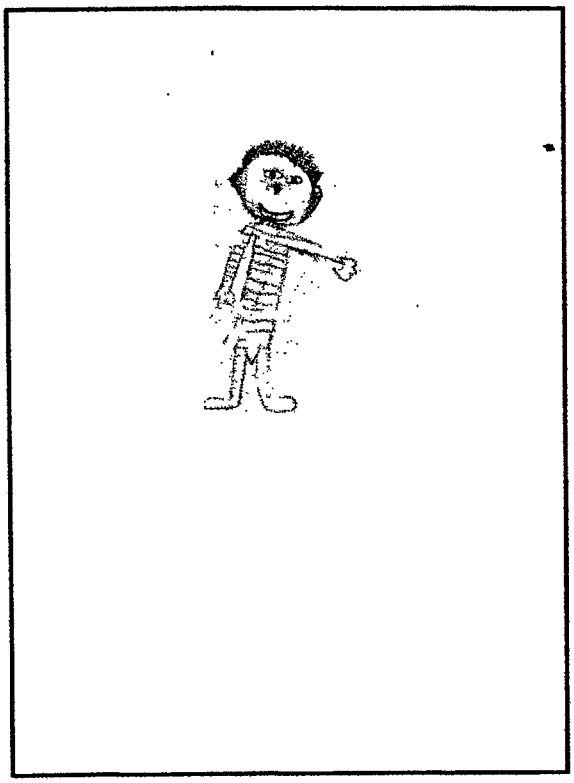

Figure 18. GH44, no Koppitz EIs

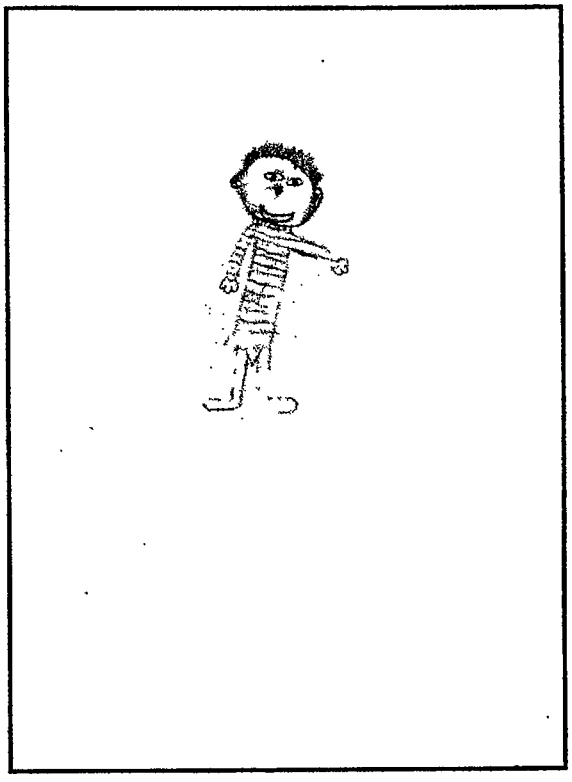

Figure 19. GH44, no revised EIs

DISCUSSION

This study attempted to discover the relationship between the indicators and the global rating of bizarreness. It was expected that the ratings would decrease when the indicators were removed if they were causing the bizarreness and were valid for assessing disturbed boy's HFDs. The results show, however, that the removal of the indicators fails to alter the ratings of bizarreness. This means that the indicators do not affect this global impression of the drawing and casts doubt on their validity.

This means that where the judges were successfully using the intuitive method, they were using the level of bizarreness of the HFDs rather than the indicators. This tends to suggest that the indicators are unnecessary and irrelevant to the interpretation of HFDs, compared with bizarreness and the intuitive method.

The absolute levels of the bizarreness ratings here are lower for some of the drawings in their original forms than they obtained elsewhere (study 9, this chapter). This is due to the smaller number of drawings that the judges had to assess and the general unwillingness by the judges to use the upper end of the scale. The consistency in ratings among the three versions of each drawing remains, however, whatever the rating of the drawing in its original form.

The drawings were randomised to avoid the original form of the drawing always being seen first and affecting the subsequent rating of the drawing in its other two forms. The within-subject nature of this design, however, may have resulted in the similarity of the ratings due to the similarity in the pictures, though the judges were told to treat each picture as if it were the first time it had been seen.

Since the indicators do not affect the drawings' level of bizarreness it can be concluded that they are of limited relevance to the differences which the judges perceive between the drawings. It is still necessary, therefore, to determine what the differences are between the drawings seen as bizarre and given clinical status, and those that are not bizarre and seen as normal. This should clarify what the judges are identifying in the picture when they assign a drawing clinical status and rate it as highly bizarre.

Study 11

Determining the differences between the human figure drawings

INTRODUCTION

The results from the previous study showed that when the emotional indicators are removed from the drawings, it does not affect the rating of bizarreness. It may be concluded, therefore, that the indicators in a drawing do not affect the level of bizarreness of that drawing. Since the bizarreness is related to the clinical status of a HFD using the intuitive method, this implies that the indicators are not related to the assumed clinical status of the drawing.

It is important, therefore, to determine what the judges saw in the drawings which might be responsible for the decisions they were making regarding the HFDs' clinical or normal status and ratings of bizarreness. If the indicators are not responsible for the level of bizarreness of a picture, then it is necessary to determine what is responsible.

Previous researchers have attempted to determine differences between drawings using systems other than the items listed by Machover and Koppitz. A special needs teacher using the intuitive method (Dieffenbach, 1977) discriminated the drawings he collected better than the Koppitz indicators. The criteria or strategy used to make the decisions, however, was not investigated.

A qualitative system devised by Tharinger and Stark (1990) was also found to be more successful than the quantitative approach of the emotional indicators in discriminating anxiety and mood disordered children from controls. The 'DAP integrative system' was constructed by sorting the drawings on a scale of 1 to 5 for psychological functioning then interviewing the raters to determine the criteria which they had used. Four characteristics were found including 'inhumanness of the drawing', 'lack of agency', 'lack of well-being of the individual in the drawing (facial expression of negative emotion)' and 'presence of hollow, vacant stilted sense in the individual portrayed'. The overall rating of the drawing was obtained through an integrative combination of the four characteristics.

Ratings of bizarreness have been used alongside ratings of adjustment and artistic quality in order to discriminate HFDs (Yama, 1990), though Yama did not attempt to determine what these variables meant. It is unclear, therefore, whether the bizarreness rating identified by Yama as a significant variable relates to the bizarreness in this thesis. By elaborating on what the bizarreness means to the judges, it clarifies the validity of the measure.

The intuitive decisions of clinicians were formalised by Hiler and Nesvig (1965) in an attempt to determine the criteria used when identifying a drawing. They also found that bizarreness was most successful though they also identified factors such as incompleteness, distortions and transparencies which were useful to the clinicians. Stricker (1967) found that the formula devised by Hiler and Nesvig was more successful than clinicians assumed to be relying on intuition alone.

Adler (1970) factor analysed 17 scales and 40 items from the literature to establish the one that accounted for the most variance in drawings from adult psychiatric patients. He concluded that it was a cognitive factor related to the "formal accuracy of the drawn figure and degree to which the figure is differentiated with regard to detail and to individuality" (p. 55). It is not clear how relevant this factor is in children's drawings, though Adler saw it as something that develops in children with age and cognitive maturity. The GH scale may be seen as a measure of this as it includes detail and proportion and the overall technical quality of the drawing.

The present study aims to investigate what differences the judges see in the drawings to examine whether any similar concepts to those of Adler or Tharinger and Stark are found. The original drawings which were considered clinical and rated highly bizarre (labelled 'disturbed') as well as a comparable set which were considered normal and rated as not bizarre (labelled 'normal') were used. Visual inspection of the drawings by judges was used to determine any differences that may be apparent. Written comments on each drawing as well as verbal discussion with the experimenter were used to investigate what distinguished the disturbed drawings from the normal ones.

METHOD

Participants

Twelve people taken from the student population at York University were used as judges in this study. All participants were aged at least 18 years. There were 5 males and 7 females.

Materials

Two sets of 14 drawings were compiled according to two criteria:

Disturbed = the HFD must have been given clinical status by a significant number of judges using the intuitive method and have been rated higher than average for bizarreness (the average is defined as the mean for the clinical and control HFDs' ratings together).

Normal = the HFD must have been given normal status by a significant number of judges using the intuitive method and have been rated lower than average for bizarreness.

Drawings were sampled from the clinical and GH matched samples in the first instance, before the CA and MA matched samples were used.

SET 1 (Disturbed) - 9 of these drawings were from the clinical sample, 3 were from the GH matched samples, 1 from the CA matched and 1 from the MA matched samples. These were all the possible drawings using the above criteria.

SET 2 (Normal) - 3 of these drawings were from the clinical sample, 11 were from the GH matched sample.

Procedure

Judges were given the two sets of drawings with the following written instructions

"You will be given two sets of drawings to look at. The drawings are all from boys (ages are specified on each drawing) who were asked to draw 'a whole person'.

SET 1 (d1-d14) These drawings have all been judged previously as from disturbed children. They have also been rated as more 'bizarre' than average.

SET 2 (n1-n14) These drawings have all been judged previously as from normally-adjusted children. They have also been rated as less 'bizarre' than average.

You are asked to look through the drawings and attempt to determine any factors which you think differentiate the two sets. There is no right answer to this and as much information as possible is needed.

Please note any details you think relevant to each drawing on the forms provided. This may include aspects of the drawing which you think are strange or odd. It may also include any particular features about the drawing which you think are different from the other set which you have looked at, as well as any similarities the drawing may have with others in the same set.

Take your time and look through the two sets once or twice before making any decisions. A brief discussion will take place afterwards in order to clarify any details you have noted."

Judges were required to look through the two sets of drawings in order to identify what differentiated them and noting any features which they thought were relevant to the drawings' group membership. The judges filled out comment forms for each of the pictures and underwent a discussion with the examiner about the drawings in general.

RESULTS

Results for this study were the written comments that the judges made for each drawing and the notes from discussion with the examiner. The comments that were made by the judges for the drawings in each set are summarised here; numbers in brackets refer to the number of judges who made the comment. Figures 20 to 48 show scaled versions of the drawings.

Disturbed set

D1 (Fig. 20) This drawing was described as 'lacking detail' (4), with 'no pupils' (4), 'no clothes' (3), 'no hair' (3), 'no ears' (3). Comments were made about the limbs: 'large arms but small skinny legs' (2), 'mis-shaped', 'arms and hands are not typically drawn', 'twisted', 'very strange arms', 'arms are disproportionate', 'arms in a funny position'. Comments were also made about the face: 'mark below the eye' (6), 'facial features disorganised', 'the face is not clear', 'expressionless'. The judges also noted aspects of the shape and proportion of the figure:

'very mis-shapen', 'deformed figure', 'very square body', 'large head', 'the body is out of proportion'. One judge thought that the child was 'not talented or unable to draw a person'.

D2 (Fig. 21) This drawing lacked many things: 'no fingers' (3), 'no feet' (2), 'no eyes' (3), 'no ears' (3), 'no clothes' (3). Comments were made about the impression of the figure: 'very angry/aggressive figure', 'angry looking face', 'the character is 'growling'' and 'looks angry' (3), 'evil expression', 'teeth' (2) make it look quite 'scary', 'monstrous teeth', 'fangs', 'spiky looking teeth', 'may represent a figure that is scary for the child', 'insecurity', 'fright', 'violently drawn hair', 'not smiling'. Other comments regarded the form of the figure: 'basic shapes' (2), 'dealing with shapes rather than images', 'no definition in limbs', 'just sausage shaped arms and legs', 'very simplified limbs', 'some attempt at showing hair but not much', 'very large head', 'deformed figure', 'simple' (2), 'minimal detail' (2), 'very out of proportion', 'very square'.

D3 (Fig. 22) Judges commented on the 'lack of facial features' (9). Comments were made about the limbs: 'strange limbs', 'almost as though the character has wings', 'child might want to fly', 'imagine himself as a bird', 'arms exaggerated', 'obscure hands', 'feet/shoes abnormally drawn', 'provides detail in the form of boots', 'no real detail apart from large hands and boots', 'arms attached to side of body', 'only 3 fingers' (2). Judges noted the proportions of the figure: 'out of proportion hands and feet', 'arms not proportional, neither are boots', 'not in proportion' (2), 'deformed figure'. The lack of detail and clothing were noted and one judge thought that this was 'not a very good drawing for a 9 year old'.

D4 (Fig. 23) This drawing was seen to lack hair (3), ears (2) and nose (2). The drawing was considered 'very small' (5) with 'long legs' (4), 'a very small head' (4) though it was 'smiling' (3). Judges noted that there was 'some detail', with 'more detail on the face and clothes' than other pictures. The body was 'well drawn', 'the fingers and hands not sufficiently defined' (2). The 'outstretched arms may appear

threatening', but also 'looks like he is flying'. One judge commented that 'the drawing does not appear very bizarre'.

D5 (Fig. 24) The judges noted that the drawing was 'not very detailed' (2). Comments were made about the arms: 'practically no arms', 'helpless', 'little' (2), 'stick' (4), 'pencil-like', 'interesting', 'spiky', 'strange', 'not defined', 'fishbones or feathers', 'attached to side'. Comments were also made about the body: 'strange/square body shape' (8), 'triangular waist', 'strange unlife-like proportions', 'deformed figure', 'big body and small stick legs', with 'undefined feet'. The drawing had a 'large head' (2), 'no ears' (3), 'no eyes' (4), 'no hair' (3), 'no teeth' and 'a very empty face', though one judge noted that 'the smile suggests the character is happy'. The judges also noted that 'the figure does not really look like a person', but 'more like an upturned house with a big chimney', or 'like a snowman'. The drawing was seen as 'very childish' and 'very basic for the child's age'; one did not see it as 'bizarre', another commented that it was 'a weird way of seeing a whole person'.

D6 (Fig. 25) The judges noted the lack of facial features (7), 'long neck' (8) and 'tiny head' (4). The definition and shape of the figure was commented on: 'no limb definition' (2), 'simple round body' (2), 'spread out arms', 'a huge body', 'out of proportion' (2), 'very basic', 'deformed figure'. The figure was seen to lack detail (3), hair (2), clothing (2) and ears. It has 'only 2 or 3 fingers', 'missing details such as fingers or toes/shoes'. Several judges also noted that 'the figure does not look like a person' (3), 'more like a turtle', 'is strange for a person' and 'looks like it's been flattened'.

D7 (Fig. 26) The judges noted the lack of clothing on the chest (3) and the added 'shading' (3) and 'nipples' (7). The judges commented on the 'smile' (3) but also the impression of the drawing: 'threatening', 'looking really sinister and evil', 'looks like a bully', 'monstrous', 'huge', 'overbearing', 'might think himself strong, ruling the world', 'very large person'. Comments were made about the hands: 'strange hands like sunflowers', 'hands look violent', 'disproportionate hands/fingers'. The judges noted the 'shaded face' (3), the 'eyes missing pupils' and the

'faint ears'. The judges noted that 'the figure was more detailed' (2), with 'added scene detail' (2), 'brand label on shoes' (2), and 'a human face on the sun as well'. One judge commented on the 'big difference in quality between this and younger pictures', which was seen as 'weird for a 10 year old'.

D8 (Fig. 27) The judges commented on the 'knife' (11) and 'shield' (5) which one noted as 'transparent' and another saw as related to a 'soldier' characterisation. The drawing had 'no ears', 'no neck', 'no facial expression' (3), 'no smile' (4) but the child had 'put in lots of detail', 'strange crosses on eyes' (3), 'added scene detail', and 'hair'. The impression of the drawing was also commented on: 'looks worried', 'emphasised teeth' (2), 'scary' (2), 'aggressive', 'violent' (2), with a 'scribbly drawing style'. The judges also noted the 'large head' (4) and' spaced out features', 'shapeless rectangular body', and one considered this drawing as 'quite normal'.

D9 (Fig. 28) The judges commented on the 'high ear position' (5) and 'undefined limbs' (6) and 'small size' (2). One noted that 'the arms are too long'. They also thought it was 'simple for the child's age' (6). Judges noted that the figure is 'smiling' (3), and has 'a wide-eyed look'. The drawing lacked a neck and clothes but ears, hair and facial features were included. The judges commented that the figure 'looks like a cartoon character', 'does not appear very bizarre', but 'is more like a mouse than a person'.

D10 (Fig. 29) The judges noted that this figure was 'very detailed' (3) though had 'no clothes' (3) apart from the presence of 'underwear' (2). They commented on the figure's form: 'like a wooden puppet' (2), 'frankenstein-esque', 'cut and paste figure', 'like a model', 'like action man', 'showing bones and joints' (2), 'segmented body parts' (3). The impression of the figure was noted: 'worried and upset', 'sad expression', 'expressionless', 'not smiling' (2). The drawing was also described as 'not too bad', and 'doesn't seem strange'.

D11 (Fig. 30) The judges noted that this was a 'nude' (4) 'woman' (3) with 'breasts' (4), a 'square body' (5) and 'feet pointing inwards' (5). The figure was 'smiling' (2) and 'happy' with 'lots of detail'. The judges thought 'the head quite good', and the 'face very detailed'. Other comments were: 'the hand/fingers not defined', 'fingers are black', 'legs not proportioned', 'hairy', 'arms don't fit at the shoulders'. One judge thought the drawing 'looks strange'; another saw it as 'rather normal' since 'it captures an image rather than a non-specific person'.

D12 (Fig. 31) This drawing was seen as lacking in detail (7), though there was 'detail on the face' and 'a hat' (3). The impression of the figure was noted: 'shy person', 'looks sad', 'smile', 'timid look', 'happy'. The drawing was seen as 'basic' (2), 'very round' and lacking in hair (2), ears (2), hand/fingers (2), feet/shoes. One judge said the drawing was 'not very bizarre', another 'like a snowman' but it was also seen as 'not good for a 10 year old' (2).

D13 (Fig. 32) The judges saw this figure as a 'tiny figure' (10), with 'stick limbs' (3) and 'no detail' (4), though it was seen as 'smiling' (2) and 'happy'. The drawing lacked clothes, but did include a neck. Comments were made that the drawing was 'childish', 'looks like a 3 or 4 year old has drawn it', 'is misproportioned' and 'unfinished'. It was seen as looking 'like a bird' or' a snowman'.

D14 (Fig. 33) The judges noted that the drawing lacked hands (5) and also commented on the 'clothing detail' (6) and 'very long body' (2). Comments were made about the face: 'dark', 'hirsute', 'unclean', 'hairy' (2), 'unrealistic', 'unfriendly', 'confused', 'nasty', 'jumbled', 'cross-eyed', 'beard seems out of place', though it was 'smiling' (3). The impression of the drawing was that it 'looks like a werewolf or bear', 'a strange drawing', 'looks scary', 'looks demonic', 'odd'. The judges noted that hair and ears were present, but also commented on the 'additional tail between the legs' (2), 'the fact that the legs and feet were in one' and that the drawing was 'out of proportion'.

Normal set

N1 (Fig. 34) Judges commented on the figure's impression: 'cheerful', 'looks friendly', 'smiling' (2), 'waving'. They also noted the level of detail - 'detailed figure' (6), 'lots of detail on clothes', 'attention to detail (drawstring, etc.)', 'detail level similar across picture'. Comments were made about the character: 'looks human', 'based on idol?', 'sense of someone being something, doing something rather than anonymous person', 'not just body', 'a more definite shape of a person', 'normal for a child's perception'. A theme was also noted: 'depicted as footballer' (2), 'typical figure for a normal boy to draw', 'healthy childhood obsession with football and marketing', 'clothed in normal sportswear', 'sporty look', 'pays attention to clothing' (3). Other comments were that 'all parts of the body are in the correct place', with 'more or less correct proportions' (5), 'a small head (or large body)', and 'the fingers were not drawn clearly' but it was seen also as 'a good picture'.

N2 (Fig. 35) Comments were made about the detail in the picture: 'attention to detail' (2), e.g. 'ear-rings' (2) 'eyelashes', 'necklace', 'lots of detail' (5), 'all facial features shown', 'attention shown to clothes (2) and face', and 'lots of hair'. Judges commented that the figure was 'well drawn', 'proportionate' (3), though also had 'a large head' (2) and 'clubbed feet'. The figure was seen as 'happy' (3), 'well balanced', 'smiling' (5), 'androgynous', 'female', and the child was seen as 'trying to create a specific image'.

N3 (Fig. 36) The judges noted that this figure was 'happy' (3), 'smiling' (5), 'detailed' (5) and 'proportionate' (3). Comments were made about the details: 'tried to get details such as lips', 'clothing drawn' (2), 'detail on clothing', e.g. 'shoes' (3), 'belt'. Judges considered that 'this figure might be a self-portrait' (2) and 'the boy likes sport a lot as he draws a person dressed in this outfit'.

N4 (Fig. 37) This figure was seen as 'detailed' (4) with 'knees' (3), 'eyebrows', 'lips', and is 'proportionate' (3). Comments were made about the impression of the figure: 'no smile' (2), 'size of the person impresses', 'strong personality', 'big shoulders like based on rugby player'. The

figure 'has clothes' (2), 'writing on clothes', 'feet and hands included', 'all facial features drawn', 'eyes more accurate', 'good shape to hands and face', 'clubbed feet', and one noted that the child had 'taken care over the picture'. The figure was 'sporty', 'looks like another footballer', 'something a normal boy would draw' and was 'possibly an idealised self-portrait'.

N5 (Fig. 38) The judges commented that this figure was 'happy' (2) and 'smiling' (5). The figure was 'proportionate' (4) with 'all limbs in correct places' and 'small' (3) which 'may symbolise shyness' (2) or 'insecurity'. The figure 'looks cute' (2), 'normal', 'friendly' and 'is like a cartoon impression', with 'lines very rounded', 'very basic', and 'not well defined'. Judges saw the figure as both 'detailed' (3) and 'lacking detail' (2), 'all essential details are there: eyes, nose 4 fingers and thumb, etc.'. The judges noted that 'the figure has large hands', 'no ears', 'clubbed feet' and is 'clothed' (2).

N6 (Fig. 39) Comments were made about the impression of the figure: 'smiling' (7), 'cheerful', 'natural-looking', 'confident', 'secure', 'appears relaxed', perhaps due to the 'hands in pockets' (2) though one judge commented that 'the kid looks evil'. The judges noted that the figure was 'detailed' (6), 'well proportioned' (4), 'clothed' (2) with 'lots of hair'. Comments were made about the quality of the picture: 'well observed', 'quite good', 'well drawn'. One judge noted that 'the teeth don't look menacing as in previous pictures', another that 'the ears are positioned very high on the head'.

N7 (Fig. 40) The judges commented that this was a 'small figure' (6) which was 'cheerful' and 'smiling' (6). The detail level was noted: 'lacking fine detail' (4), 'only has essential details' e.g. 'nose', 'hair', 'eyes', 'mouth', 'ears' (2), 'lacks hands but has clothes which are coloured in'. The figure is 'in proportion' (2) but is 'not very clear'.

N8 (Fig. 41) This figure was described as 'cheerful', 'friendly looking' 'cute', 'smiling' (5), 'cartoonish' (3), 'quite sophisticated somehow' and 'detailed' (3). The figure is 'clothed' (2), though 'the clothing is not as

clearly marked as in some of the other drawings'. The figure is 'a bit out of proportion' (2) and has 'very large legs/feet' and 'no ears'. Other comments were that 'all the features look normal', 'it is a normal way of drawing another person', but 'is not very talented'.

N9 (Fig. 42) This figure was described as 'cheerful', and 'smiling' (5) but comments were made about the 'odd limbs' (7). Comments were made about the body: 'missed out stomach!', 'very small body' (2) ('which is square') 'no real body'. The figure is 'not proportional' (2), 'very mis-shapen', 'very basic', 'particularly simple' and 'very small' (2). Judges noted that the figure 'has face details' (2), 'includes hair' and 'facial features' plus 'fingers and feet', that 'the child tried to get fingers on hands' and 'the arms/hands are shown as bubbles (not sticks)', and 'the figure has no clothes'. One noted that it was 'a weird looking person'.

N10 (Fig. 43) The judges commented on the 'smile' (7), which was seen as both 'cheerful', and 'leering'. The figure was described as 'quite detailed' (4), 'proportionate' (4) and 'with clothes shown' (2) and 'good facial features' (2). The shape of the figure was noted: 'square face and body', 'quite angular'; as were the missing hands, and integrated legs and feet. The drawing was described as 'a big, confident figure', which 'looks normal', with 'everything in the right place', and an 'absence of anything scary'.

N11 (Fig. 44) This figure was described as 'detailed' (6) and 'proportionate' (2). The judges noted the 'thumbs up' (2) as a 'welcoming', 'positive gesture'. The child had 'given attention to clothing' (2), drawing the figure of 'a footballer' (2), 'wearing sportswear', 'something a normal boy would draw', and had 'attention shown to the face', which was 'happy' and 'smiling'. The judges commented that this was 'a good drawing for this age', and that 'lots of effort had been used'.

N12 (Fig. 45) The judges commented that this drawing was 'very detailed' (11) e.g. 'watch' (2), 'features', 'hairstyle' (2), 'twiddling thumbs', 'shoelaces' and 'specific clothes' (3) and was' proportionate' (4) with a

'smiling', 'simple but realistic face'. Comments were made about the quality of the drawing: 'well-observed', 'skilful drawing', 'talented child', 'really good drawing', 'very well drawn'. The figure was described as 'distinct' and a 'very normal looking person', 'very different from the others'.

N13 (Fig. 46) The judges described this figure as 'happy' (3) and 'smiling' (2). Comments were made about the limbs: 'weird limbs', 'rubber arms', 'missing hand/finger detail' and comments were made about the detail in general (2), in the facial features (4) and ears in particular (2) and in the clothes (2) which were possibly sportswear (2). The judges noted that this drawing was 'a bit strange' (2) and 'a very funny picture', but 'normal for this age' (3).

N14 (Fig. 47) The judges described this as 'a small figure' (5), 'fairly proportionate' (3) and 'not very detailed' (3) though it is 'clothed' (2). Comments were made about the impression of the figure: 'insecurity', 'shyness', 'inferiority', 'confused', 'unsure', 'anxious' although 'smiling' (2). The figure lacked ears (2), a neck and 'has only 3 fingers on right hand' (3). 'All body parts are present', and 'the figure has got the basic shape of a person' as well as 'good facial features' but 'the fingers are not well defined' and 'the figure has clubbed feet'. One judge noted 'interesting attention to the crotch' and the figure was seen as 'bizarre' and 'childish for an 11 year old'.

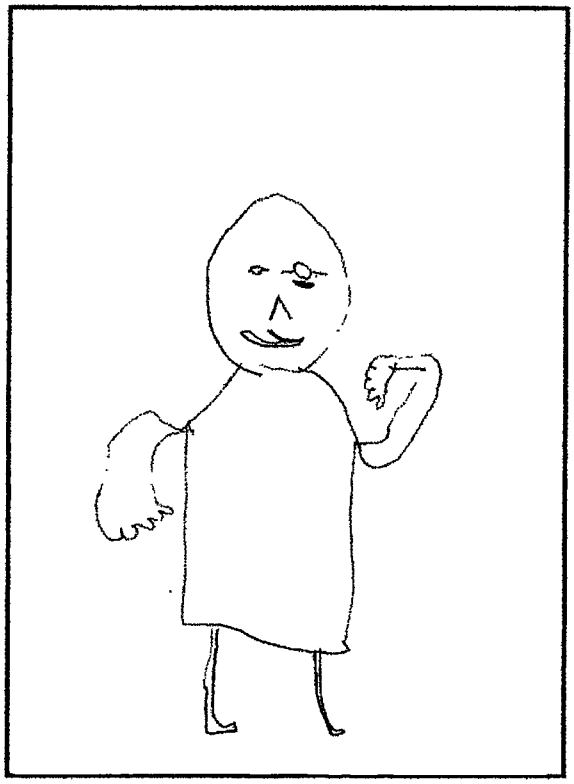

Figure 20. D1 (C2), age 7;7

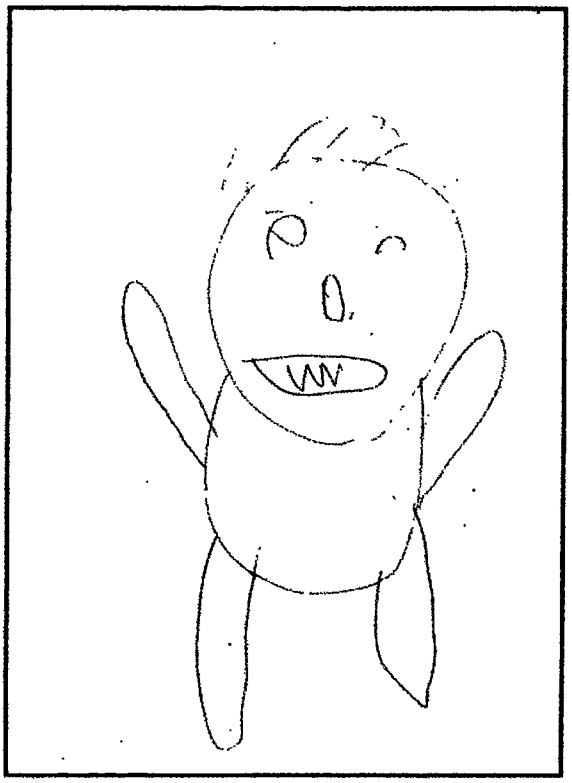

Figure 21. D2 (C7), age 8;9

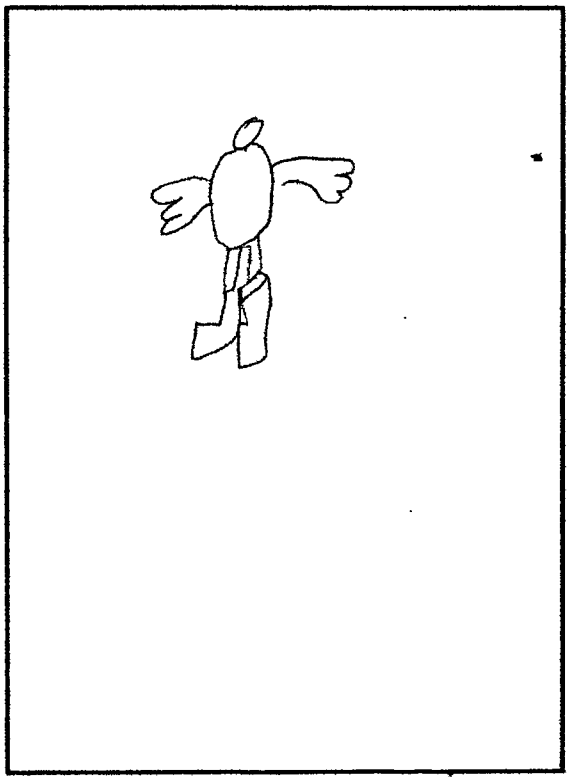

Figure 22. D3 (C12), age 9;2

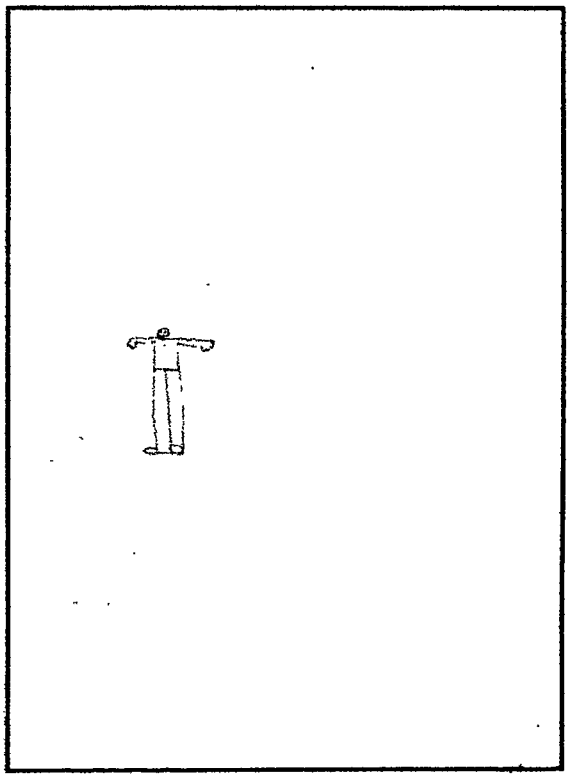

Figure 23. D4 (MA13), age 9;0

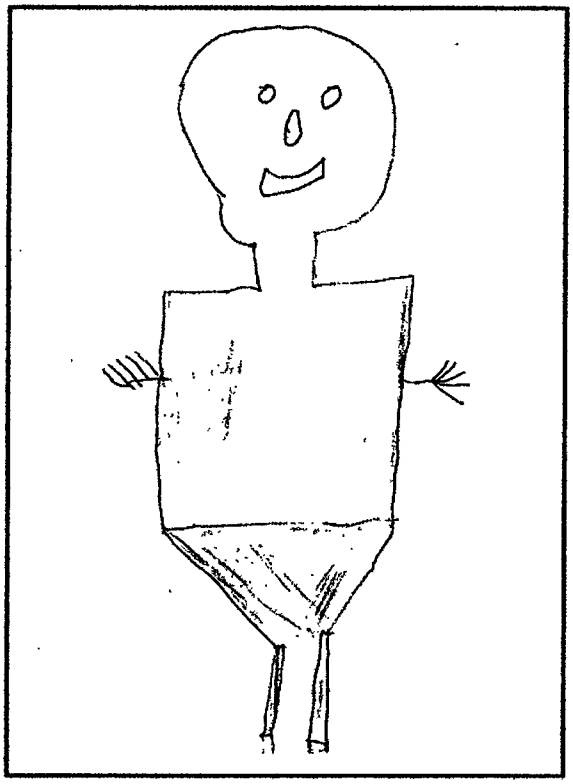

Figure 24. D5 (GH14), age 11

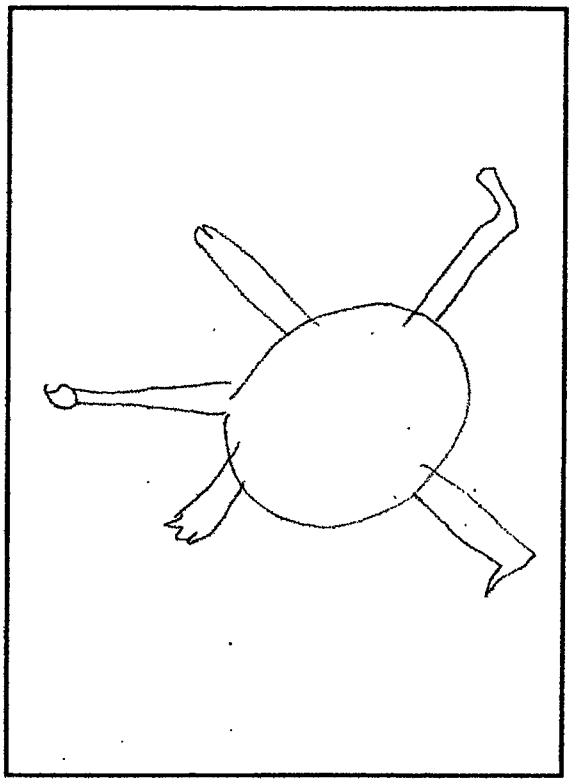

Figure 25. D6 (C18), age 9;7

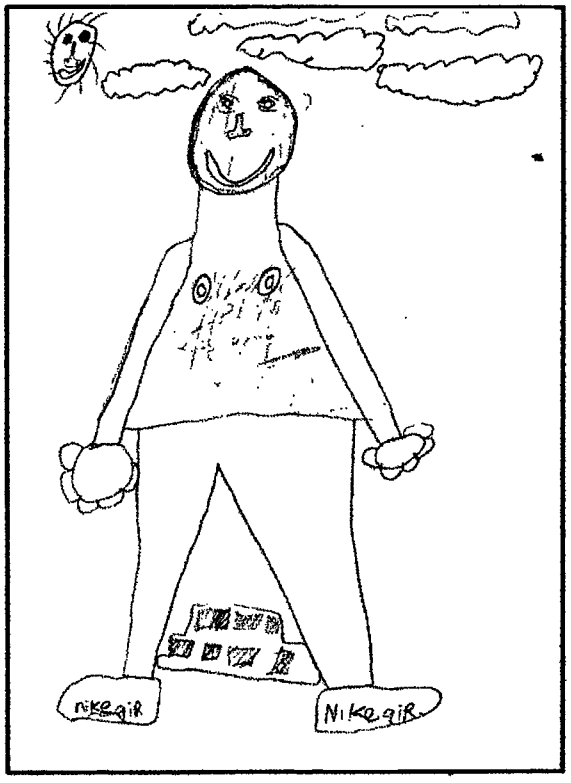

Figure 26. D7 (C22), age 10;2

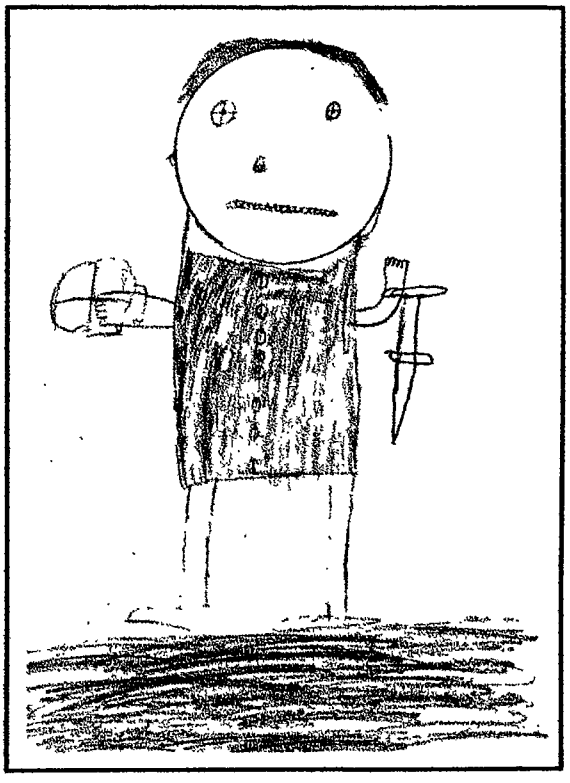

Figure 27. D8 (GH25), age 10

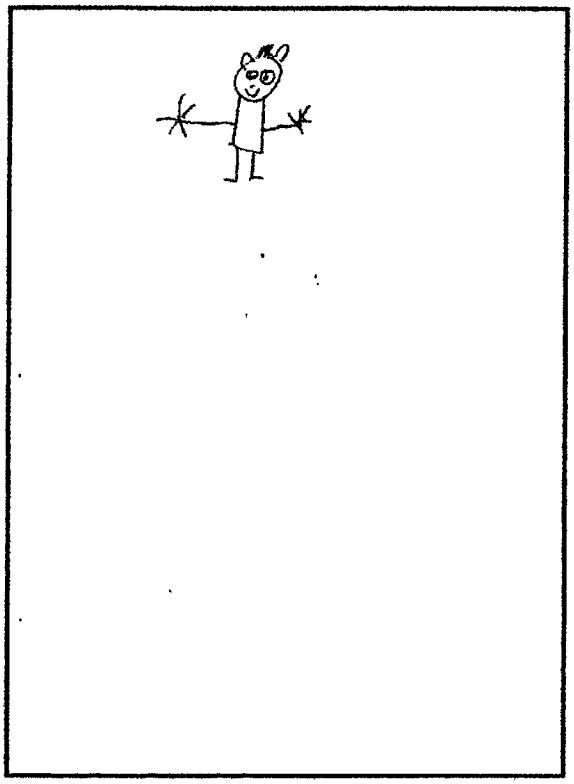

Figure 28. D9 (C28), age 10;4

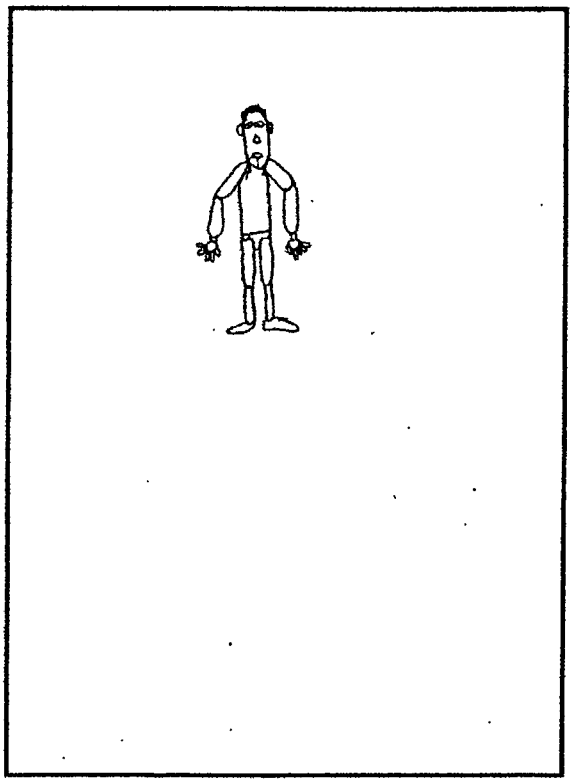

Figure 29. D10 (CA28), age 10;4

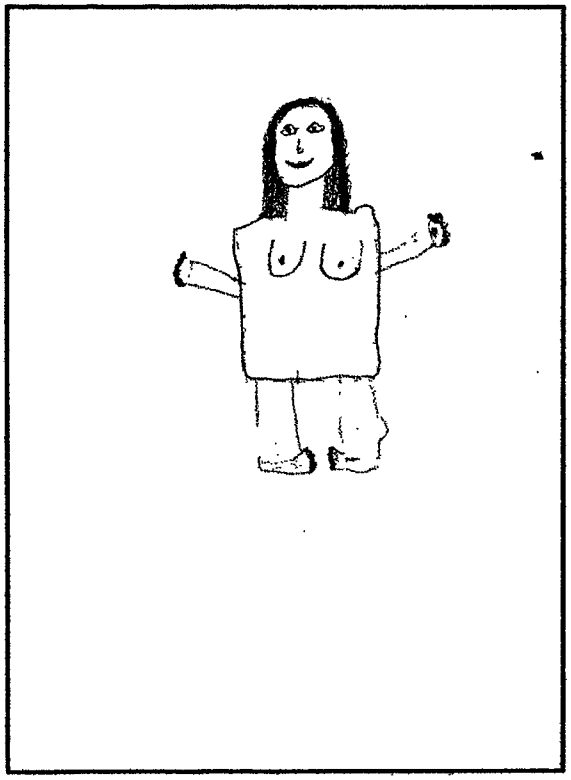

Figure 30. D11 (C29), age 10;5

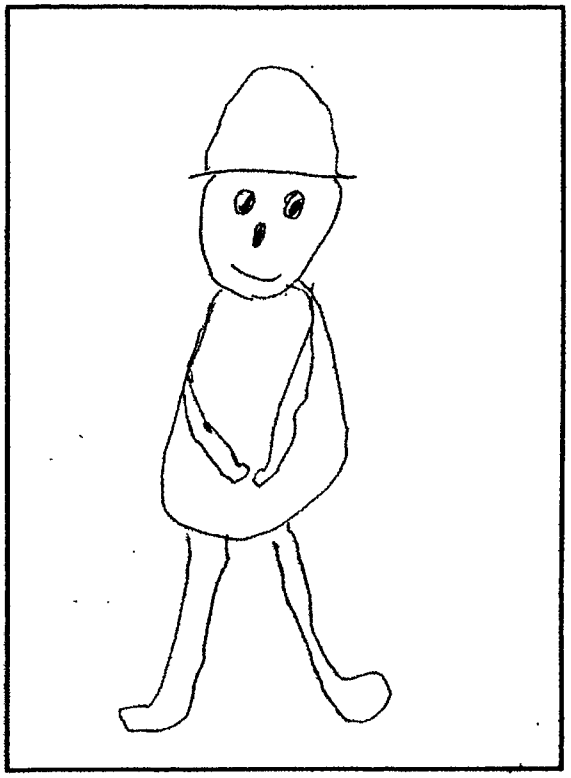

Figure 31. D12 (C31), age 10;6

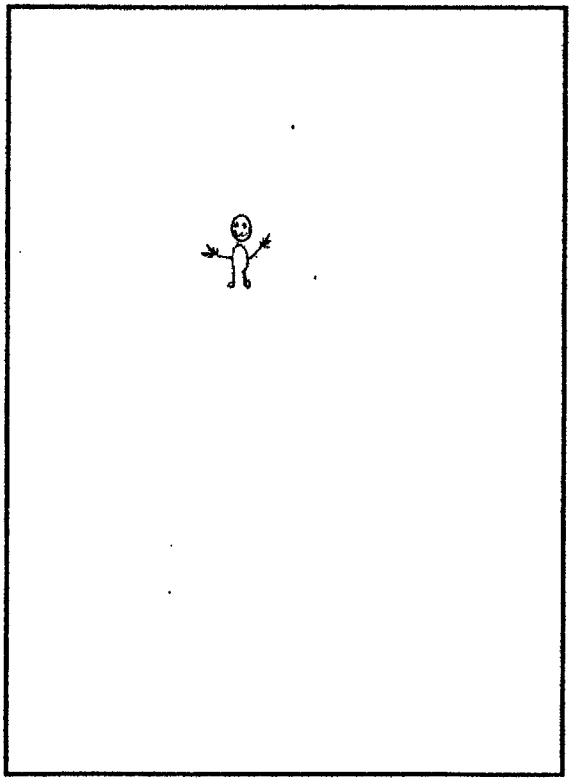

Figure 32. D13 (C33), age 10;8

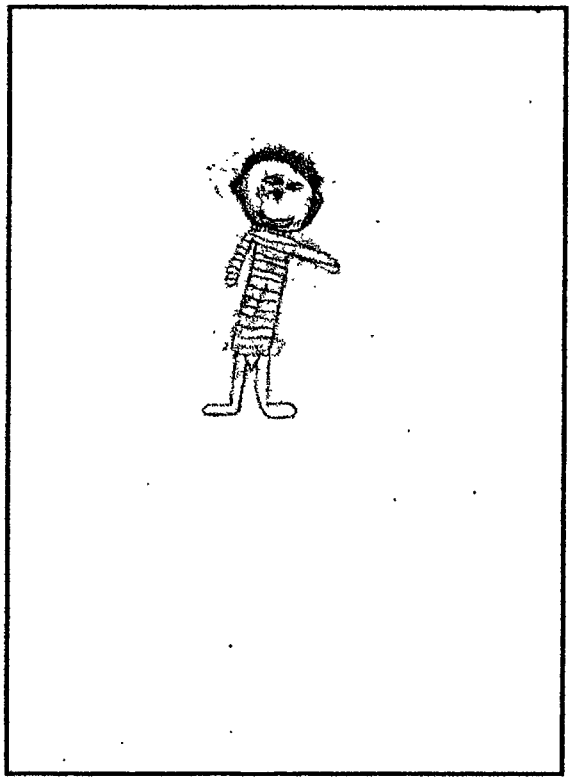

Figure 33. D14 (GH44), age 11

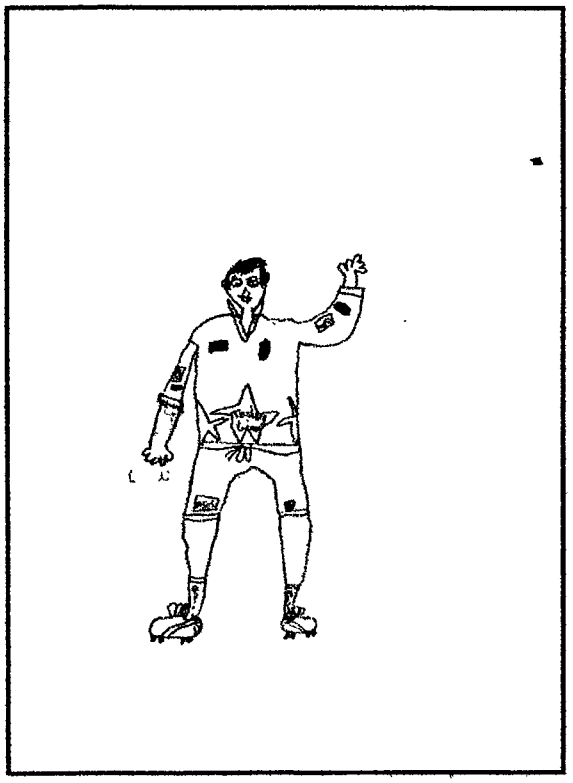

Figure 34. N1 (GH2), age 9

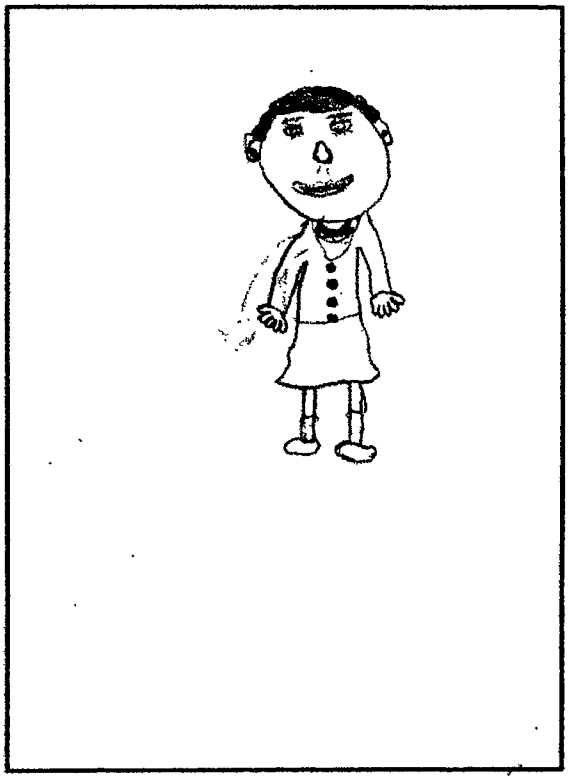

Figure 35. N2 (GH11), age 8

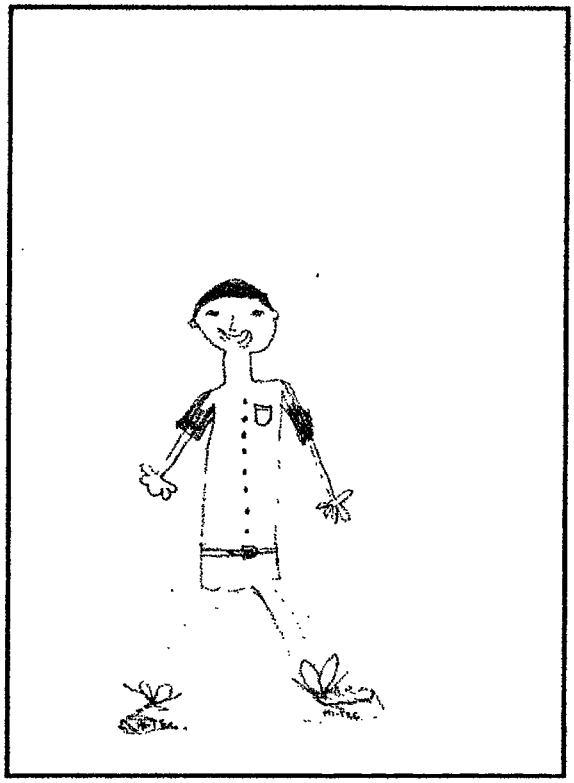

Figure 36. N3 (GH19), age 9

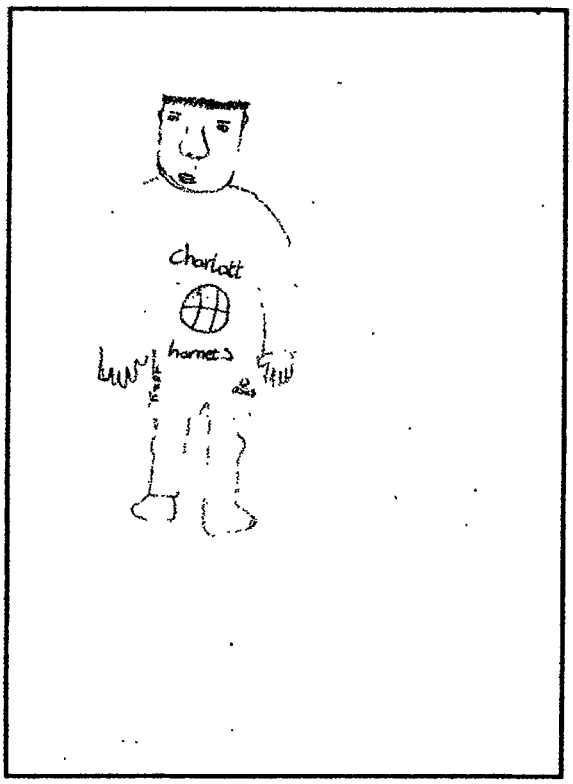

Figure 37. N4 (GH22), age 10

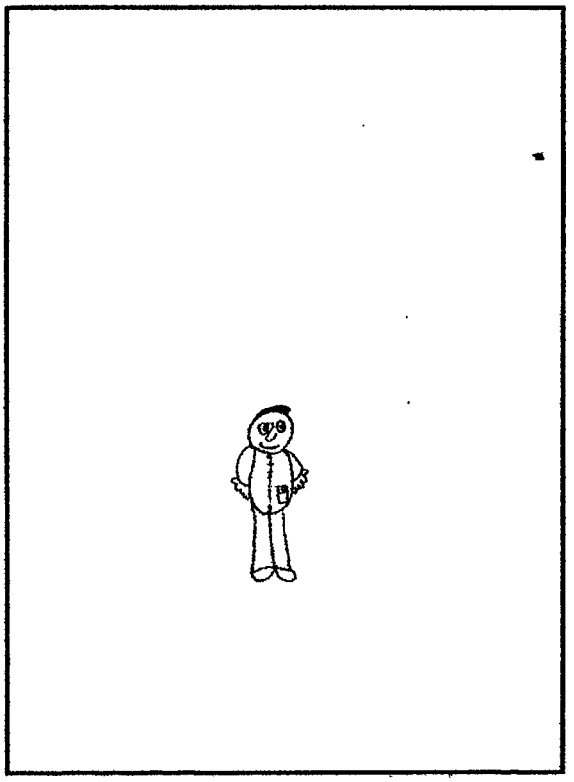

Figure 38. N5 (C24), age 10;3

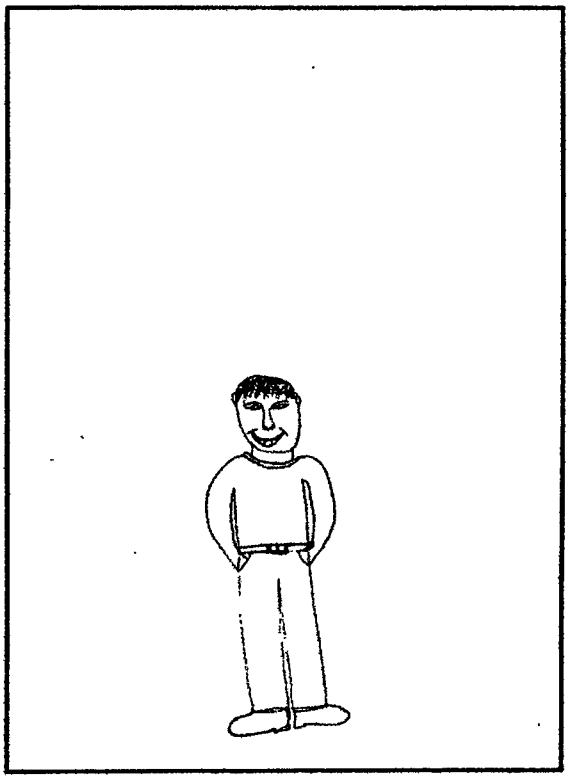

Figure 39. N6 (C27), age 10;4

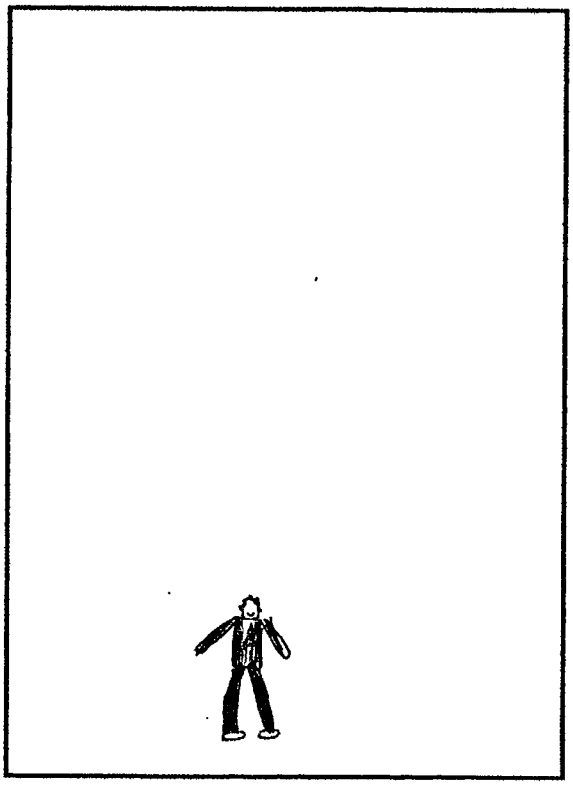

Figure 40. N7 (GH28), age 9

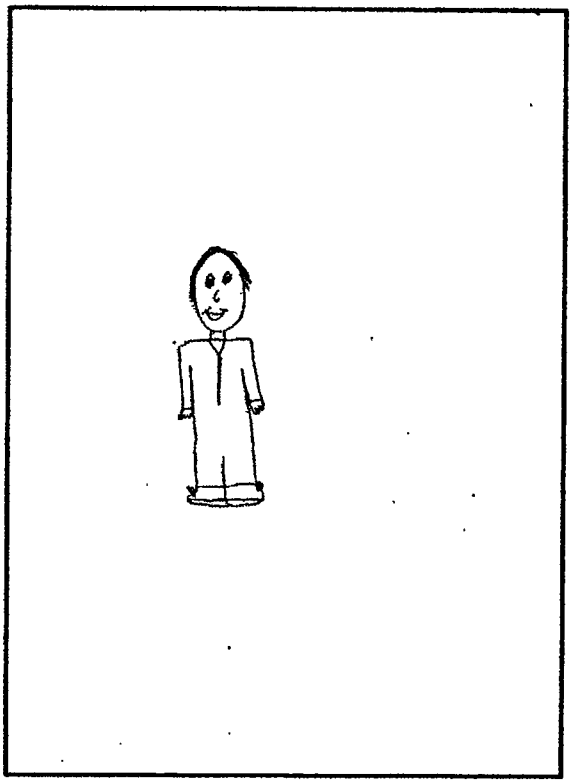

Figure 41. N8 (GH29), age 11

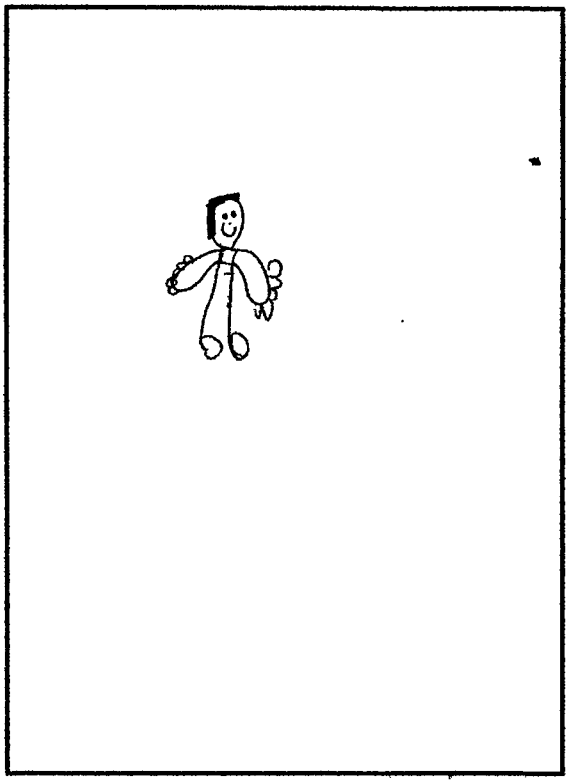

Figure 42. N9 (GH31), age 7

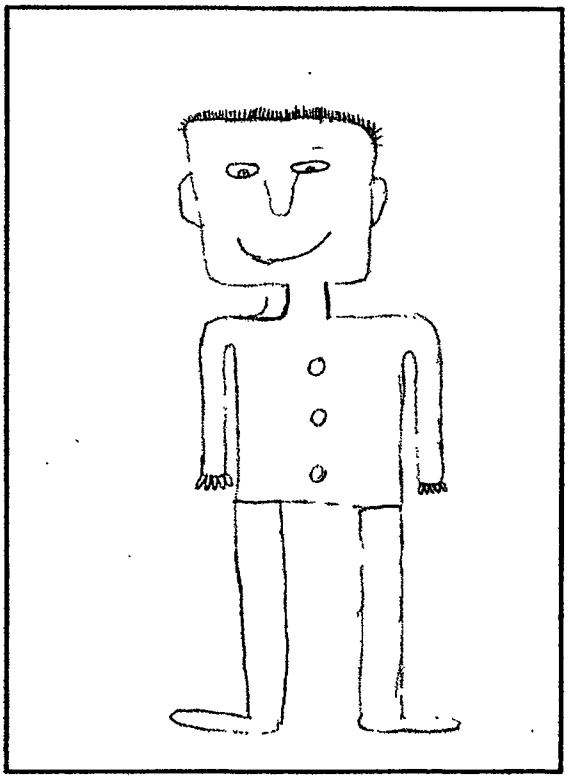

Figure 43. N10 (GH32), age 8

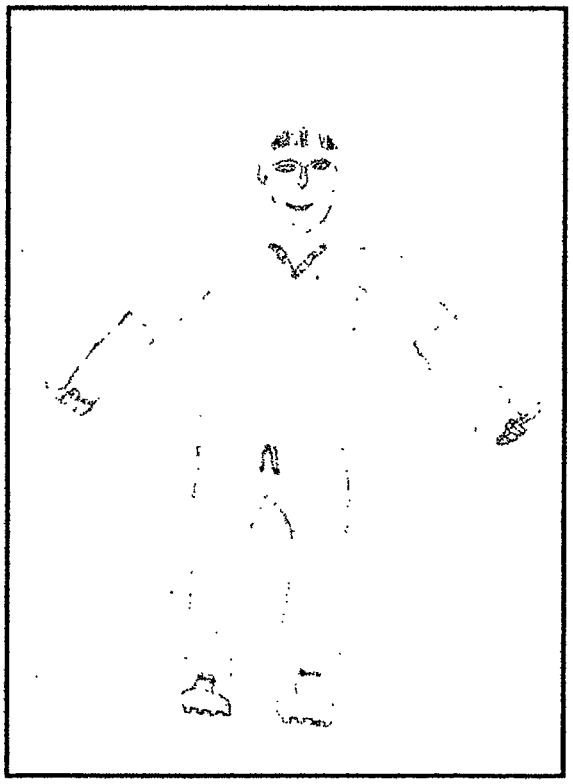

Figure 44. N11 (GH37), age 10

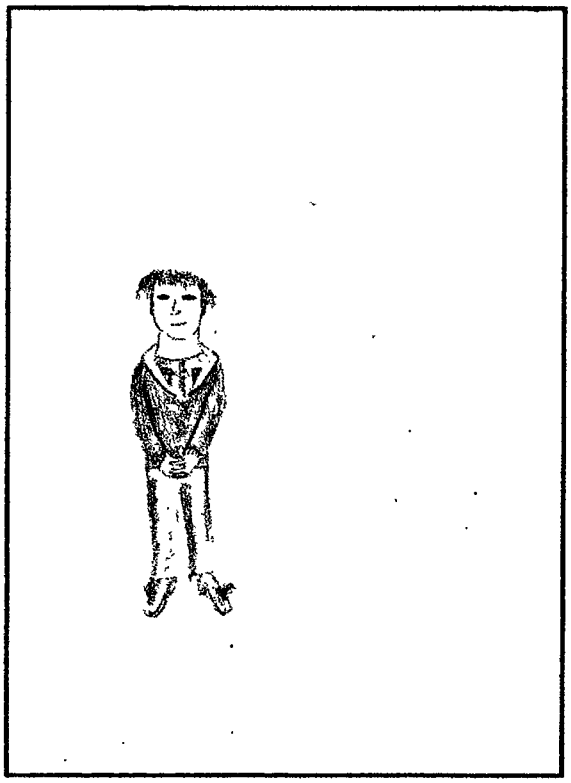

Figure 45. N12 (GH39), age 10

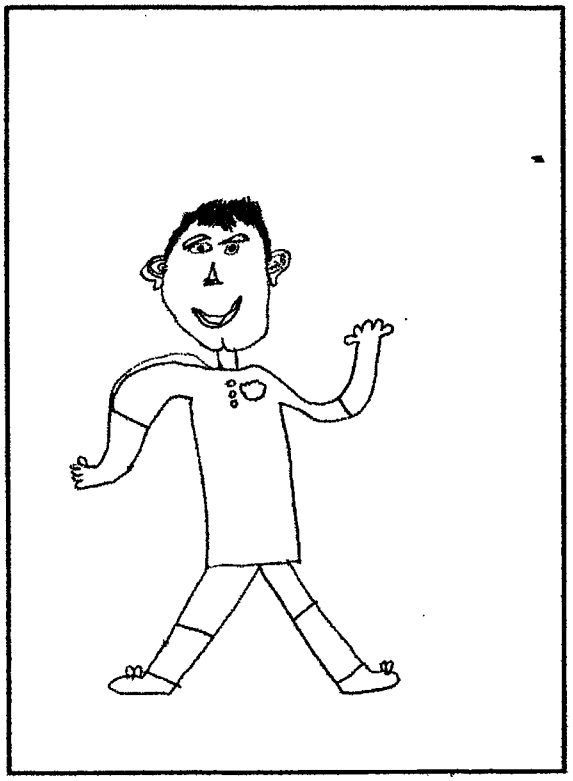

Figure 46. N13 (GH42), age 8

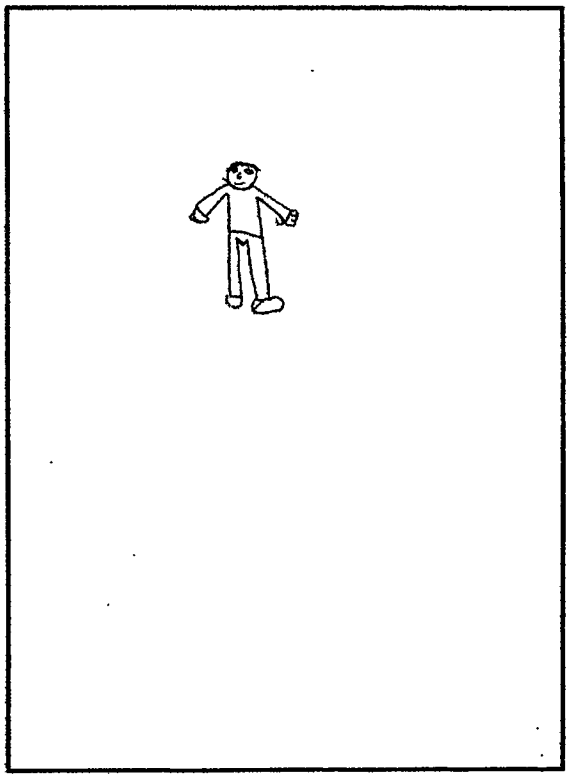

Figure 47. N14 (C44), age 11;8

In the discussion with the experimenter (MC) the judges confirmed that the disturbed set of drawings suffered from a lack of detail in general, as well as specifically on the face and clothing. The figures also had poor proportions. Strangely shaped and poorly defined limbs characterised the disturbed set. These drawings showed a lack of facial features and detail such as pupils in the eyes. They showed fewer smiles and facial expressions in general and often did not have clothing or hair depicted. The figures were described as basic and immature.

The normal HFDs were described as cheerful, happy figures, depicted as a specific person, often footballers and often doing something. These HFDs were more detailed with brand logos on clothing and more detail on faces such as pupils in the eyes and had hair present. Smiling faces were common on the normal HFDs and they also depicted clothing more often. These HFDs were seen as age appropriate, typical and looking more like people.

Similarities between the drawings were that the core features of the human figure (i.e. head, body, arms and legs) were included in both sets. The children had all mainly drawn male figures.

Some of the drawings were identified as not belonging to the sets as well as the others. These are listed with any reasons which were given.

<u>Disturbed set</u>

D4 no reasons

D7 'could be normal but the figure proportions are wrong'

D8 'the dagger could be media influenced so it could be normal'

D9 'friendly looking', 'cartoon like'

D10 'puppet-like', 'arty', 'different', 'might be normal, like he's copying a wooden model'

D12 'not as bizarre as the others, lacks detail though'

D13 no reasons

<u>Normal set</u>

N5 'basic for the child's age'

N7 'doesn't belong due to its size, but fits due to the smile'

N9 'not as well defined as the others', 'similar to older D group ones', 'a bit basic', 'has facial detail but out of proportion'
N14 'basic for 11', 'different to the others', 'lack of detail', 'looks anxious, not happy'

It is also possible to look at the differences between the drawings on the scoring methods already used in this thesis (indicator scores and GH scores), both to check that the drawings were different on these measures and look for confirming evidence for the judges' statements. Summary statistics for these scores are shown below. (For full data, see appendix 7). The drawings were scored for the indicators without the confounding items from the GH scale.

	Disturbed HFDs			Normal HFDs		
	GH	EI	Revised	GH	EI	Revised
Mean	85	2.4	1.1	104	1.5	0.4
St.Dev	15.8	1.4	0.9	18.2	0.9	0.5

Table 39. Summary statistics for the disturbed and normal drawings.

This table shows that the GH scores are considerably different. A t-test showed the GH scores are significantly different ($t(26)=-2.97$, $p<0.01$). Due to a skewed distribution of scores a Mann-Whitney test was used to compare the EI scores. The results showed no significant difference between the disturbed and normal drawings' EI scores ($z=-1.9$, ns) or revised indicator scores ($z=-1.9$, ns).

The data can also be analysed as Koppitz did, comparing the number of drawings showing 0 or 1 indicators with the number showing 2 or more indicators. These data are shown in the following tables.

	Disturbed	Normal
0-1	4	6
2+	10	8

Table 40. Number of HFDs in each group showing 0 or 1 and 2+ Koppitz indicators.

A Chi-Square analysis showed no significant difference between the number of HFDs with 0 or 1 and the number with 2 or more indicators (X^2=0.156, df=1, ns).

	Disturbed	Normal
0-1	8	14
2+	6	0

Table 41. Number of HFDs in each group showing 0 or 1 and 2+ revised indicators.

A Fisher Exact test for the revised indicator scores showed significantly more normal drawings with 0 or 1 indicators ($p<0.01$).

DISCUSSION

The results of this study show that the judges did see a difference between the disturbed and normal drawings. First, the drawings were seen to differ in detail. The normals had more detail in the face, e.g. eye detail and also on the clothing such as shoelaces and brand logos. The disturbed drawings were seen to lack detail. The level of detail overall in the drawing can be related to the amount of attention paid to the drawing. Some detail which the judges commented on as missing from the disturbed drawings is not necessarily expected on the drawings for that age (e.g. ears) which may show a naiveté on the part of the judges about what to expect from children's drawings. These are details that are added to the HFD later in development and are not necessarily problematic for the disturbed drawings, especially if they are considered as delayed rather than deviant in nature. Indeed, this can be seen as reinforcing the idea that the indicators of disturbance are more cognitive than emotional.

The normal HFDs were seen to depict clothing more often than the disturbed ones. This may also be a sign of developmental delay since the addition of clothing on a drawing is part of the developmental process. It may also be the result of an attention problem for the boys who simply did not spend enough time adding detail to their drawings. This may be the result of the clinical children's disturbance, but could

be evident in normal children too since some of their drawings were included in the disturbed group.

Smiling is a common theme noted in the normal HFDs, which was seen less often in the disturbed drawings. Facial expression has been noted by other researchers (e.g. Tharinger & Stark, 1990) as a guide to discriminating clinical HFDs, but the major problem is that a very large proportion of all children's drawings, even clinical ones, have a smiling face. Golomb (1992) noted this in drawings of scary dreams, where the child still drew a smiling face on the scary character. This may be more to do with the schema or symbol of the face which is automatically drawn with a smile. This schema is learnt and it may be that some children have not learnt it, or do not use it appropriately. However, when all the clinical drawings are compared with the control HFDs, they show similar numbers of smiles. Facial expression is therefore a very poor predictor of disturbance for the drawings collected in this thesis because it does not discriminate well between the samples, even where differences are usually found such as between the clinical and CA and MA matched samples.

Some added detail (e.g. knife, breasts) was seen in the disturbed set which may be seen as indicative of disturbance, but it was highly specific to individual drawings and sometimes misleading since the drawing (e.g. knife added to HFD) was actually drawn by a normal child. Therefore the clinical validity of these items is severely limited.

The complete lack of faces entirely on two drawings was always seen as indicative of disturbance. This cannot be due to a developmental delay for these boys, since even a 3 year old would be expected to put eyes in a drawing and the face is usually included before the body or arms. It is possible that the children attended to other parts of the drawing (e.g. boots) and did not go back over the drawing to check they had included everything, but the face is normally completed during the initial part of the drawing sequence, so this could be a sign of a deviancy in the development of these boy's drawings. Equally, it could relate to an attention problem if they simply did not spend enough time attending to the task and completing the drawing properly and comprehensively. It is possible, but perhaps less likely, that they did not see the face as necessary for the representation of the

person. This relates to the issue of the purpose of the drawing, for if it is recognisable as a person without the face, then the child may see it as unnecessary to draw one in. The problem with this explanation is that the judges commented that these faceless drawings do not look like people but look more like a turtle and a bird, not because of the lack of a face, but because of the shape of the limbs. If the face were added, it would not necessarily alter the judges' perception since when this was done in study 10 (this chapter), the rating of bizarreness did not change. The mis-shapen limbs were overpowering and the presence or absence of the face did not make any difference to its perceived 'humanness'. It remains odd that the children did not include a face, and this may be indicative of their EBD, but the practical usefulness of such an indicator is questionable due to the fact that it occurs so rarely in a clinical population and may not have been used as a general strategy which will generalise to other situations and other children.

The size of the figures was also mentioned by the judges, with a tendency to see the disturbed drawings as smaller than the normal ones. However, there is no difference overall between the heights of the clinical and control or disturbed and normal drawings, and the reliability of size in HFDs is questioned by many researchers (e.g. Jolley, 1995).

The judges in this study mentioned items which are similar to the Koppitz emotional indicators such as teeth present, lack of hands, shaded face/body, lack of face and figure size, though several of these are not revised indicators. Some of these are also already involved in the GH scale and were removed due to confounding in study 6, chapter 5. The use of these items alone as indicators of disturbance would not be very sensitive to the clinical children's drawings, and suffers from the problem that the fewer items which are used, then the poorer the discriminatory performance becomes. There was no significant difference between the Koppitz and revised EI scores of the two groups when the confounded GH scale items are discounted. Using the cut-off score of two indicators, no difference between the disturbed and normal drawings was found for the Koppitz indicators, though significantly more normal HFDs showed 0 or 1 revised indicators. The judges did not mention many features associated with the revised indicator list,

however, and they did not seem to use those items when assigning the drawings clinical status or rating them as bizarre, limiting their involvement.

Judges also noted that the shape and proportions of the normal HFDs were better than the disturbed ones. The normal pictures are described as having 'okay' or 'good' proportions whereas proportions are either not mentioned at all for the disturbed drawings or they are out of proportion. The disturbed drawings suffer with deformities in shape and lack of definition in the limbs for example. 'Odd limbs', 'poor limb definition', 'tiny head', 'strange body shape', 'long neck', 'large head', 'high ear position', 'segmentation', 'square shape', 'in-turning feet', 'stick arms' and 'long body' have all been used to describe a disturbed drawing. Judges characterise the disturbed drawings as poorly defined and immature which creates an odd impression of the drawings whereas the normal drawings have good proportions and normally shaped limbs and bodies, creating an impression of normality.

The normal drawings tended to depict specific figures, as if the children had an image in mind and wanted to fit the detail to the character. These drawings were of a personality or character rather than just a person formed from the amalgamation of a head, body, arms and legs, which describes the disturbed drawings. The football theme in the normal boy's drawings relates to the influence of local culture on drawings. The disturbed drawings did not contain such images that may have been due to the children not attending to such influences or not utilising them in their drawings. Where the disturbed drawings did contain detail it was seen as odd such as the mark under the eye, the inclusion of breasts or a knife or the segmentation of the body. The detail was misplaced and odd, espccially in the context of what was omitted on the figure, such as the face.

Of the drawings which the judges thought did not belong so obviously to the sets, it is interesting to note that some of them were the clinical ones that had been mistaken for normal and the normal ones which had been mistaken for clinical. The reasons the judges gave for why the drawings did not seem to fit show what criteria they were using to assess the ones that did belong to the set. Proportions and detail were mentioned, with three out of the four normal ones that did not fit

being due to their being too basic for their age and lacking in detail, suggesting they were suffering a developmental delay. Those disturbed ones that did not belong to that set contained a feature which could be seen as normal. The puppet-like drawing (D10) could be seen as arty, and the dagger (D8) may be media influenced and the result of the child attempting to draw a character (soldier?), a tendency seen in normal drawings which usually resulted in a footballer. These two drawings were still seen as belonging to the disturbed set, however, due to their general lack of detail and poor proportions.

Much of what is discussed above is involved in the GH scale and therefore the differences which the judges picked up on is likely to be differences in the GH scores of the drawings. Detail, proportion and body shape are highly implicated in the GH scale, especially in terms of clothing which can earn points and good proportions which can earn points too. Details which the judges specifically mentioned such as pupils in the eyes also increases the GH score of a drawing. The use of 2-D for limbs is also something that the GH scale awards points for and the use of stick arms in the disturbed set is something the judges mentioned. This means that the normal HFDs achieved more success in including features awarded points by the GH scale, and the disturbed drawings suffered on these items. This is confirmed by looking at the results of the GH score t-test analysis which showed a significant difference between the GH scores of the disturbed and normal drawings. This result also corresponds with the previous results showing no differences between the drawing of the clinical and control HFDs when the GH scores are controlled.

Tharinger and Stark's (1990) variables can be related to the results presented in this study. The 'inhumanness of the drawing' can be seen in the descriptions given to the two drawings without faces which were described as not looking like people but more like a bird and a turtle. The normal drawings, by contrast, are seen as drawings of specific characters or personalities. 'Lack of agency' is seen in the disturbed drawings whereas the normal pictures contain images of footballers and people in action. Facial expression was identified by Tharinger and Stark, and is mentioned by the judges in this study, but it is more a case of complete lack of expression at all on the disturbed

HFDs and smiles occurring more commonly on the normal pictures than a 'facial expression of negative emotion' being evident. 'The presence of a hollow, vacant stilted sense in the individual portrayed' is difficult to relate to what the judges said in this study, but may be associated with the disturbed HFDs' lack of facial expressions and failure to show specific characters showing action and emotion.

Hiler and Nesvig (1965) identified bizarreness as the most important variable used by their judges, but also listed 'incompleteness' which may be seen in this study from the lack of detail on the disturbed drawings to the complete lack of faces on two drawings. 'Distortions' were also identified on the disturbed drawings here, mainly in terms of poor proportions and deformed body and limb shapes. Hiler and Nesvig also considered transparencies to be useful in discriminating the drawing of clinical and normal drawings, but the relevance of this item to the disturbed and normal drawings here is questionable. Only one judge mentioned one transparency and very few are actually present in the drawings. Interestingly, Hiler and Nesvig found that clothes and proportions were not a useful discriminatory cue in their study, contrasting with the results found here.

The data seem to best fit a factor described by Adler (1970). Though this was discovered using drawings from adult psychiatric patients, it is relevant to the drawings in this study. The "formal accuracy of the drawn figure and degree to which the figure is differentiated with regard to detail and to individuality" (p. 55) encompasses the comments made by the judges in terms of details and shape and proportions which are also measured by the GH scale. This factor is able to discriminate between the disturbed and normal drawings more successfully than the other researchers' variables and it is important to note that it is a cognitive rather than emotional factor.

The results of this study have implications for the validity not only of the emotional indicators but the DAP test too. The features that the judges identified as differentiating the two groups of pictures are cognitive related features, rather than the emotional disturbance related items listed by Koppitz. The missing details, poor proportion and deformed shapes are features which may be the result of behavioural characteristics such as rushing the drawing process, failing to attend to

the task properly and failing to rigorously check the finished picture. All these behavioural characteristics go into the process of the drawing and it may be misleading if only the final product is looked at. The behaviour that produced the drawing may be more indicative of the disturbance than the drawing itself, which is too vague, confounded by other variables and liable to be misunderstood.

Summary of Chapter 6

The clinical HFDs were rated as more bizarre than the CA and MA matched samples, but not more bizarre than the GH matched control sample. This pattern of results is the same as was found with the emotional indicators, which may mean that the bizarreness of the drawings is related to the number of indicators it contains; however, no clear correlations were obtained between these two variables.

The same pattern of results was also found as with the intuitive method, which suggests that the judges may have been using something similar to bizarreness to make their decisions. Drawings that had previously been identified as clinical had higher ratings of bizarreness than those which had previously been identified as normal. This reinforces the view that the judges may have been using a more global impression such as bizarreness in making their decisions, rather than features such as the emotional indicators.

The bizarreness ratings altered when the boy's ages were available and the ratings were negatively correlated with the GH scores which seems to suggest that the level of bizarreness is related to the child's developmental level. The bizarreness level of a drawing may be an indication of the deviancy of a drawing but it may equally be an indication of delay, since an immature picture could be classed as bizarre for the child's age.

The ratings of bizarreness did not change when some or all of the indicators were removed from a selection of the drawings that were considered clinical and rated higher than average for bizarreness. This implies that the emotional indicators have no effect on the ratings of bizarreness. This also supports the idea that the judges were not using the emotional indicators to make their decisions. It was important therefore to investigate what the differences were between the drawings

that gave them clinical status and higher levels of bizarreness, if it was not the emotional indicators.

When judges were asked to determine what the differences were between the disturbed and normal drawings, differences were noted in the amount of detail, the shape, definition and proportions of the drawings. These factors are highly involved in the GH scale and again suggest that it is a cognitive delay related to drawing development that causes the differences between the drawings.

Facial differences in terms of detail and expression were also noted as well as odd features specific to particular clinical drawings. The discriminatory ability of the facial expression is questionable, however, since so many of all children's drawings have a smiling face.

The differences which were seen were often related to delay and immaturity in the disturbed drawings, but strange features which would not be expected even if the child were younger were also noted, such as the lack of faces on two drawings, or the addition of a knife. These items may be related to the level of bizarreness and clinical status of the drawing. The clinical usefulness of these items is questionable, however, since they occurred in odd instances which may not be replicated and were highly specific to the drawing itself rather than being used as a general strategy which may occur in different situations by different children. The clinical validity of the DAP test as a measure of emotional disturbance is questioned due to the obvious involvement of cognitive influences and developmental delay in the interpretation of a drawing.

CHAPTER SEVEN

GENERAL DISCUSSION AND CONCLUSIONS

Review of results

The main aim of this thesis has been to evaluate the DAP test and the Koppitz emotional indicators as a measure of emotional disturbance by comparing the HFDs of children with EBD with normally adjusted children. The influence of chronological age, mental age and Goodenough-Harris scores was investigated as well as the intuitive method of identification and global ratings of bizarreness.

The first aim of the thesis was to replicate Koppitz (1968) who found significant differences between the numbers of indicators on the drawings of disturbed and normal children, using this as evidence for the validity of the indicators. Mental age delay in the clinical sample could explain the higher number of emotional indicators found in their drawings. Therefore, a mental age control was also introduced to examine whether the clinical sample children showed a mental age delay and responded in their drawings like younger, normally developing children. Study 1 showed that the inter-rater reliability of the indicator scoring system is good, but no differences existed between the indicator scores of children with EBD compared with CA and MA matched control samples. This result failed to replicate Koppitz and queries the validity of the indicators. The question of whether the mental ages of the children could explain the differences Koppitz found could not be answered with these results. Suspicion over the severity of the sample used in study 1, compared with Koppitz's clinical sample, and the validity of the normative data upon which the indicators are based suggested further investigations were necessary, before concluding that the indicators are not a valid method for interpreting HFDs.

Chapter 2 also examined the use of a different method for identifying disturbed children's drawings, namely the 'intuitive method of identification'. This involves using visual inspection of the drawings and implicit criteria for discriminating the HFDs of clinical and control samples. Study 2 showed that neither expert nor novice judges could

discriminate the drawings of the clinical sample children from their CA and MA matched controls, using this method. These results may be problematic for the validity of this technique, but equally may be due to the similarity of the drawings, since no differences could be found between the HFDs on other measures, such as the emotional indicators and compositional variables. The performance of the experts did not differ from the novices, implying that their experience was not useful for this task, but since neither set of judges could discriminate the samples above chance, probably for reasons other than personal ability, it is hard to come to any conclusions about the gains available from having expertise in the area. It is concluded from chapter 2, therefore, that the indicators may not be valid for interpreting the HFDs of children who are disturbed but still attending mainstream schools. The drawings of these children cannot be separated from their peers of similar chronological or mental age.

The question over the severity of the clinical sample in study 1 was assessed in chapter 3, by employing a sample of boys who were attending non-mainstream schools for their special educational needs relating to emotional/behavioural difficulties (EBD). Study 3 showed significant differences between the indicator scores of the boys with EBD in special schools compared with both CA and MA matched control samples. This contrasts with the findings of study 1 and suggests that the differences between the results was due to differences in the severity of the clinical sample disturbance. This means that the indicators may be valid for interpreting HFDs, since they are able to discriminate the drawings of disturbed and normal samples. The issue of whether the disturbed children showed higher levels of indicators due to their mental age delay was resolved by these results. The comparison between the clinical and MA control sample results was identical to the clinical and CA control sample results, showing that the clinical sample children were not simply drawing like younger children of a comparable mental age. However, there is a question over the practical utility of the indicators, even given the significant differences which are found, since they occur in such small numbers, even on the clinical sample of study 3. Using the cut-off score of two indicators which Koppitz proposed as significant, a large proportion of the clinical

sample children would be missed. As noted in chapter 1, Lingren (1971) acknowledged that it is possible to find statistically significant differences between groups of children's drawings but these generally have limited practical significance.

The intuitive method of identification was used again in chapter 3, in order to determine whether the judges would succeed at the task when the drawings of the more severely disturbed boys were used. Study 4 showed that both experts and novice judges could discriminate the clinical sample HFDs from both the CA and MA matched control HFDs with accuracy levels which exceeded chance. No difference was found between the experts and novices, reinforcing the results of previous research and showing that experience and familiarity with children's drawings is not necessary in order to be able to discriminate the HFDs of disturbed and normal boys. The success of the judges using this method gives validity to the DAP test but not necessarily to the indicators. The judges may be using these features, since where the drawings differ on this measure, the judges have succeeded in discriminating them, but it is not clear that the judges are employing them to make their decisions. The results of study 4 suggests that there is a visually apparent difference between the drawings of disturbed and normal boys, but it is not clear at this stage what this difference involves.

Chapter 3, therefore, has shown that the indicators may be valid since they occur in higher numbers on clinical sample drawings, and that this is not necessarily due to delay in mental age on the part of the clinical sample, since the MA matched control sample responded in a similar way to the CA matched control sample. The clinical usefulness of the indicators is limited, since they occur in small numbers and do not identify the majority of the clinical sample. The sensitivity of the indicators is also limited by the severity of the clinical sample disturbance, since study 1 did not show differences between the drawings of disturbed and normal children, but study 3 did. Either the indicators are not sensitive enough and therefore lack clinical validity for all but the most severe levels of disturbance, or the drawings of the children in study 1 do not actually differ from those of normally adjusted children. The retest reliability of the indicators was significant

for boys with EBD, but not for normally adjusted boys of the same CA. The mean scores for the two groups were similar over the two test sessions. The lower number of indicators and small range of scores which occur on normally adjusted children's HFDs may have adversely influenced this reliability coefficient. The DAP test and the use of HFDs is not necessarily invalid, since the judges could discriminate the drawings of the disturbed and normal boys above chance levels. It is unclear what the judges are using to make their decisions; they may be using the emotional indicators, since they are only successful on the task where differences have also been found between the drawings on this measure.

The validity of the indicators is assessed using three criteria, two of which relate to the normative data upon which the indicators are based - that the items included must occur rarely in the normal population, and not increase in occurrence with age. Chapter 4 examined whether the normative data which Koppitz collected were still relevant in terms of whether the 30 emotional indicators are still rare and not related to age or maturity. Study 5a showed certain changes in the normative data upon which the Koppitz indicators were based. Some indicators are no longer valid due to their showing a relationship with age or occurring too often on the drawings of normal children. Indicators decreasing in occurrence with age were included once they fell below the criteria 16% occurrence, and the minimum age was seen to have altered for some indicators. A revised list of 23 indicators was compiled.

A question over whether Koppitz had fully accounted for the relationship between the indicators and age is raised, since those decreasing in occurrence might also be related to development. The development of the human figure in children's drawing was described in chapter 1. In general, the HFD develops by the accumulation of detail and differentiation of parts. Poor integration, tiny figure, hands cut off and the omission of nose and neck are all items common on very young children's drawings. These items decrease in occurrence with age as the child becomes better at integrating their figure, including necessary features and drawing at a reasonable size to include those features. The occurrence of these items on older children's drawings might

simply reflect a delayed developmental status rather than a deviant nature. In order to maintain comparability with other research, however, these items were still included in further analysis. The question which has been raised serves simply to cast more doubt on the validity of the Koppitz indicators as a measure of emotional health.

The question of new items which may be considered indicators was not addressed here. The Koppitz indicators were based on Machover's work and her own clinical experience, and it is therefore difficult to determine how new items would be chosen. Many of the Machover items have failed to be validated in research. It may be possible to determine items using the perceptions of the judges who discriminated the drawings using the intuitive method, though any new items would then need to be validated themselves.

Chapter 4 also included an examination of the results from studies 1 and 3, using the revised indicators. Studies 5b and 5c obtained the same pattern of results as studies 1 and 3, finding significant differences present only between the indicator scores of the more severely disturbed children and their CA and MA matched control samples. This means that using the revised indicators, and thereby accounting for changes in the way children draw the human figure, the same differences are seen as using the original Koppitz list. This revision of the normative data, however, limits the practical usefulness of the indicators, because of the reduction in number of valid items. The average number of revised indicators present even on the clinical sample drawings was reduced to a level close to one, a number which Koppitz admitted was not clinically significant.

The collective use of individual features such as the emotional indicators, for clinical populations, is restricted by the incidencc in which those features occur in the normal population. Subtle changes in the form of the HFD over time and between cultures means that the features which are taken as significant can vary. Chapter 4 showed that the normative data upon which the indicators are based had changed enough to cancel the validity of certain items and alter the age at which others may be interpreted as meaningful. This revision of the normative data, and the subsequent alteration of the results shows that the original list of emotional indicators should not be relied on too

heavily, especially for contemporary British children's drawings. The Koppitz norms have changed and the indicators are no longer all valid. Though a revised form of the indicators showed similar statistical differences between the samples as the original list, the mutability of the form of the HFD questions the reliance on individual items, even used collectively. The alterations which were made to the list of valid indicators in chapter 4 also placed constraints on the practical usefulness of the indicators, given the lower number of items which could be considered valid. Care must be taken in using the remaining items, since the lower number of valid indicators has reduced the likelihood of their occurrence in numbers which have a practical usefulness. This result also cautions against wholesale use of other people's normative data which may not be relevant to the population being examined, and reliance on data from research which used the emotional indicators, without considering the effects of the normative data.

The involvement of cognitive maturity, measured by the GH scale, for the interpretation of drawings and the assessment of emotional disturbance was assessed in chapter 5, following a question over the relevancy for drawings of the WISC measure of MA. The use of the WISC items as a measure of drawing ability and its related skills may explain the failure to obtain the expected results using the MA matched control sample. The verbal subtests such as vocabulary and similarities relate more to verbal intelligence than to drawing skills and therefore using these subtests would not give a reasonable estimation of a child's expected performance on the DAP test. The performance subtests are more relevant as they utilise spatial and problem solving skills. Object assembly may be similar to a manikin-type production task used in drawing research; however, this usually reveals a higher level of representation than the children are capable of drawing themselves and the overestimation in a prediction of DAP test performance may result from using this task.

The GH scale measures the intellectual maturity of the child, using the details, differentiation and proportions of their HFD. Harris (1963) considered the scale as a measure of intellectual maturity, related to concept development. Many studies have shown a correlation

between the scale and IQ measures and it is generally accepted that the scale is a cognitive measure. The scaled score measures the quality and formal accuracy of the drawing using the body parts and associated detail, features and proportions, as well as the quality of line, differentiation of parts and shape of the figure. Thus it can be seen as a measure of both recall memory of the body image for the HFD and the artistic and technical ability of the child to produce the image on paper. Motivation and concentration may both improve the scores which can be obtained.

Items on the indicator list are confounded with items on the GH scale. Items associated with head and limb proportion and the presence of body parts occur on the GH scale and the indicator list. With these items removed, study 6 showed no significant differences between the indicator scores of the clinical sample from study 3 and control HFDs matched for GH scores. This suggests that the differences previously found in study 3 between the drawings of the clinical and CA/MA samples may have been due to differences in GH scores and confounding between the indicators and the GH scale. A negative correlation between the indicator scores and the GH scores of the drawings when the confounded items are not removed, supports a view that the indicators are an indirect, cognitive measure of drawing ability rather than a measure of emotional status.

Chapter 5 also evaluated whether the intuitive method of identification would be successful, using the GH matched control HFDs. Study 7 showed that judges could not discriminate the clinical and GH matched control HFDs above chance level. The judges were not able to separate the drawings of disturbed boys from those of normal boys when the indicator scores of the drawings are similar, and the GH scores are matched. This suggests that the previously successful achievement of the judges was either due to the differences in indicator scores or the GH scores of the drawings. Equally, the failure of the judges in study 2 may be explained by either the similarities in indicator scores, or the fact that the GH scores of the three samples were all very close to the normal average of 100, therefore not from poor drawers. The items included on the GH scale are more obvious and less obscure than the indicator features, and therefore are more likely to

have been used, but it is also possible that the judges utilised some other factor more connected to the global impression of the drawing, than the collection of individual features.

The question of whether the expert judges who took part in studies 2 and 4 were lacking in practical experience using artwork in a clinical setting was also examined in chapter 5. Previous research (Ulman & Levy, 1973) had shown that art therapists were no more successful than novice judges, though this study used adult artwork and changes in art therapy training may have occurred since then. Study 8 employed art therapists as judges who attempted to discriminate the drawings of the clinical sample drawings from study 3 from their CA, MA and GH matched control sample drawings, in order to determine whether they could be more successful than the previous judges in identifying the clinical sample HFDs. The results showed that the art therapists did not perform differently from the other expert and novice judges previously used, as they did not identify any more of the drawings correctly than the other experts or novices. The art therapists could identify the clinical HFDs when compared with CA and MA matched control drawings, but not the GH matched control HFDs. Previous experience with artwork in a clinical setting does not seem to affect the success in using the intuitive method of identification, since the art therapists did not achieve higher scores than the other types of judge. This reinforces the work of Ulman and Levy (1973) and extends the findings to children's drawings.

Chapter 5 has shown, therefore, that the GH scores may be responsible for the differences seen between the severely disturbed clinical sample and their CA and MA matched control sample drawings. This supports the possibility that the HFDs acquired using the DAP test are a measure of cognitive maturity rather than emotional disturbance. A problem with the use of the emotional indicators is the influence of cognitive maturity, measured not necessarily by the WISC, but by the GH scale. The differences between the drawings in terms of the emotional indicators in study 3 may have been due to the confounding between them and the GH scale items. The judges may also have been using the GH scores to discriminate the drawings since they are not successful when the scores of the drawings are similar. However, since

the indicator scores are also equal in these cases, it is still possible that the judges are using the indicators to tell the drawings apart. It is also feasible that the judges are using some third factor gained through a global impression of the drawings, which the indicator and the GH scores could be connected to.

The interpretation of HFDs using the appraisal of a collective number of features such as the Koppitz emotional indicators does not receive much support from the results of this thesis. This adds to the basic problem with feature methods in general as the evaluation of Machover's signs in chapter 1 illustrated. The sign interpretation as a method for analysing HFDs shows poor reliability and validity and many of the signs are confounded with production problems. An alternative method to the collective use of features, is the use of intuitive criteria for discriminating the HFDs of clinical and normal samples. Studies 2, 4 and 7 evaluated the 'intuitive method of identification' (Dieffenbach, 1977), which involved using a visual inspection of the drawings and intuitive criteria to discriminate the clinical HFDs from their CA, MA and GH matched control samples. This technique yielded a similar pattern of differences between the drawings of disturbed and normal boys as with the indicators. The judges were successful in discriminating the pictures when the factors of CA and MA were taken into account but no differences were seen between the clinical sample drawings and their GH matched controls.

The subjective nature of the intuitive method is acknowledged, but experience did not improve the accuracy and objectivity of the decisions. The consistency of the judges on many pictures suggests good agreement among them regarding the clinical or normal status of the HFDs, though sometimes the judges were consistently incorrect. The accuracy of the judges suggests a hit rate which may be better than the indicators, in those circumstances where differences between the drawings were seen. On average, the judges could identify approximately 70% of the drawings correctly when the CA and MA of the children were controlled. In some conditions (e.g. 'CA control, pair format, with ages'), all of the judges were able to discriminate the drawings above chance levels. However, the accuracy of individual judges fell to chance levels when GH scores were controlled across the

clinical and normal samples, again implicating the influence of the children's cognitive maturity on their drawings. The problems which the judges had in discriminating the samples when GH scores were controlled implies that the judges were not sensitive to the child's emotional disturbance in the drawing, but their cognitive maturity, questioning the use of this method for the assessment of emotional status.

Chapter 6 evaluated whether the drawings differed on a global impression, measured using ratings of bizarreness. Quantifying the global impression of the drawings into a rating of bizarreness also produced a similar pattern of results as the intuitive method and the indicators. Study 9 showed that the clinical HFDs were rated as more highly bizarre than both the CA and MA matched control HFDs, but equal to the GH matched drawings. The ratings changed when the boy's ages were made available to the judges, such that when the judges knew how old the boys were, they made their rating higher for the clinical drawings and similar for the control samples. The ratings did not seem to relate to the indicator scores or the boy's ages, though the small numbers of children at some ages cautions against the reliability of these data. A negative correlation is obtained, however, between the bizarreness ratings and the GH scores of the drawings. Only 2 out of 44 clinical HFDs (C2 and C22) had age appropriate GH scores but were also rated above the average for bizarreness. Other drawings which were rated above average for bizarreness also had lower than average GH scores, which may explain the rating. The interpretation of the drawings for emotional disturbance, therefore, is again confounded with the effect of cognitive maturity on the drawings. Drawings which had previously been assigned clinical status in studies 4 and 7 were shown to have higher ratings of bizarreness than those assigned normal status. This suggests that the judges were using the global impression of the drawing, rather than the emotional indicators to make their decisions.

The indicators, however, may have contributed to the global impression. This was tested in study 10, by removing the indicators from the drawings to determine what effect it had on the ratings of bizarreness. Results showed that these ratings did not change when

either the original Koppitz indicators, or the revised indicators were removed from the drawing. The indicators are therefore not related to the level of bizarreness in a drawing, suggesting a dissociation between the presence of features such as the emotional indicators and the global impression of how bizarre the figure looks. This casts doubt on the view that the global impression of a drawing is related to the indicators it contains and suggests that the differences between the drawings in terms of numbers of indicators are not being utilised by the judges when deciding which drawings to assign clinical status. It is therefore necessary to determine what does characterise the differences between those HFDs which are seen as clinical and bizarre and those which are seen as normal and not bizarre.

Study 11 aimed to determine those differences which could be seen by the judges and were relevant to the distinction between certain drawings in terms of assigned status (clinical/normal) and ratings of bizarreness. The difference between the drawings was found to be associated with the detail, shape, differentiation and proportion of the figures. Occasional mentions were made of features similar to the emotional indicators, but they are only implicated in very limited fashion due to their highly specific and non-strategically placed nature, giving them virtually no practical, clinical value. The lack of detail in the drawings which were in the disturbed group can be seen as relating to their developmental status. In chapter 1, a description of the development of the human figure showed how the development of the HFD, once the basic form of the canonical representation has been established, undergoes a process of addition of parts and details. Up to about the age of 12, children add more and more parts to their HFD, a fact which led to the development of the GH scale. The 'normal' drawings reflected this, with the judges noting that lots of detail was included in them. The 'disturbed' drawings, on the other hand, were lacking in such detail as pupils in the eyes, hair and clothing, perhaps reflecting a delayed developmental status. The lack of faces entirely on two of the disturbed drawings does not concur with this delayed status, however, since the depiction of eyes normally occurs before the body or arms are produced, such as in a tadpole figure, and both these figures had a body and all four limbs. However, the practical value of this lack

of facial features is limited as with those features similar to the emotional indicators, due to their severely limited occurrence and non-strategic nature.

The qualities which the judges in study 11 isolated are highly related to the GH scale items. They are also implicated in the factor described by Adler (1970) which shows enormous similarity with the items on the GH scale. In a review and factor analysis of the criteria used by clinicians to interpret drawings, Adler discovered that this factor could account for most of the variance in the drawings of adult patients. Described as the "formal accuracy... and degree to which the figure is differentiated with regard to detail and individuality" (p. 55), this factor develops in children with age and cognitive maturity and relates to the "maturity or sophistication of the body image representations" (p. 56). Items associated with pathology in drawings loaded highly on this factor and since Adler stressed that this factor was a cognitive one, indicators of pathology could therefore be understood in terms of cognitive immaturity. The confounding of many Koppitz emotional indicators with items on the GH scale also supports this argument that aspects of a drawing thought to reveal emotional status are more likely to be associated with cognitive maturity. The differences between the drawings in study 11 are best described by cognitive rather than emotional variables, casting doubt on the assumption that the DAP test, and the use of the emotional indicators in particular, is valid for measuring emotional status in the artist.

Chapter 6 has shown that the drawings differ in respect to their global impression, measured by ratings of bizarreness. The clinical HFDs were rated as more bizarre than the CA and MA, but not GH matched controls, reinforcing previous evidence which suggests that the GH scores are responsible for the differences between the disturbed and normal children's drawings. The indicators are unrelated to the level of bizarreness of a drawing and those which are seen as bizarre and clinical differ from those which are seen as normal and not bizarre in terms of the details, proportions, shape and differentiation of the figures. These differences are highly related to the GH scale and the factor described by Adler, and are cognitive developmental variables, not measures of emotional disturbance. The relationship between a

child's emotional functioning and their HFD is less clear when cognitive functioning is taken into consideration.

Returning to the question of new indicators, raised at the end of chapter 4, this was partly assessed in study 11. The differences between the drawings which the judges identified could be formulated into new indicators. However, it is now questionable whether this would be an appropriate method, given the problems which have been found with the collective use of indicators and the issue over whether the HFD reflects the child's emotional health at all, whichever items are used.

The influence of cognitive maturity

The first aim of this thesis was to replicate Koppitz, controlling for mental age, a variable many researchers had failed to acknowledge. It was expected that the difference in indicator scores between the clinical sample HFDs and normal children matched only for CA was due to the disturbed children drawing like younger children of comparable mental ages. However, when the mental ages of the children were matched using WISC psychometric test data, the differences between the drawings in terms of indicator scores remained. The results from both the intuitive method and bizarreness ratings reflected this, with the HFDs of the MA matched children responding in a similar way as the CA matched sample. It was therefore concluded that the clinical sample children were not simply drawing like younger children of similar mental ages.

The delay itself in the disturbed children's mental ages might be useful for explaining the differences seen between their drawings, and those of the CA and MA matched samples. The clinical sample boys had a mean CA of 9;10, and mean MA of 8;2. This discrepancy between their CA and MA rather than the absolute level of their MA could be the cause of the differences in their drawings. The drawings of the CA and MA control samples may have reflected their age appropriate cognitive levels. The clinical children therefore, were not simply drawing like younger children, but like delayed children, that is, children who are not functioning at their chronological ages. This delay is clearly reflected in the GH scores of their drawings which were below the

standard mean of 100, for a large proportion (73%) of the clinical HFDs. The GH matched control children's drawings were similar to the clinical sample HFDs in this respect, showing equivalent developmental delay in terms of their drawing of a human figure. These normally adjusted boy's drawings are also identical in respect of emotional indicators, intuitive method and ratings of bizarreness. By using delay as the determining factor for the differences seen between the drawings, it is possible to understand why the differences appeared in the pattern they did.

It is possible that the delay in cognitive functioning in the clinical sample children may be specific to their drawing ability. Assessing drawing as an independent cognitive ability incorporating symbolic and/or spatial intelligence may account for the similarity between the clinical sample and GH matched controls and dissimilarity between the clinical sample and MA matched controls performance on the DAP test. The GH scale may be a more accurate measure of drawing ability than the WISC items and thus predict the performance of the children on the DAP test with greater accuracy. The clinical sample children may show a specific delay in so-called 'graphic intelligence' (Gardner, 1985) for reasons related to their education, or cultural influences from their peers. Alternatively, Goodnow et al. (1986) found that requesting a 'good' drawing from a child focused their attention on detail and resulted in higher GH scores than spontaneous drawings. The children with EBD have general attention problems and therefore may not respond in this way which might have resulted in their poorer drawing performance, as measured by the GH scale. Freeman (1980) saw drawing as problem solving and working memory has also been implicated (Bensur & Eliot, 1993) suggesting that specific delays in these drawing related abilities might explain the DAP test performance of the clinical sample children.

The GH scale is heavily implicated in the results of this thesis. It can account for the differences between the HFDs of the disturbed and normally adjusted children in terms of emotional indicators scores, assigned clinical status and ratings of bizarreness. The emotional indicators are confounded with the GH scale items and the criteria which the judges were using to distinguish between the disturbed and

normal drawings in the final study are also highly implicated in the GH scale. When the GH scores of the drawings were controlled, the judges consistently and correctly agreed on 18 out of the 88 HFDs, when they were given the ages of the boys. Eight of these were clinical sample drawings, 10 were from normally adjusted boys. These drawings presumably contain more of those qualities which the judges used to discriminate the drawings in the other conditions. The differences between some of these pictures was assessed in study 11. It was found that even in these drawings the differences relate to developmental status variables such as the level of detailing and sense of proportion, rather than items such as emotional indicators.

The clinical use of drawings

The results of this thesis have implications for the clinical use of drawings. Though the DAP test may be popular, the results of this thesis suggest the involvement of cognitive measures affecting the interpretation of drawings which cautions against their diagnostic use for the assessment of personality and emotional health.

The Koppitz indicators have been found to occur more often on disturbed boy's drawings, when compared with normally adjusted boys of similar chronological and mental ages. This gives them a possible use as a screening measure, though their generally low rates of occurrence and poor sensitivity means that their usefulness is limited, something other researchers have noted (e.g. Lingren, 1971). Using a global impression of a drawing in order to interpret whether it is from a normal or disturbed child appears to have more success than the indicators in terms of the numbers of drawings that could correctly be identified. Previous researchers (Arkell, 1976; Dieffenbach, 1977; Hiler & Nesvig, 1965; Stricker, 1967; Ulman & Levy, 1973) have also had similar success using this method. As with the indicators, the intuitive method may have uses as a screening measure, given that simple visual inspection can isolate the disturbed children's drawings above chance levels when compared with drawings from children of either similar CA or MA. The improvement in the 'pair format' and 'with age' conditions suggests that the use of an age standard as a guide might increase the success of this technique and allow individual children to be assessed.

The clinical sample drawings were also found to be more bizarre which, again, could be useful for interpreting the HFD, since it quantifies and objectifies the global impression.

The DAP test, however, is susceptible to the influence of cognitive maturity, whichever method is employed for analysis and interpretation. The effect of controlling for GH scores has cast doubt on the validity of the DAP test as a measure of emotional health. It appears from the results presented here, that the indicators may not measure emotional disturbance at all, but may be a broader measure of cognitive functioning, assessed through the detail and accuracy of the HFD. If this is so, then the indicators have no clinical validity for the interpretation of emotional health. Feature methods in general are impractical due to the rarity of their occurrence and the relationship between the items and developmental level, as well as production problems associated with each feature. The actual number of indicators which appear in any one HFD is very low which limits the practical usefulness of them, something which has been noted before (Eno et al., 1981). Though their rarity is what supposedly gives the indicators their clinical significance, it also means that they do not occur consistently enough for them to be used with any confidence. In normal and clinical populations the indicators show a floor effect with large numbers of children only scoring one indicator at most, limiting their sensitivity and specificity

The discriminative ability of the judges, as with the indicators, is affected by the developmental status of the child's drawing and therefore the issue of what the judges are sensitive to is questioned. The rating of bizarreness method also suffered from the problem created by the poor performance of the clinical sample children on the drawing task. Each method employed is related to the GH scale scores so may simply assess the form and content of the drawing as it relates to mental functioning rather than emotional status.

The overlapping area between the distribution of the clinical and control samples drawings in terms of emotional indicators and bizarreness ratings as well as the disagreement among the judges using the intuitive method is quite large. Separating the drawings using emotional function measures leads to many wrong diagnoses. Only

59% of the clinical sample would be correctly diagnosed using the Koppitz cut-off score of two indicators (34% using revised indicators). The judges seemed to have hit a ceiling in terms of the number they could correctly identify, and the bizarreness ratings also obtained a large grey area between the normal and clinical drawings as the standard deviation statistic reveals. The GH scale as a cognitive measure achieved a much more successful discrimination (73% of clinical HFDs were below 100), but as a measure of cognitive rather than emotional functioning.

Factors were isolated in the last study which were not necessarily implicated in the GH scale, such as the presence of breasts, a long neck, the complete absence of facial details, the presence of a knife in the hand of the figure, or a specific mark on the face, but the clinical usefulness of these is limited. These items were rare in the main body of drawings or occurred on specific drawings only and therefore were not being used as a general strategy by the children to display a particular attitude. The variables might not occur again in the same situation or by the same child even under the same conditions, questioning their clinical validity. Symbols and schemas in drawings can be used for the communication of attitudes; the main problem is with their interpretation. The projective use of HFDs is that of an analogic form of communication through symbols. As discussed in chapter 1, this suffers from a lack of negation (Mortensen, 1991) and it would therefore be very difficult to establish whether those items did or did not have a particular symbolic significance. Study 5 also showed that the schemas and symbols in the HFD can change over time and between societies. Copying is a possible route for the transmission of the symbols and schemas. The clinical sample drawings may be poor because they copied the older children's whose drawings were also poor. Though this could explain the maintenance of the difference, it does not explain its origin. The interpretation of symbols in the use of drawings for clinical purposes is problematic and adds to the weight of evidence against the use of the DAP test.

The issue of whether expert or novice judges should be used to interpret the DAP test protocols is important for the clinical use of the DAP test. The results presented here, and supported elsewhere (Arkell,

1976; Hiler & Nesvig, 1965; Stricker, 1967; Ulman & Levy, 1973), suggest that experience is not necessary to be able to discriminate a disturbed child's drawing from normal children's. Undergraduate students were found to be as competent as academic experts and art therapists. Ulman and Levy (1973) saw the ability to make a simple discrimination as the foundation of interpreting drawings in art therapy, since it involves the ability to interpret graphic messages. If this is the case, then it would have been expected that the expert judges, especially art therapists, would have shown higher rates of accuracy, which they did not. The success of the novices on the task suggests that the differences between the drawings were relatively obvious and visually apparent, rather than implying some intuitive ability or skill being present in the students.

Differences between the drawings supports the clinical validity of the DAP test, but since the judges were unable to discriminate the drawings when the GH scores were controlled, these differences are limited. They appear to be more related to the cognitive maturity of the children as it is measured by the GH scale, rather than their emotional status as reflected by features such as the emotional indicators. Previous research showing that clinicians, teachers or other judges were able to identify the drawings of clinical samples may also have failed to consider the effect of the GH scores of the drawings. It is therefore difficult to evaluate those studies which fail to account for, or control, the quality of the drawings and the drawing ability of the artists and recommends caution against relying too heavily on their conclusions that the DAP test is valid.

These results leave the DAP test in a precarious position. The HFD may be useful as a general cognitive measure, through the child's drawing development, but it is probably not useful as a measure of emotional disturbance. Occasionally, features of a child's drawing may seem to given an indication of a troubled emotional health, but this is likely to occur rarely, and behavioural observations might show a problem much more clearly than the child's HFD. The differences on the task for the children with EBD which results in the bizarre and clinical status images seems to be due to reasons local to the drawing task itself. The children's attempts to produce an accurate

representation of the human figure are limited by their cognitive abilities such as problem solving and working memory as well as technical drawing skill involving perceptuo-motor skills. Though their emotional health may be associated with poorer cognitive abilities, a child does not necessarily need to be disturbed to produce a picture with two or more indicators present, which is seen as bizarre and given clinical status. Equivalent drawings were found in this thesis which were from normally adjusted children.

It is always necessary to acknowledge the cognitive influences on the quality of the HFD and its effect on the judgement and interpretation that is made. Researchers (e.g. Engle & Suppes, 1990; Hammer, 1969; Pfeffer & Richman, 1991; Thomas & Jolley 1998, Wanderer, 1969) usually consider that the DAP test must not be used in isolation and interpretations should always be confirmed by other evidence. This allows for the possibility that the emotional status of the child can be seen in their drawings, but may simply be masked by other factors. The results presented here suggest that cognitive maturity may be one of those masking factors, if not the most important one, though it is difficult to determine whether emotional disturbance can be isolated and reliably assessed, independently of this mask. Without knowledge of the cognitive functioning of the child it would be foolish to assess his/her drawing and indeed it might be safe only to assess cognitive abilities via the drawing and leave emotional status to other measures. A delay in drawing development and a bizarre looking drawing may be used as a cue to a problem, but it is unlikely that the drawing can be used as anything more such as confirmation of a diagnosis.

Theories of drawing

According to researchers such as Luquet (1927) and Piaget and Inhelder (1956), the drawing which a child produces is a reflection of their internal representation of the object. The HFD, therefore, can be seen as reflecting the body image and self concept of the artist as they look inwards in order to 'capture the essence of a person' (Koppitz, 1968). Children with emotional problems will reflect this disturbance in their self-concept onto their drawings, in the form and content of the

figure that is produced on the paper, where it can be identified using methods such as the Koppitz indicators. Evidence from this thesis to support this argument comes from the differences seen between the clinical and CA and MA control HFDs. The significant differences between these drawings in terms of emotional indicator scores, intuitive method and bizarreness ratings seems to show that the children with EBD are reflecting their disturbance into their drawings, whereas the normally adjusted samples reflect their healthy emotional state. This disturbance in the clinical sample's self concept is also independent of their mental age, since they were found not to be simply drawing like children of comparable mental ages.

The retest reliability of the indicator scores can also be used as evidence that they are a reliable aspect of the child's drawings, representing their emotional status in the HFD each time the child draws. The Koppitz emotional indicators were found to be reliable for disturbed boys, but not for normally adjusted boys. DiLeo (1973) thought that normally adjusted children, unlike disturbed ones, are not emotionally involved in their drawings, which could explain the different result for the two samples. The revised indicators, however, were reliable for both samples, so it may simply be that those Koppitz items which were later found to be invalid were produced in an erratic fashion by the normally adjusted boys. The GH scores of the HFDs from both the clinical and normal samples were extremely reliable over time, suggesting a relationship with the internal representation. However, the confounding with the GH scores questions the assumption that the indicators represent the self concept or body image, and may instead be an indirect result of cognitive immaturity and lack of drawing skill.

The level of bizarreness of the HFDs may also be seen as a reflection of the internal representation. However, the negative relationship between the GH scores and bizarreness ratings suggests they may be an indirect measure of cognitive maturity and drawing skill rather than emotional health. Also, the finding in study 10, that the emotional indicators are not related to the level of bizarreness, is problematic if both are supposed to be reflective of the internal representation. The judges using the intuitive method appeared to be using the level of bizarreness as a guide to clinical status, rather than

the indicators. The role of the indicators in the judges' success using the intuitive method is limited, if not irrelevant, casting doubt on them as a symbol of the child's disturbed internal representation of their self.

The problem which the results of this thesis present for Luquet and Piaget and Inhelder mainly lies in the involvement of the GH matched control sample HFDs. These drawings came from boys who were normally adjusted, but their HFDs were identical to the clinical sample in terms of emotional indicators, intuitive method and bizarreness ratings. The GH matched sample were not reflecting a faulty internal representation of their self concept in their drawings, because they were normally adjusted. This suggests a problem with the task itself and with depicting the human figure on paper, rather than with the child's internal representation of a person. Given evidence presented in chapter 1 (Sitton & Light, 1992; Thomas, 1995), the drawing procedure and the context rather than the internal representation is more likely to constrain the HFD. It is probable, therefore, that the difficulties which the EBD children have with the DAP test relate to the task demands of drawing rather than their internal representation of the self or human figure. The evidence presented in this thesis better fits the constructivist approach to drawing. This focuses more on the production processes involved and considers the errors and distortions as a result of faulty drawing strategies related to cognitive functions rather than as a reflection of an internal representation. Researchers such as Freeman (1980) consider drawing as a problem solving task and the difficult nature of the drawing process for children is what causes the odd effects sometimes seen in their HFDs. This approach emphasises the cognitive aspects of drawing rather than the emotional.

This means that the children with EBD have a problem with the drawing task and its related abilities. These abilities are assessed by the GH scale but not by the WISC, since the MA matched sample performed differently on the DAP test to the GH matched control sample. The problem with the task which resulted in poorer GH scores may lie in the clinical sample children's attention difficulties, since good attention to detail can acquire higher GH scores (Goodnow et al., 1986). The GH matched sample HFDs were selected from an extremely large

number of drawings and this level of delay in GH scores is rarer in the normally adjusted population. The retest reliability coefficients for the GH scores of the clinical sample HFDs suggests that their style of drawing is stable; it is not possible to determine whether this would also be the case for the boys who produced the GH matched drawings, though the normally adjusted boys who were retested did show good retest reliability for their GH scores. The sample of behaviour assessed using the DAP test is highly associated with the child's motivation and concentration, qualities which are more transient in the children with EBD and the clinical sample's poor performance on the DAP test in terms of GH scores may be the result of such factors, rather than their emotional disturbance itself.

By referring to cognitive factors, and influences local to the drawing task, rather than emotional disturbance, the constructivist or process approach to children's drawing is capable of explaining how the GH matched control sample can be normally adjusted, but produce similar drawings to the disturbed children. The low occurrence of such low GH scoring normally adjusted children, however, is an issue which must be acknowledged and the discrepancy between the results using WISC measures of cognitive ability, and using GH scores needs to be resolved. The fact remains, however, that if the problems in the drawings of the clinical children are viewed as resulting from cognitive function specific to the drawing task, such as their poor concentration and lack of drawing skills, then it invalidates the indicators as a measure of emotional disturbance.

Human figure drawing is just one example of children's drawings. It appears that the drawing that a child produces, whether of a human figure or other object, is not simply a direct reflection of the internal model of the object but is altered as a function of the child's ability to construct a two dimensional image of that representation, using materials such as paper and pencil.

Future research

Some issues raised in the first chapter of this thesis have not been resolved by the results presented in this thesis. Regarding the development of drawing in children with EBD, the results suggest a

delay rather than deviancy in their nature, but it is not clear whether this delay manifests itself in other drawings than the HFD, where the delay originates, how it is maintained and how it is related to other cognitive abilities. Regarding the clinical use of the HFD, the results suggest that the body image hypothesis may be invalid, but this has not been directly assessed.

This thesis has therefore raised some interesting issues for future research. The suggestion of a delay in the drawing ability of disturbed children might benefit from investigations of more specific and related abilities to drawing, such as spatial and symbolic intelligence, co-ordination, planning and problem solving. The role of inhibition and executive function in the drawings of children with EBD has not been adequately assessed and may play a role in the production of the HFD in the DAP test situation. Given the involvement of the GH scale scores in the interpretation of HFDs, factors which might affect these scores could be investigated, to determine what effect they might have. The problems which some clinical groups may present in the process of drawing itself, rather than in their finished HFDs, is something which needs to be researched more fully in the future.

It would also be important to determine whether the delay on the part of the clinical children could be remedied. Given the challenging behaviour of these children, their art education and drawing development may suffer due to a lack of practise in drawing skills. Training studies for younger children have shown a lasting advantage (Cox, 1992, unpublished data) and drawings of children have been improved through education (Pfeffer & Olowu, 1986). The disturbed children who were still in mainstream education in study 1 showed no differences in their drawings which may have been due to their adopting the drawing styles of their normally adjusted peers and spending comparable amounts of time practising their drawing skills. The similarities between the clinical and GH matched sample drawings were likely to be due to similar levels of drawing ability in the children. The remediation of the delay in drawing development and skill would seriously undermine the view that the HFD is a symbol of the child's internal representation and a reflection of their emotional state, questioning whether the DAP test has clinical validity.

It would be useful to determine whether the delay seen in the DAP test pictures of the disturbed children is specific to the HFD or occurs in all their drawings. It would be assumed that the problems that the children might have in drawing skill would present for all topics, unless the HFD is an isolated case. Freeman (1975) criticised the use of a single drawing as evidence of development and the limited information which can be obtained from one HFD is something which psychologists have criticised in the past (Golomb, 1992); the use of another type of drawing or other drawing task may give more information on the differences which might occur between normal children's drawings and those of children with emotional disturbance.

Given the problems for the clinical validity of the DAP test discussed above, it is necessary in future research to determine what role drawings might have in clinical settings. The incremental validity of drawings was not evaluated in this thesis and is an issue which has concerned researchers (e.g. Gresham, 1993). The involvement of the child in the interpretative process could utilise the drawing more as a cue, for instance, in discussing sensitive or difficult issues, for example in the evaluation of sexual abuse. The presence of the interpreter and focus more on the drawing process itself would also allow judges/clinicians to be more aware of the influences local to the drawing task which may affect the final form of the picture. It is unclear how the finished product of the drawing relates to the conditions under which it is produced, especially for older children. The request to 'draw a person' in a particular setting may create its own effects in children's drawings, specific to the artificial situation in which the children are placed. It is therefore necessary to determine the purpose which the drawing serves for the child in the assessment of the validity of the DAP test. This could be done by interviewing the children, though they may have problems verbalising what the drawing means to them, or it could be done by comparing the pictures which are obtained under various instructions, to determine whether different intentions can affect the drawings which are produced and to determine whether the results differ from those obtained during the original DAP test.

Appendix 1

Koppitz (1968) scoring manual for 30 emotional indicators

1. Poor integration of parts (boys 7, girls 6): One or more parts not joined to rest of figure, part only connected by a single line or barely touching.
2. Shading of face: deliberate shading of whole face or part of it, including "freckles," "measles," etc.; an even, light shading of face and hands to represent skin colour is *not scored.*
3. Shading of body and/or limbs (boys 8, girls 7)
4. Shading of hands and/or neck (boys and girls 7)
5. Gross asymmetry of limbs: One arm or leg differs markedly in shape from the other arm or leg. This item is *not scored* if arms or legs are similar in shape but just a bit uneven in size.
6. Slanting figures: Vertical axis of figure tilted by 15° or more from the perpendicular.
7. Tiny figure: Figure 2 inches or less in height.
8. Big figure (boys and girls 8): Figure 9 inches or more in height.
9. Transparencies: Transparencies involving major portions of body or limbs.
10. Tiny head: Height of head less than one-tenth of total figure.
11. Crossed eyes: Both eyes turned in or out.
12. Teeth: Any representation of one or more teeth.
13. Short arms: Short stubs for arms, arms not long enough to reach waistline.
14. Long arms: Arms excessively long, arms long enough to reach below knee or where knee should be.
15. Arms clinging to body: No space between body and arms.
16. Big hands: Hands as big or bigger than face of figure.
17. Hands cut off: Arms with neither hands nor fingers; hands hidden behind back of figure or in pocket *not scored.*
18. Legs pressed together: Both legs touch with no space in between, in profile drawings only one leg is shown.
19. Genitals: Realistic or unmistakably symbolic representation of genitals
20. Monster or grotesque figure: Figure representing nonhuman, degraded or ridiculous person: the grotesqueness of figure must be deliberate on part of the child and not the result of his immaturity or lack of drawing skill.
21. Three or more figures spontaneously drawn: Several figures shown who are not interrelated or engaged in meaningful activity: repeated drawing of figures when only "a" figure was requested; drawing of a boy and a girl or the child's family is *not scored.*
22. Clouds: Any representation of clouds, rain, snow or flying birds.
23. No eyes: Complete absence of eyes; closed eyes or vacant circles for eyes are *not scored.*
24. No nose (boys 6, girls 5)
25. No mouth
26. No body
27. No arms (boys 6, girls 5)
28. No legs
29. No feet (boys 9, girls 7)
30. No neck (boys 10, girls 9)

Appendix 2

Emotional indicators on the HFDs of the three samples in study 1

HFD	Clinical	CA control	MA control
1	none	shading body/limb teeth no nose asymmetry	teeth shading face
2	none	short arms	crossed eyes short arms hands cut off
3	teeth no neck	teeth arms clinging	teeth arms clinging
4	tiny figure	none	none
5	shading face shading hand/neck	none	teeth arms clinging no nose no neck
6	teeth	tiny head	long arms hands cut off
7	poor integration big figure asymmetry long arms no body	tiny head teeth short arms	arms clinging legs together
8	none	short arms	asymmetry poor integration
9	none	shading face	shading face
10	hands cut off	none	none
11	big hands no neck	none	legs together
12	hands cut off short arms	none	none
13	short arms	none	none
14	shading face no mouth no arms	shading face shading body/limb	shading face shading body/limb
15	none	slanting figure	tiny head
16	short arms no neck	tiny figure legs together	no neck
17	shading hand/neck teeth short arms	tiny figure legs together	none
18	poor integration	poor integration no arms	poor integration no arms

Table 42. Indicators scored on the children's drawings of the clinical, CA and MA matched sample HFDs.

Appendix 3

Emotional indicators on the HFDs of the three samples in study 3

HFD	Clinical	CA control	MA control
1	tiny figure	none	none
2	asymmetry poor integration shading face	none	asymmetry
3	transparencies teeth no nose	arms clinging legs together no nose	asymmetry shading face
4	poor integration short arms	none	short arms
5	poor integration	none	eyes crossed short arms
6	poor integration teeth short arms	big hands	arms clinging
7	big figure teeth hands cut off	none	none
8	hands cut off	teeth shading body/limb	poor integration
9	slanting figure no nose	teeth	poor integration
10	no nose	legs together	short arms legs together arms clinging transparencies
11	shading body/limb arms clinging	none	big figure
12	poor integration big hands no eyes no nose no mouth	legs together	slanting figure transparencies poor integration
13	arms clinging	tiny figure teeth arms clinging	tiny figure legs together no nose
14	tiny figure short arms hands cut off no feet	poor integration arms clinging short arms tiny figure	shading face slanting figure
15	teeth	transparencies	teeth
16	none	shading face	short arms
17	slanting figure	none	none
18	poor integration tiny head no eyes no nose no mouth	none	big hands
19	shading body/limb	crossed eyes teeth	teeth legs together
20	poor integration tiny figure	poor integration	poor integration
21	short arms arms clinging	big figure short arms shading body/limb teeth	short arms

Table 43

22	shading face shading body/limb big figure arms clinging clouds	teeth	legs together
23	poor integration shading body/limb	short arms	no neck
24	no neck arms clinging legs together short arms	short arms	shading face short arms no neck teeth
25	none	shading body/limb	tiny figure
26	no neck hands cut off short arms legs together	none	teeth poor integration big figure
27	teeth transparencies	teeth	none
28	long arms no neck	no neck	teeth no neck
29	shading body/limb shading hand/neck	none	no neck short arms
30	no neck	shading body/limb teeth	arms clinging transparencies
31	poor integration big figure arms clinging no neck	transparencies teeth	no neck
32	none	arms clinging	none
33	tiny figure no nose long arms	teeth	short arms
34	arms clinging no nose	none	short arms no neck
35	shading body/limb teeth long arms no neck	none	none
36	tiny figure arms clinging legs together short arms teeth	none	no nose
37	none	teeth legs together	shading face shading hand/neck teeth short arms
38	tiny figure short arms asymmetry hands cut off no neck	no neck	teeth shading body/limb no neck
39	teeth	teeth no neck	none
40	none	teeth crossed eyes	no neck
41	none	teeth short arms	big figure teeth no neck

Table 43

42	shading face shading hand/neck	none	shading face arms clinging
43	shading face	short arms	transparencies legs together
44	no neck	arms clinging	none

Table 43. Emotional indicators on the HFDs of the clinical, CA and MA matched samples in study 3.

Appendix 4

Study 5a Normative data - raw percentage data

Males

Age (years)	5	6	7	8	9	10	11
N=	220	109	136	94	113	90	56
Poor integration	18	8	1	4	1	0	0
Shading face	3	9	4	19	21	10	13
Shading body/limb	10	21	18	19	28	21	36
Shading hand/neck	2	6	1	0	3	1	0
Asymmetry	4	7	2	2	1	4	0
Slanting figure	6	2	4	3	1	0	2
Tiny figure	20	20	12	4	4	2	0
Big figure	4	6	4	20	18	13	7
Transparencies	4	5	4	5	2	1	4
Tiny head	1	2	1	1	2	0	0
Crossed eyes	0	3	0	2	0	3	5
Teeth	2	7	12	29	23	31	23
Short arms	9	17	19	19	16	22	23
Long arms	8	1	1	1	2	2	0
Arms clinging	0	2	4	11	5	4	4
Big hands	5	6	1	1	0	1	0
Hands cut off	24	15	10	4	4	3	5
Legs together	3	4	5	4	5	10	14
Genitals	1	0	1	0	0	0	0
Monster	0	0	0	0	0	1	0
3+ figures	1	0	0	1	0	0	0
Clouds etc.	1	0	0	0	0	1	0
No eyes	1	4	1	0	0	0	0
No nose	23	26	18	5	5	0	2
No mouth	5	7	2	1	0	0	0
No body	6	1	0	0	0	0	0
No arms	17	9	2	1	0	0	0
No legs	2	1	0	0	0	0	0
No feet	35	10	10	1	1	3	2
No neck	66	73	54	38	46	33	16

Table 44. Percentage of boy's drawings showing each of the 30 emotional indicators at each age level.

Females

Age (years)	5	6	7	8	9	10	11
N=	223	96	124	110	105	69	53
Poor integration	15	7	3	3	1	1	0
Shading face	2	0	6	8	9	10	15
Shading body/limb	11	20	9	8	21	17	11
Shading hand/neck	1	1	2	0	0	3	0
Asymmetry	6	6	2	2	2	1	0
Slanting figure	7	0	5	3	3	1	0
Tiny figure	20	19	15	1	4	0	0
Big figure	4	4	10	15	12	14	17
Transparencies	4	2	4	3	4	0	2
Tiny head	1	3	1	0	0	0	0
Crossed eyes	1	1	1	3	2	0	2
Teeth	2	0	6	5	8	3	4
Short arms	26	30	30	20	29	22	19
Long arms	4	2	0	0	1	0	0
Arms clinging	1	3	10	10	11	14	11
Big hands	3	3	0	0	0	0	0
Hands cut off	26	20	14	6	5	4	2
Legs together	3	2	6	8	16	6	4
Genitals	0	0	0	0	0	0	0
Monster	0	0	0	0	0	0	0
3+ figures	0	1	0	0	1	0	0
Clouds etc.	1	0	0	0	1	0	0
No eyes	3	0	0	0	0	0	0
No nose	17	9	6	2	1	1	0
No mouth	5	2	1	1	1	0	0
No body	5	0	0	0	0	0	0
No arms	7	1	1	1	0	0	0
No legs	1	3	3	0	1	1	0
No feet	19	16	8	5	6	3	6
No neck	75	66	49	37	28	16	11

Table 45. Percentage of girl's drawings showing each of the 30 emotional indicators at each age level.

Appendix 5

Revised indicators on the HFDs for study 5b

HFD	Clinical	CA control	MA control
1	none	no nose asymmetry	none
2	none	none	crossed eyes
3	none	arms clinging	arms clinging
4	none	none	none
5	shading hand/neck	none	arms clinging no nose
6	none	tiny head	long arms hands cut off
7	poor integration asymmetry long arms no body	tiny head	arms clinging
8	none	none	asymmetry poor integration
9	none	none	none
10	none	none	none
11	big hands	none	none
12	hands cut off	none	none
13	none	none	none
14	no mouth no arms	none	none
15	none	slanting figure	tiny head
16	none	tiny figure	none
17	shading hand/neck	tiny figure	none
18	poor integration	poor integration no arms	poor integration no arms

Table 46. Revised indicators on the clinical, CA and MA matched samples' HFDs in study 1.

Revised indicators on the HFDs for study 5c

HFD	Clinical	CA control	MA control
1	tiny figure	none	no feet
2	asymmetry poor integration	none	asymmetry
3	transparencies	arms clinging	asymmetry
4	poor integration	none	none
5	poor integration no feet	none	eyes crossed
6	poor integration	big hands	arms clinging
7	hands cut off no feet	none	none
8	hands cut off no feet	none	poor integration
9	slanting figure no nose	none	poor integration
10	no nose	none	arms clinging transparencies
11	arms clinging	none	none
12	poor integration big hands no eyes no nose no mouth	none	slanting figure transparencies poor integration

Table 47

13	arms clinging	tiny figure arms clinging	tiny figure no nose
14	tiny figure hands cut off no feet	poor integration arms clinging tiny figure	slanting figure
15	none	transparencies	none
16	none	none	none
17	slanting figure	none	none
18	poor integration tiny head no eyes no nose no mouth	none	big hands
19	none	crossed eyes	none
20	poor integration tiny figure	poor integration	poor integration
21	arms clinging	none	none
22	arms clinging clouds	none	none
23	poor integration	none	none
24	arms clinging	none	none
25	none	none	tiny figure
26	hands cut off	none	poor integration
27	transparencies	none	none
28	long arms	none	none
29	shading hand/neck	none	none
30	none	none	arms clinging transparencies
31	poor integration arms clinging	transparencies	none
32	none	arms clinging	none
33	tiny figure no nose long arms	none	none
34	arms clinging no nose	none	none
35	long arms	none	none
36	tiny figure arms clinging	none	no nose
37	none	none	shading hand/neck
38	tiny figure asymmetry hands cut off	none	none
39	none	none	none
40	none	crossed eyes	none
41	none	none	none
42	shading hand/neck	none	arms clinging
43	none	none	transparencies
44	none	arms clinging	none

Table 47. Revised indicators on the clinical, CA and MA matched samples' HFDs in study 3.

Appendix 6

Koppitz and revised emotional indicators on the HFDs in study 6
NB. HFDs are scored for EIs without the confounding GH scale items.
Koppitz indicators

HFD	Clinical	GH control
1	tiny figure	none
2	asymmetry poor integration shading face	shading body teeth
3	transparencies teeth	poor integration
4	poor integration	teeth
5	poor integration	transparencies crossed eyes
6	poor integration teeth	gross asymmetry legs together
7	big figure teeth	none
8	none	poor integration shading face
9	slanting figure	shading face transparencies shading body/limb
10	none	none
11	shading body/limb arms clinging	shading body/limb teeth
12	poor integration big hands	tiny figure
13	arms clinging	shading hand/neck clouds
14	tiny figure	shading body big figure
15	teeth	gross asymmetry
16	none	shading face
17	slanting figure	none
18	poor integration	tiny figure
19	shading body/limb	shading body/limb
20	poor integration tiny figure	poor integration
21	short arms arms clinging	shading body/limb teeth
22	shading face shading body/limb big figure arms clinging clouds	no neck
23	poor integration shading body/limb	teeth legs together
24	arms clinging legs together	none
25	none	shading body/limb big figure teeth
26	short arms legs together	teeth poor integration
27	teeth transparencies	teeth
28	none	shading body/limb

Table 48

29	shading body/limb shading hand/neck	arms clinging legs together
30	none	shading face
31	poor integration big figure arms clinging	shading face big figure
32	none	big figure
33	tiny figure	none
34	arms clinging	none
35	shading body/limb teeth	teeth
36	tiny figure arms clinging legs together teeth	teeth
37	none	big figure
38	tiny figure asymmetry	gross asymmetry
39	teeth	shading body/limb arms clinging
40	none	teeth
41	none	none
42	shading face shading hand/neck	none
43	shading face	legs together arms clinging
44	none	crossed eyes

Table 48. Koppitz emotional indicator scores of the clinical and GH matched sample drawings.

Revised indicators

HFD	Clinical	GH control
1	tiny figure	none
2	asymmetry poor integration	none
3	transparencies	poor integration
4	poor integration	none
5	poor integration	transparencies crossed eyes
6	poor integration	gross asymmetry
7	none	none
8	none	poor integration
9	slanting figure	transparencies
10	none	none
11	arms clinging	none
12	poor integration big hands	tiny figure
13	arms clinging	shading hand/neck clouds
14	tiny figure	none
15	none	gross asymmetry
16	none	none
17	slanting figure	none
18	poor integration	none
19	shading (body/limb)	none
20	poor integration tiny figure	poor integration
21	arms clinging	none

Table 49

22	arms clinging clouds	none
23	poor integration	none
24	arms clinging	none
25	none	none
26	none	poor integration
27	transparencies	none
28	none	none
29	shading hand/neck	arms clinging
30	none	none
31	poor integration arms clinging	none
32	none	none
33	tiny figure	none
34	arms clinging	none
35	none	none
36	tiny figure arms clinging	none
37	none	none
38	tiny figure asymmetry	gross asymmetry
39	none	arms clinging
40	none	none
41	none	none
42	shading hand/neck	none
43	none	arms clinging
44	none	crossed eyes

Table 49. Revised indicator scores for the clinical and GH matched sample drawings.

Appendix 7

GH scaled scores, Koppitz and revised indicator scores for study 11
NB. HFDs are scored for EIs without the GH scale confounding items.

	Disturbed			Normal		
HFD	GH	Koppitz	Revised	GH	Koppitz	Revised
1	121	3	2	121	1	0
2	77	2	0	110	2	0
3	74	2	2	110	2	0
4	85	3	2	105	1	0
5	72	3	0	86	2	1
6	70	1	1	107	2	1
7	105	5	2	77	3	1
8	82	4	0	88	2	1
9	76	0	0	77	2	1
10	108	1	0	112	1	0
11	89	2	1	124	1	0
12	75	3	2	122	2	1
13	70	1	1	131	0	0
14	84	4	2	85	0	0
Mean	85	2.4	1.1	104	1.5	0.4
St.Dev	15.8	1.4	0.9	18.2	0.9	0.5

Table 50. GH scores, Koppitz and revised emotional indicator scores for the disturbed and normal HFDs used in study 11.

Glossary

DAM	Draw-A-Man test
DAP	Draw-A-Person test
EBD	Emotional/behavioural difficulties
EI	Emotional indicators
GH	Goodenough-Harris (1963) scale
HFD	Human figure drawing
HTP	House-Tree-Person Test (Buck, 1948)
KFD	Kinetic Family Drawing (Burns & Kaufman, 1970)
MMPI	Minnesota Multiphasic Personality Inventory

References

Abell, S. C., Von Briesen, P. D. & Watz, L. S. (1996) Intellectual evaluations of children using human figure drawings. Journal of Clinical Psychology, 52, 67-74.

Abell, S. C., Heiberger, A. M. & Johnson, J. E. (1994) Cognitive evaluations of young adults by means of human figure drawings. Journal of Clinical Psychology, 50, 901-905.

Adler, P. T. (1970) Evaluation of the figure drawing technique: reliability, factorial structure and diagnostic usefulness. Journal of Consulting and Clinical Psychology, 35(1), 52-57.

Aikman, K. G., Belter, R. W. & Finch, A. J. (1992) Human Figure Drawings: Validity in assessing intellectual level and academic achievement. Journal of Clinical Psychology, 48, 114-120.

Allik, J. & Laak, T. (1985) The head is smaller than the body: but how does it join on? In Freeman, N. H. & Cox, M. V. (Eds.) Visual Order: The nature and development of pictorial representation. Cambridge University Press.

Alschuler, R. H. & Hatwick, LB. W. (1947) Paintings and personality. Chicago: Chicago University Press.

Ames, L. B. (1945) Free drawing and completion drawing: A comparative study of pre-school children. Journal of Genetic Psychology, 66, 161-165.

Andersson, S. (1995) Social scaling in children's family drawings: a comparative study in three cultures. Child Study Journal, 25, 97-121.

Andersson, S. (1995) Projection systems and x-ray strategies in children's drawings. British Journal of Educational Psychology, 65, 455-464.

Apfeldorf, M. & Smith, J. (1965) The representation of the body self in human figure drawings. Presented at annual meeting of Eastern Psychological Association, Atlantic City, NJ.

Arbit, J., Lakin, M. & Mathis, A. G. (1959) Clinical psychologists' utilisation of human figure drawings. Journal of Clinical Psychology, 15, 325 - 327.

Arkell, R. N. (1976) Naive prediction of pathology from human figure drawings. Journal of School Psychology, 14(2), 114-117.

Arnheim, R. (1974) Art and Visual Perception: A psychology of the creative eye; the new version. Berkeley, CA: University of California Press.

Aronsson, K. & Andersson, S. (1996) Social scaling in children's drawings of classroom life: A cultural comparative analysis of social scaling in Africa and Sweden. British Journal of Developmental Psychology, 14, 301-314.

Babiker, G. (1993) Projective testing in the evaluation of the effects of sexual abuse in childhood: A review. British Journal of Projective Psychology, 38(2), 45-53.

Bachara, G. H., Zaba, J. N. & Raskin, L. M., (1975) Human figure drawings and LD children. Academic Therapy, 11(2), 217 - 222.

Ballard, P. B. (1912) What London children like to draw. Journal of Experimental Pediatrics, 1, 185-197.

Barnett, A. & Henderson, S. E. (1992) Some observations on the figure drawings of clumsy children. British Journal of Educational Psychology, 62, 341-355.

Barrett, M. D. & Light, P. H. (1976) Symbolism and intellectual realism in children's drawings. British Journal of Educational Psychology, 46, 198-202.

Bassett, E. (1976) Production strategies in the child's drawing of the human figure: Towards an argument for a model of syncretic perception. In Butterworth, G. (Ed.), The Child's Representation of the World. New York and London: Plenum.

Bennett, V. (1966) Combinations of figure drawing characteristics related to drawers self-concept. Journal of Projective Techniques, 30, 192-196.

Bensur, B. & Eliot, J. (1993) Case's developmental model and children's drawings. Perceptual and Motor Skills, 76, 371-375.

Berger, R. (1994) Children draw their stepfamilies. Journal of Family Psychotherapy, 5(4), 33-48.

Berman, S. & Laffal, J. (1953) Body type and figure drawing. Journal of Clinical Psychology, 9, 368-370.

Black, J. & Niven, C. (1993) Children's drawings of topics differing in significance: Effects of the stimulus object. Poster presented at the meeting of the developmental section of the British Psychological Society, Birmingham, September.

Bodwin, R. R. & Bruch, M. (1960) The adaptation and validation of the DAP test as a measure of self-concept. Journal of Clinical Psychology, 16, 427-429.

Bombi, A. S. & Pinto, G. (1994) Making a dyad: Cohesion and distancing in children's pictorial representation of friendship. British Journal of Developmental Psychology, 12(4), 563-575.

Bradshaw, D. H. (1952) A study of group consistencies on the Draw-A-Person test in relation to personality projection. Unpublished Master's thesis, Catholic University.

Brewer, T. M. (1995) An examination of untutored thematic and observational drawings made by third and seventh grade students. Visual Arts Research, 21(2), 57-65.

Bricetti, K. A. (1994) Emotional indicators of deaf children on the Draw-A-Person test. American Annals of the Deaf, 139(5), 500-505.

Brittain, W. L. & Chien, Y. C. (1983) Relationship between pre-school children's ability to name body parts and their ability to construct a man. Perceptual and Motor Skills, 57, 19-24.

Brown, E. V. (1977) Reliability of children's drawings in the Goodenough-Harris DAM test. Perceptual and Motor Skills, 44, 739-742.

Buck, J. N. (1948) The H-T-P Technique: A qualitative and quantitative scoring manual. Journal of Clinical Psychology, 4, 319-396.

Buhler, K. (1930) The Mental Development of the Child. London: Routledge and Kegan Paul Ltd.

Burgess, A. W. & Hartman, C. R. (1993) Children's drawings. Special issue: Clinical recognition of sexually abused children. Child Abuse and Neglect, 17(1), 161-168.

Burgess, A. W., McCausland, M. P. & Wolbert, W. A. (1981) Children's drawings as indicators of sexual trauma. Perspectives in Psychiatric Care, 19(2), 50-57.

Burns, R. C. & Kaufman, S. H. (1970) Kinetic family drawings (KFD): An introduction to understanding children through kinetic drawings. New York: Brunner/Mazel.

Buros, O. K. (Ed.) (1972) The Seventh Mental Measurement Yearbook, Vol. 1. Highland Park, NJ: The Gryphon Press.

Burt, C. (1921) Mental and Scholastic Tests. London: King and Son.

Cates, J. A. (1991) Comparison of human figure drawings by hearing and hearing-impaired children. Volta-Review, 93(1), 31-39.

Celotta, B. K. (1973) Knowledge of the human figure as measured by two tasks. Developmental Psychology, 8, 377-381.

Chambers, G. V. (1954) An investigation of the validity of judgements based on "blind" Rorschach records. Dissertation Abstracts International, 14, 2399-2400.

Claparède, E. (1907) (Plan for the collective experiments of the drawing of children). Archives de Psychologie, 6, 276-278.

Cohen, D. N. (1933) The Goodenough scale applied to thirteen year old children. Unpublished Masters thesis, Colombia University.

Cohen-Liebman, M. S. (1995) Drawings as aids in child sexual abuse litigation: A composite list of indicators. Special issue: Sexual abuse. Arts in Psychotherapy, 22(5), 475-483.

Cooke, E. (1885) Art teaching and child nature. London Journal of Education.

Coopersmith, S., Sakai, D., Beardslee, B. & Coopersmith, A. (1976) Figure drawings as an expression of self-esteem. Journal of Personality Assessment, 40(4), 370-375.

Cordell, A. S. & Bergman-Meador, B. (1991) The use of drawings in group intervention for children of divorce. Journal of Divorce and Remarriage, 17(1-2), 139-155.

Costall, A. P. (1989) Another look at Luquet: Stages in our understanding of children's drawings. Presented at the British Psychological Society's Developmental Section Annual Conference, University of Surrey, Guildford, 8-11th September.

Cox, M. V. (1981) One thing behind another: problems of representation in children's drawings. Educational Psychology, 1, 275-287

Cox, M. V. (1992) Children's Drawings. London: Penguin Books.

Cox, M. V. (1993) Children's Drawings of the human figure. Hove, UK: Lawrence Erlbaum Associates.

Cox, M. V. & Batra, P. (n. d.) The position of the torso in children's drawings of the human figure in the UK and in India. Unpublished manuscript, University of York.

Cox, M. V. & Bayraktar, R. (1989) A cross-cultural study of children's human figure drawings. Presented at the Tenth Biennial Conference of the International Society for the Study of Behavioural Development, University of Jyvaskyla, Finland, July.

Cox, M. V. & Lambon Ralph, M. (1996) Young children's ability to adapt their drawings of the human figure. Educational Psychology, 16, 245-255.

Cox, M. V. & Moore, R. (1994) Children's depictions of different views of the human figure. Educational Psychology, 14, 427-436.

Cox, M. V. & Parkin, C. (1986) Young children's human figure drawing: Cross sectional and longitudinal studies. Educational Psychology, 6, 353-368.

Craddick, R. A. (1961) Size of Santa Claus drawings as a function of the time before and after Christmas. Journal of Psychological Studies, 12, 121-125.

Craddick, R. A. (1962) DAP characteristics of psychopathic prisoners and college students. Perceptual and Motor Skills, 15, 11-13.

Craddick, R. A. (1963) The self image in the Draw-A-Person test and self-portrait drawings. Journal of Projective Techniques and Personality Assessment, 27, 288-291.

Crook, C. (1985) Knowledge and appearance. In Freeman, N. H. & Cox, M. V. (Eds.) Visual Order: The nature and development of pictorial representation. Cambridge: Cambridge University Press.

Cummings, J. (1986) Projective drawings. In Knoff, H. M. (Ed.) The Assessment of Child and Adolescent Personality. New York: Guildford.

Currie, S. F., Holtzman, W. H. & Swartz, J. D. (1974) Early indicators of personality traits viewed retrospectively. Journal of School Psychology, 12, 51-59.

Dalby, J. T. & Vale, H. L. (1977) Self-esteem and children's human figure drawings. Perceptual and Motor Skills, 44, 1279-1282.

Darwin, C. (1877) A biographical sketch of an infant. Mind, 11, 286-294.

Daum, J. M. (1983) Emotional indicators in drawings of aggressive or withdrawn male delinquents. Journal of Personality Assessment, 70, 243 - 249.

Delatte, J. G. & Hendrickson, N. J. (1982) Human figure drawing size as a measure of self-esteem. Journal of Personality Assessment, 46, 603-606.

Dickson, J. M., Saylor, C. F. & Finch, A. J. (1990) Personality factors, family structure and sex of drawn figure on the Draw-A-Person test. Journal of Personality Assessment 55(1-2), 362-366.

Dieffenbach, E. W. (1977) Koppitz HFD test: The reliability and clinical validity of its emotional indicators. Dissertation Abstracts International, 38(10-A), 6053.

DiLeo, J. H. (1973) Children's drawings as diagnostic aids. New York: Brunner/Mazel, Inc.

Landsman, M. & Dillard, H. (1967) Evanston Early Identification Scale, manual. Chicago: Follett Educational Corp.

Engle, P. L. & Suppes J. S. (1990) The relation between HFDs and test anxiety in children. Journal of Projective Techniques and Personality Assessment, 34, 223-231.

Eno, L., Elliot, C. & Woehlke, P. (1981) Koppitz emotional indicators in the human-figure drawings of children with learning problems. Journal of Special Education, 15(4), 459 - 470.

Exner, J. E. (1962) A comparison of the human figures drawings of psychoneurotics, character disturbances, normals and subjects experiencing experimentally-induced fear. Journal of Projective Techniques, 26, 392-397.

Falchikov, N. (1990) Youthful ideas about old age: An analysis of children's drawings. International Journal of Ageing and Human Development, 31(2), 79-99.

Fayol, M., Barrouillet, P. & Chevrot, C. (1995) Judgement and production of drawings by three to ten year olds: comparison of declarative and procedural drawing knowledge. European Journal of Psychology of Education, 10, 303-313.

Fenson, L. (1985) The transition from construction to sketching in children's drawings. In Freeman, N. H. & Cox, M. V. (Eds.) Visual Order: The nature and development of pictorial representation. Cambridge: Cambridge University Press.

Fisher, S. & Fisher, R. (1950) Test of certain assumptions regarding figure drawing analysis. Journal of Abnormal and Social Psychology, 45, 727-732.

Flannery, K. A. & Watson, M. W. (1991) Perceived competence in drawing during the middle childhood years. Visual Arts Research, 17(34, no 2), 66-71.

Forrest, M. & Thomas, G. V. (1991) An exploratory study of drawings by bereaved children. British Journal of Clinical Psychology, 30(4), 373-374.

Fox, T. (1989) Some effects of emotional significance of the topic on the size of children's drawings. Unpublished Master's thesis, University of Birmingham.

Fox, T. & Thomas, G. V. (1990) Children's drawings of an anxiety eliciting topic: Effects on the size of the drawing. British Journal of Clinical Psychology, 29(1), 71-81.

Freeman, N. (1975) Drawing conclusions. New Behaviour, 2, 17-19.

Freeman, N. (1980) Strategies of representation in young children: analysis of spatial skills and drawing processes. London: Academic Press.

Fu, V. R. (1981) Analysis of children's self-drawings as related to self-concept. Psychological Reports, 49(3), 941-942.

Fuller, G. B., Preuss, M. & Hawkins, W. F. (1970) Validity of HFDs with disturbed and normal children. Journal of School Psychology, 8, 54-56.

Gardner, H. (1985) Frames of Mind. London: Paladin Books.

Gayton, W. F., Tavormina, J., Evans, H. E. & Schuh, J. (1974) Comparative validity of Harris and Koppitz scoring systems for human figure drawings. Perceptual and Motor Skills, 39, 369-370.

Gellert, E. (1968) Comparison of children's self drawings with their drawings of other persons. Perceptual and Motor Skills, 26, 123-138.

Gillespie, J. (1994) The Projective Use of Mother and Child Drawings. New York: Brunner/Mazel.

Glutting, J. J. & Nester, A. (1986) Koppitz emotional indicators as predictors of kindergarten children's learning related behaviour. Contemporary Educational Psychology, 11, 117-126.

Goh, D. S. & Fuller, G. B. (1983) Current practices in the assessment of personality and behaviour by school psychologists. School Psychology Review, 12, 240-243.

Goldman, R. K. & Warren, D. (1976) Preliminary analyses of human-figure drawings as a kindergarten screening measure. Perceptual and Motor Skills, 43, 58.

Golomb, C. (1973) Children's representation of the human figure: the effects of models, media, and instruction. Genetic Psychology Monographs, 87, 197-251.

Golomb, C. (1974) Young Children's Sculpture and Drawing: A study in representational development. Cambridge, Mass.: Harvard University Press.

Golomb, C. (1992) The child's creation of a pictorial world. Oxford, England: University of California Press.

Goodenough, F. L. (1926) Measurement of intelligence by drawings. Chicago: World Book Company.

Goodenough, F. L. & Harris, D. B. (1950) Studies in the psychology of children's drawings: II 1928-1949. Psychological Bulletin, 47, 369-433.

Goodnow, J. (1977) Children's Drawing. London: Open Books.

Goodnow, J. (1978) Visible thinking: Cognitive aspects of change in drawings. Child Development, 49, 637-641.

Goodnow, J. Wilkins, P. & Dawes, L. (1986) Acquiring cultural forms: cognitive aspects of socialisation illustrated by children's drawings. International Journal of Behaviour Development, 9, 485-505.

Gordon, N., Lefkowitz, M. M. & Tesiny, E. P. (1980) Childhood depression and the Draw-A-Person test. Psychological Reports, 47(1), 251-257.

Gray, D. M. & Pepitone, A. (1964) Effect of self-esteem on drawings of the human figure. Journal of Consulting Psychology, 28, 452-455.

Gresham, F. M. (1993) "What's wrong with this picture?": Response to Motta et al's review of human figure drawings. School Psychology Quarterly, 8, 182-186.

Groves, J. R. & Fried, P. A. (1991) Developmental items on children's human figure drawings: A replication and extension of Koppitz to younger children. Journal of Clinical Psychology, 47(1), 140-148.

Gutman, B. (1952) An investigation of the applicability of the human figure drawing in predicting improvement in therapy. Unpublished Ph.D. Thesis, New York University.

Haas, M. (1978) Private Communication. In Golomb, C. (1992) The Child's Creation of a Pictorial World. Oxford, England: University of California Press.

Hagood, M. M. (1992) Diagnosis or dilemma: Drawings of sexually abused children. British Journal of Projective Psychology, 37(1), 22-33.

Hammer, E. F. (1958) The Clinical Application of Projective Drawings. Springfield, IL: Charles C Thomas.

Hammer, E. F. (1969) DAP: back against the wall? Journal of Consulting and Clinical Psychology, 33, 151-156.

Hammer, E. F. (1976) Acting out and it's prediction by projective drawing assessment. In Abt, L. E. & Weissmen, S. L. (Eds.) Acting Out. New York: Jason Aronson.

Hammer, M. & Kaplan, A. M. (1966) The reliability of children's Human figure drawings. Journal of Clinical Psychology, 22, 316 - 319.

Handler, L. & Reyher, J. (1964) The effects of stress on the Draw-A-Person test. Journal of Consulting Psychology, 28, 259-264.

Handler, L. & Reyher, J. (1966) Relationship between GSR and anxiety indices in projective drawings. Journal of Consulting Psychology, 30, 60-67.

Harris, D. B. (1963) Children's Drawings as Measures of Intellectual Maturity. Harcourt Brace & World, Inc.

Harris, D. B. (1971) The case study. In Gensler, G. (Ed.) A Report on Pre-Conference Education Research Training Programs for Descriptive Research in Art Education. National Art Education Association.

Henderson, J. A. & Thomas, G. V. (1990) Looking ahead: Planning for the inclusion of detail affects relative sizes of head and trunk in children's human figure drawings. British Journal of Developmental Psychology, 8, 383-391.

Hibbard, R. A. & Hartman, G. L. (1990) Emotional indicators in human figure drawings of sexually victimised and nonabused children. Journal of Clinical Psychology, 46(2), 211 - 219.

Hiler, E. W. & Nesvig, D. (1965) An evaluation of criteria used by clinicians to infer pathology from figure drawings. Journal of Consulting Psychology, 29, 520-529.

Holmes, R. (1995) Children's artwork and social distance: An examination of children's racial beliefs. Child Study Journal, 25(1), 23-43.

Holmes, C. B. & Wiederholt, J. (1982) Depression and figure size on the Draw-A-Person test. Perceptual and Motor Skills, 55(3, Pt 1), 825-826.

Holzberg, J. D. & Wexler, M. (1950) The validity of human form drawings as a measure of personality deviation. Journal of Projective Techniques, 14, 343-361.

Hulse, W. C. (1952) Childhood conflict expressed through family drawings. Journal of Projective Techniques, 16, 66-79.

Hurlock, E. B. (1943) The spontaneous drawings of adolescents. Journal of Genetic Psychology, 63, 141-156.

Ivanoff, E. (1909) (Experiments on the drawing of Swiss school children: the correlation between drawing aptitude and other capacities). Archives de Psychologie, 8, 97-156.

Johnson, G. S. (1989) Emotional indicators in the human figure drawings of hearing-impaired children: A small sample validation study. American Annals of the Deaf, 134(3), 205-208.

Joiner, T. E., Schmidt, K. L. & Barnett, J. (1996) Size, detail and line heaviness in children's drawings as correlates of emotional distress: (more) negative evidence. Journal of Personality Assessment, 67(1), 127-141.

Jolley, R. P. (1995) Children's production and perception of visual metaphors for mood and emotion in line drawings and art. Unpublished Doctoral thesis, University of Birmingham.

Jones, S. (1972) An investigation into the conservative nature of children's drawings. Unpublished Honours thesis, University of New South Wales, Australia.

Jones, C. J. (1992) Human figure drawings of mildly handicapped students: Learning disabled, mildly mentally retarded, emotionally disturbed and speech/language impaired. Springfield, IL: Charles C Thomas.

Jones, L. W. & Thomas, C. B. (1961) Studies on figure drawings: A review of the literature (1949 - 1959). The Psychiatric Quarterly Supplement, 35, 212-261.

Kahill, S. (1984) Human figure drawing in adults: An update of the empirical evidence 1967-1982. Canadian psychology, 25, 269-292.

Kahn, M. W. & Jones, N. F. (1965) Human figure drawings as predictors of admission to a psychiatric hospital. Journal of Projective Techniques, 29, 319-322.

Kamano, D. K. (1960) An investigation of the meaning of human figure drawing. Journal of Clinical Psychology, 16, 429-430.

Kamphaus, R. W. & Pleiss, K. L. (1991) DAP techniques: Tests in search of a construct. Journal of School Psychology, 29, 395-401.

Karmiloff-Smith, A. (1990) Constraints on representational change: Evidence from children's drawing. Cognition, 34, 57-83.

Kellogg, R. (1970) Analysing children's art. Palo Alto, CA: Mayfield.

Kellogg, R. (1979) Children's drawings, children's minds. New York: Avon.

Kennedy, J. M. (1980) Blind people recognising and making pictures. In Hagen, M. A. (Ed.) The Perception of Pictures. New York: Academic Press.

Kennedy, J. M. (1983) What can we learn about pictures from the blind? American Scientist, 71, 19-26.

Kidd, A. H. & Kidd, R. M. (1995) Children's drawings and attachment to pets. Psychological Reports, 77(1), 235-241.

Klein, R. G. (1986) Questioning the clinical usefulness of projective psychological tests for children. Journal of Developmental and Behavioural Pediatrics, 7(6), 378-382.

Klopfer, W. G. & Taulbee, E. S. (1976) Projective Tests. Annual Review of Psychology, 27, 543-567

Koppitz, E. M. (1966) Emotional indicators on human figure drawings of shy and aggressive children. Journal of Clinical Psychology, 22, 466-469.

Koppitz, E. M. (1968) Psychological evaluation of children's human figure drawings. The Psychological Corporation, Harcourt Brace Jovanovich, Inc.

Koppitz, E. M. (1984) Psychological evaluation of human figure drawings of middle school pupils. New York: Grune & Stratton.

Kosslyn, S. M., Heldmeyer, K. H. & Locklear, E. P. (1977) Children's drawings as data about internal representations. Journal of Experimental Child Psychology, 23(2), 191-211.

Kotkov, B. & Goodman, M. (1953) The Draw-A-Person tests of obese women. Journal of Clinical Psychology, 9, 362-364.

Krampen, M. (1991) Children's Drawings: Iconic coding of the environment. New York: Plenum Press.

Kurdek, L. A. & Darnel-Goetschel, G. (1987) Young adolescents human figure drawings as indicators of psychopathology. Journal of Adolescent Research, 2, 69-74.

Lamprecht, K. (1906) (Children's drawings as an historical source). Bulletin de l'Academie Royale de Belgique, Classe des Lettres, 9-10, 457-469.

Lehner, G. F. J. & Gunderson, E. K. (1952) Reliability of graphic indices in a projective test (Draw-A-Person). Journal of Clinical Psychology, 8, 125-128.

Lehner, G. F. J. & Gunderson, E. K. (1953) Height relationships on the Draw-A-Person test. Journal of Personality, 21, 392-399.

Leichtman, S. R., Burnett, J. W. & Robinson, H. M. (1981) Body image concerns of psoriasis patients as reflected in human figure drawings. Journal of Personality Assessment, 45(5), 478-484.

Levy, L. R. (1931) The function of the Goodenough drawing scale in the study of high school freshmen. Unpublished Masters thesis, Colombia University.

Lewinsohn, P. M. (1964) Relationship between height of figure drawings and depression in psychiatric patients. Journal of Consulting Psychology, 28, 380-381.

Lewinsohn, P. M. (1965) Psychological correlates of overall quality of figure drawings. Journal of Consulting Psychology, 29, 504-512.

Lingren, R. H. (1971) Attempted replication of emotional indicators in HFDs by shy and aggressive children. Psychological reports, 29(1), 35-38.

Lubin, B., Larsen, R. M. & Matarazzo, J. D. (1984) Psychological test usage in the United States: 1935-1969. American Psychologist, 39, 451-454.

Lubin, B., Larsen, R. M., Matarazzo, J. D. & Seever, M. (1985) Psychological test usage in five professional settings. American Psychologist, 40 (7), 857-861.

Lukens, H. (1896) A study of children's drawings in the early years. Pedagogical Seminary, 4, 79-110.

Lourenso, S., Greenberg, J. & Dunn, M. B. (1965) Personality characteristics in drawings of deprived children who differ in school achievement. Journal of Educational Research, 59 63-67.

Luquet, G. H. (1913) Les Dessins d'un Enfant. Paris: Alcan.

Luquet, G. H. (1927) Le dessin enfantin. Paris: Alcan.

Lyons, S. J. (1993) Art psychotherapy evaluations of children in custody disputes. Arts in Psychotherapy, 21(2), 153-159.

Machover, K. (1949) Personality projection in the drawing of the human figure. Springfield, IL: Charles C Thomas.

Machover, K. (1953) HFDs of children. Journal of Projective Techniques, 17, 85-91.

Maitland, L. (1895) What children draw to please themselves. Inland Educator, 1, 87.

Malchioldi, C. (1990) Breaking the Silence: Art therapy with children from violent homes. New York: Brunner/Mazel.

Maloney, M. P. & Glasser, A. (1982) An evaluation of the clinical utility of the DAP test. Journal of Clinical Psychology, 38, 183-190.

Mann, B. S. & Lehman, E. B. (1976) Transparencies in children's human figure drawings: A developmental approach. Studies in Art Education, 18, 41-48.

Matthews, J. (1983) Children Drawing: Are young children really scribbling? Paper presented at the British Psychological Society International Conference on Psychology and the Arts, Cardiff.

McCarthy, D. (1944) A study of the reliability of the Goodenough Drawing Test of intelligence. Journal of Psychology, 18, 201-216.

McHugh, A. (1966) Children's figure drawings in neurotic and conduct disturbances. Journal of Clinical Psychology, 22, 219 - 221.

McNeish, T. J. & Naglieri, J. A. (1993) Identification of individuals with serious emotional disturbance using the Draw A Person: Screening Procedure for Emotional Disturbance. Journal of Special Education, 27, 115 - 121.

Melamed, B. G. & Siegel, L. J. (1975) Reduction of anxiety in children facing hospitalisation and surgery by use of filmed modelling. Journal of Consulting and Clinical Psychology, 43, 511-521.

Millar, S. (1975) Visual experience or translation rules? Drawing the human figure by blind and sighted children. Perception, 4, 363-371.

Miller, T. W., Veltkamp, L. J., Lane, J. & Janson, D. (1987) Projective measures in the clinical evaluation of sexually abused children. Child Psychiatry and Human Development, 18, 47-57.

Morris, W. W. (1949) Methodological and normative considerations in the use of drawings of human figures as a projective method. American Psychologist, 4, 267.

Mortensen, K. V. (1984) Children's human figure drawings. Copenhagen: Dansk Psykologisk Forlag.

Mortensen, K. V. (1991) Form and content in children's human figure drawings: Development, sex differences and body experience. New York: New York University Press.

Motta, R. W., Little, S. G. & Tobin, M. I. (1993) The use and abuse of human figure drawings. School Psychology Quarterly, 8(3), 162 - 169.

Motta, R. W., Little, S. G. & Tobin, M. I. (1993) A picture is worth less than a thousand words: Response to reviewers. School Psychology Quarterly, 8(3), 197 - 199.

Naglieri, J. A. (1988) Draw-A-Person: A Quantitative Scoring System. The Psychological Corporation, Harcourt Brace Jovanovich, Inc.

Naglieri, J. A. (1993) Human figure drawings in perspective. School Psychology Quarterly, 8(3), 170 - 176.

Naglieri, J. A., McNeish, T. J. & Bardos, A. N. (1993) Draw-A-Person: Screening Procedure for Emotional Disturbance. Austin, TX: ProEd.

Naglieri, J. A. & Pfeiffer, S. I. (1992) Performance of disruptive behaviour disordered and normal samples on the DAP: SPED. Psychological Assessment, 4, 156-159.

Neale, E. L. & Rosal, M. L. (1993) What can art therapists learn from the research on projective drawing techniques for children? A review of the literature. The Arts in Psychotherapy, 20, 37-49.

Nichols, R. & Strumpfer, D. (1962) A factor analysis of DAP test scores. Journal of Consulting Psychology, 26, 165-171.

Norford, B. C. & Barakat, L. P. (1990) The relationship of human figure drawings to aggressive behaviour in pre-school children. Psychology in the Schools, 27(4), 318-325.

Oakley, C. A. (1930) The interpretation of children's drawings. British Journal of Psychology, 21, 256-270.

Olsen, J. (1992) Evaluating young children's cognitive capacities through computer versus hand drawings. Scandinavian Journal of Psychology, 33(3), 193-211.

Oullette, S. E. (1988) The use of projective drawing techniques in the personality assessment of prelingually deafened young adults: A pilot study. American Annals of the Deaf, 133(3), 212-218.

Paget, G. W. (1932) Some drawings of men and women made by children of certain non-European races. Journal of the Royal Anthropological Institute, 62, 127-144.

Paine, S. (1992) Conflicting Paradigms of vision in drawing development research. In Thistlewood, D. (Ed.) Drawing, Research and Development. Longman/NSEAD.

Payne, M. A. (1990) Effects of parental presence/absence on size of children's HFDs. Perceptual and Motor Skills, 70, 843-849.

Pemberton, E. F. (1990) Systematic errors in children's drawings. Cognitive Development, 5(4), 395-404.

Pemberton, E. F. & Nelson, K. E. (1987) Using interactive graphic challenges to foster young children's drawing ability. Visual Arts Research, 13, 29-41.

Perry, L. (1992) Towards a definition of drawing. In Thistlewood, D. (Ed.) Drawing, Research and Development. Longman/NSEAD.

Pfeffer, K. & Olowu, A. (1986) Effects of socio-economic differences on the sophistication of Nigerian children's HFDs. Perceptual and Motor Skills, 62, 771-774.

Pfeffer, C. R. & Richman, J. (1991) Human figure drawings: An auxiliary diagnostic assessment of childhood suicide potential. Comprehensive Mental Health Care, 1(2), 77-90.

Phillips, W. A., Inall, M. & Lauder, E. (1985) On the discovery, storage and use of graphic descriptions. In Freeman, N. H. & Cox, M. V. (Eds.) Visual Order: The nature and development of pictorial representation. Cambridge: Cambridge University Press.

Piaget, J. & Inhelder, B. (1956) The child's conception of space. London: Routledge & Kegan Paul.

Pihl, R. O. & Nimrod, G. (1976) The reliability and validity of the Draw-A-Person test in IQ and personality assessment. Journal of Clinical Psychology, 32(2), 470-472.

Poster, E. C., Betz, C., McKenna, A. & Mossar, M. (1986) Children's attitudes towards the mentally ill as reflected in human figure drawings and stories. Journal of the American Academy of Child Psychiatry, 25(5), 680-686.

Prout, H. (1983) School psychologists and social-emotional assessment techniques: Patterns in training and use. School Psychology Review, 12, 377-383.

Prout, H. & Phillips, P. D. (1974) A clinical note: The kinetic school drawing. Psychology in the Schools, 11(3), 303-306.

Prytula, R. E. & Thompson, N. D. (1973) Analysis of emotional indicators in human figure drawings as related to self-esteem. Perceptual and Motor Skills, 37, 795-802.

Rae, W. A. (1991) Analysing the drawings of children who are physically ill and hospitalised, using the ipsative method. Children's Health Care, 20(4), 198-207.

Reuning, H. & Wortley, W. (1973) Psychological studies of the Bushmen. Psychologia Africana Monograph Supplement, No. 7.

Richman, J. & Pfeffer, C. R. (1977) Figure drawings in the evaluation of suicidal latency age children. Presented at the ninth annual meeting of the International Association of Suicide Prevention and Crisis Intervention, Helsinki, Finland, June, 1977.

Richey, M. H. (1965) Qualitative superiority of the "Self" figure in children's drawings. Journal of Clinical Psychology, 21, 59-61.

Riordan, R. J. & Verdel, A. C. (1991) Evidence of child abuse in children's art products. School Counsellor, 39(2), 116-121.

Roback, H. B. (1968) Human figure drawings: Their utility in the clinical psychologists armamentarium for personality assessment. Psychological Bulletin, 70, 1-19.

Rouma, G. (1913) Le Language Graphique de l'Enfant. Paris: Misch et Thron.

Rubin, J. A., Schachter, J. & Ragins, N. (1983) Intra-individual variability in human figure drawings: A developmental study. American Journal of Orthopsychiatry, 53(4), 654-667.

Sadowski, P. M. & Loesch, L. C. (1993) Using children's drawings to detect potential child sexual abuse. Elementary School Guidance and Counseling, 28(2), 115-123.

Scarlett, G., Fucigna, C. & Finkelstein, E. (1980) The expression of qualities and feelings in young children's drawings and block constructions. Unpublished manuscript, Harvard Project Zero.

Schetky, D. & Benedek, E. (1992) Clinical handbook of child psychiatry and the law. Baltimore, MD: Williams & Wilkins.

Schuyten, M. C. (1904) De oorspronkelijke 'Ventjes' der Antwerpsche Schoolkindern. (The early 'little men' of Antwerp school children). Paedologisch Jaarboek, 5, 1-87.

Sechrest, L. & Wallace, J. (1964) Figure drawings and naturally occurring events: Elimination of the expansive euphoria hypothesis. Journal of Educational Psychology, 53, 42-44.

Selfe, L. (1983) Normal and anomalous representational drawing ability in children. London: Academic Press.

Silverstein, A. B. & Robinson, H. A. (1956) The representation of orthopedic disability in children's figure drawings. Journal of Consulting Psychology, 20, 333-341.

Sims, J., Dana, R. H. & Bolton, B. (1983) The validity of the Draw-A-Person test as an anxiety measure. Journal of Personality Assessment, 47(3), 250-257.

Sitton, R. & Light, P. (1992) Drawing to differentiate: flexibility in young children's HFDs. British Journal of Developmental Psychology, 10, 25-33.

Smith, P. M. (1993) Young children's depiction of contrast in human figure drawing: Standing and walking. Educational Psychology, 13(2), 107-118.

Soliscamara, P. & Mata, L. G. (1985) Children's HFDs and impulsive style at two levels of SES. Perceptual and Motor Skills, 61, 1039-1042.

Solley, C. M. & Haigh, G. (1957) A note to Santa Claus. Topeka Research Papers, The Meninger Foundation, 18, 4-5.

Soutter, A. (1994) A comparison of children's drawings from Ireland and Oman. Irish Journal of Psychology, 15(4), 587-594.

Stefanatou, A. & Bowler, D. (1997) Depiction of pain in the self drawings of children with sickle cell disease. Child: Care, Health and Development, 23(2), 135-155.

Stricker, G. (1967) Actuarial, naive clinical and sophisticated clinical prediction of pathology from figure drawings. Journal of Consulting Psychology, 31, 492-494.

Sturner, R. A., Rothbaum, F., Visintainer, M. & Wolfer, J. (1980) The effects of stress on children's human figure drawings. Journal of Clinical Psychology, 36, 324-331.

Sully, J. (1895) Studies of Childhood. London: Longman's, Green & Co.

Swensen, C. H. (1957) Empirical evaluations of human figure drawings. Psychological Bulletin, 54, 431-466.

Swensen, C. H. (1968) Empirical evaluations of human figure drawings: 1957-1966. Psychological Bulletin, 70, 20-44.

Taylor, M. & Bacharach, V. R. (1981) The development of drawing rules. Metaknowledge about drawing influences performance on non-drawing tasks. Child Development, 52, 373-375.

Tharinger, D. J. & Stark, K. (1990) A qualitative versus quantitative approach to evaluating the DAP and KFD: A study of mood and anxiety-disorder children. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2(4), 365-375.

Thomas, G. V. (1995) The role of drawing strategies and skills. In Thomas, G. V. & Lange-Küttner, C. (Eds.) Drawing and Looking: Theoretical approaches to pictorial representation in children. London: Harvester Wheatsheaf.

Thomas, G. V., Chaigne, E. & Fox, T. J. (1989) Children's drawings of topics differing in significance: Effects on size of drawing. British Journal of Developmental Psychology, 7(4), 321-331.

Thomas, G. V. & Gray, R. (1992) children's drawings of topics differing in emotional significance - effects on placement relative to a self drawing: a research note. Journal of Child Psychology and Psychiatry, 33, 1097-1104.

Thomas, G. V. & Jolley, R. P. (1998) Drawing conclusions: A re-examination of empirical and conceptual bases for psychological evaluation of children from their drawings. British Journal of Clinical Psychology, 37(2), 127-139.

Thomas, G. V. & Silk, A. M. J. (1990) An Introduction to the Psychology of Children's Drawings. London: Harvester Wheatsheaf.

Thomas, G. V. & Tsalimi, A. (1988) Effects of order of drawing head and trunk on their relative sizes in children's human figure drawings. British Journal of Developmental Psychology, 6, 191-203.

Tolor, A. & Digrazia, P. V. (1977) The body image of pregnant women as reflected in their HFDs. Journal of Clinical Psychology, 33(2), 566-571.

Trautner, H. M., Lohaus, A., Sahm, W. B. & Helbing, N. (1989) Age-graded judgements of children's drawings by children and adults. International Journal of Behavioural Development, 12(4), 421-431.

Trevisan, M. S. (1996) Review of the Draw-A-Person: Screening Procedure for Emotional Disturbance. Measurement and Evaluation in Counselling and Development, 28, 225-228.

Trowbridge, M. M. (1995) Graphic indicators of sexual abuse in children's drawings: A review of the literature. Special Issue: Sexual abuse. Arts in Psychotherapy, 22(5), 485-493.

Truhon, S. A. (1990) Drawings of the president: 1980 and 1984. Perceptual and Motor Skills, 71(2), 411-414.

Tryphon, A. & Montangero, J. (1992) The development of diachronic thinking in children: Children's ideas about changes in drawing skills. International Journal of Behavioural Development, 15(3), 411-424.

Ulman, E. & Levy, B. I. (1973) Art therapists as diagnosticians. Paper presented at the annual meeting of the American Art Therapy Association, Colombus, Ohio.

Unruh, A. McGrath, P., Cunningham, J. S. & Humphreys, P. (1983) Children's drawings of their pain. Pain, 17(4), 385-392.

Urban, W. H. (1963) The Draw-A-Person Catalogue for Interpretive Analysis. LA: Western Psychological Services.

Van Dyne, W. T. & Carskadon, T. G. (1978) Relationships among three components of self concept and same-sex and opposite-sex human figure drawings. Journal of Clinical Psychology, 34, 537-538.

Van Sommers, P. (1984) Drawing and Cognition. Cambridge University Press.

Vane, J. R. & Eisen, V. W. (1962) The Goodenough Draw-a-Man test and signs of maladjustment in kindergarten children. Journal of Clinical Psychology, 18, 276 - 279.

Vane, J. R. & Kessler, R. T. (1964) The Goodenough Draw-A-Man test: Long term reliability and validity. Journal of Clinical Psychology, 20, 487-488.

Wallach, M. A. & Leggert, M. I. (1972) Testing the hypothesis that a person will be consistent: Stylistic consistency versus situational specificity in size of children's drawings. Journal of Personality, 40, 309-330.

Wanderer, Z. W. (1969) Validity of clinical judgements based on human figure drawings. Journal of Consulting and Clinical Psychology, 33, 143-150.

Weber, J. A., Cooper, K. & Hesser, J. L. (1996) Children's drawings of the elderly: Young ideas abandon old age stereotypes. Art Therapy, 13, 114-117.

Wechsler, D. (1990) Wechsler Pre-school and Primary Scale of Intelligence. The Psychological Corporation.

Wechsler, D. (1992) Wechsler Intelligence Scale for Children. The Psychological Corporation.

Whitmyre, J. W. (1953) The significance of artistic excellence in the judgement of adjustment inferred from human figure drawings. Journal of Consulting Psychology, 17, 421-422.

Williams, J. H. (1935) Validity and reliability on the Goodenough intelligence test. School and Society, 41, 653-656.

Wilson, B. & Ligtvoet, J. (1992) Across time and cultures: Stylistic changes in the drawings of Dutch children. In Thistlewood, D. (Ed.) Drawing, Research and Development. Longman/NSEAD.

Wilson, B. & Wilson, M. (1985) The artistic tower of Babel: Inextricable links between culture and graphic development. Visual Arts Research, 11(1), 90-104.

Winner, E. & Gardner, H. (1981) The art in children's drawings. Review of Visual Arts Research, 14, 91-105.

Wright, S. K. & Ashman, A. F. (1991) The use of symbols in drawings by children, nondisabled adults and adults with an intellectual disability, Special issue: Perspectives on exceptionality. Developmental Disabilities Bulletin, 19(2), 105-128.

Wysocki, B. & Whitney, E. (1965) Body image of crippled children as seen in DAP test behaviour. Perceptual and Motor Skills, 21, 499-504.

Yama, M. F. (1990) The usefulness of human figure drawings as an index of overall adjustment. Journal of Personality Assessment, 54, 78-86.

Yepsen, T. N. (1929) The reliability of the Goodenough Drawing Test with feeble-minded subjects. Journal of Educational Psychology, 20, 448-451.

Zuk, G. H. (1962) Relation of mental age to size of figures on the Draw-A-Person test. Perceptual and Motor Skills, 14, 410.