
Indicators of Feminine Gender Identity in Latency-aged Boys in the Draw a Person and the Rorschach Tests



Hagit Benziman and Selina Marodes

Hebrew University, Jerusalem

Seven diagnostic indicators were examined for their potential to identify latency aged boys who had feminine gender identity. Nineteen emotionally disturbed boys (average age 10.2) with feminine gender identity were compared on these indicators with 21 control boys. The indicators which were found to be valid were drawing the female figure with a greater value than the male figure in the DAP; difficulty in identification with the male figure in stories to the DAP; a low relative rate of male human responses and/or a high relative rate of female human responses in the Rorschach; and initial butterfly response to card V in the Rorschach. Using these indicators in combination reduced the false positive and negative rates associated with each individual indicator. © 1997 John Wiley & Sons, Inc.

Most authors in the field of gender identity disturbance have focused on boys. According to the literature, professional help is most often sought for boys with a feminine gender identity (Zucker, 1985). Effeminate behavior in males is less acceptable in society than masculine behavior in females; boys showing cross-gender behaviors attract more attention than do such girls and are thus more likely to be brought for treatment (Rekers, 1981). There is therefore an especial need for the development of tools suitable for the identification of feminine gender identity in boys.

When discussing and developing methods for assessment of gender identity, it is important to understand that gender identity is not a unitary concept. It differs in meaning according to which aspect of the individual one is assessing (Biller, 1974; Tyson, 1982). Tyson (1982) defined three components of sexual identity: First, in order of development, comes the unconscious and conscious sense of belonging or wishing to belong to one particular sex, the "core gender identity." Second, the behavioral aspect of gender identity develops—the so-called "gender role," defined as the manner in which the individual interacts with others with respect to cultural guidelines for sex role. Finally comes the "sexual partner orientation"—that is, the sex of the chosen love object. Clearly, these aspects of gender identity do not develop in discrete stages: The core gender identity continues to be consolidated even after gender role has

begun to develop; likewise, gender role continues to develop while sexual partner orientation is being established. Similarly, the foundations of gender role and sexual partner orientation are laid in the infant's earliest object relationships, even though these aspects of gender identity are only firmly established much later.

Despite the longstanding awareness of the above distinctions between these strands of gender identity, many researchers in the field have not heeded them in their work. Many have not distinguished between results of behavioral assessment, assessment of personality traits, and assessment using projective tests, although each type of test taps a different aspect of gender identity. Most of the tests developed specifically to assess gender identity, in both adults and children, are in the form of questionnaires that concentrate on the appraisal of masculine and feminine personality traits and behavior, primarily assessing gender role. Following the understanding that masculinity and femininity are not necessarily opposite poles of a linear scale, questionnaires were developed that also made it possible to classify individuals as androgynous or undifferentiated. The most commonly used of these tests are the Bem Sex Role Inventory (Bem, 1974) and the Personality Attributes Questionnaire (Spence et al., 1979), both of which have been adapted for use with children.

Tests have also been developed specifically for children, concentrating mostly on the assessment of the child's behavior either by the use of questionnaires on behavior in general or favorite games in particular, or by observation of the child in structured play situations (Brown, 1956; Meyer-Bahlburg et al., 1985; Rekers & Morey, 1989). Though popular, these tests are problematic for two main reasons. First, as suggested above, they assess only the behavioral aspects of gender identity, the gender role (with the exception of Brown [1956]); yet they are often used to draw conclusions about gender identity as a whole. Second, many of these tests are appropriate only for use with young children, since they rely on the use of games or questions about games suitable only for younger children. Older children are liable to understand the point of the test and therefore hide, or lie about, their natural behavior. It should be noted, too, that these tests are especially unreliable for the assessment of the gender identity of girls of latency age, as it is normal for girls of this age to show ambivalence or even a preference for the masculine role while actually being identified with the feminine role (Brown, 1956).

Partly in response to these problems, researchers in the field of gender identity have also used traditional projective tests, primarily the Draw-a-Person and the Rorschach tests. The nature of the Draw-a-Person Test (DAP) makes it, among the major projective tests, a particularly appropriate tool for the assessment of gender identity. Subjects are asked first to draw a person—that is, a male or female figure—and then a figure of the opposite sex from the one first drawn.

Much of the research on gender identity and the DAP test has concentrated on examining the significance of the sex of the first figure drawn, following the claim of Machover (1949) that drawing a person of the opposite sex first is symptomatic of an inverted sexual identity. Most research has concentrated on the general population or on boys with a feminine sexual identity of homosexual men, and has shown equivocal results (Abraham, 1979, 1989; Baker et al., 1953; Brown & Tolor, 1957; Granick & Smith, 1953; Green et al., 1972; Hammer, 1954; Kurtzburg et al., 1966; Lefkowitz, 1962; Levy, 1959; de Martino, 1954; Skilbeck et al., 1975; Wysocki & Wysocki, 1977; Zucker et al., 1983). As a general conclusion from the studies, it seems that feminine-identified boys as a group draw a female figure first more often than controls, but it is clear that this indicator gives rise to many false positives and negatives—especially among older latency children—and should not be used as the sole indicator of inverted gender identity.

Other possible indicators of inverted gender identity which were subject for research were the relative size of the drawing (Skilbeck et al., 1975; Zucker et al., 1983) and the number of articles of clothing drawn (Skilbeck et al., 1975). But the studies using these indicators have

also shown equivocal results. Thus, on their own, neither drawing a female figure first, nor drawing it larger or more clothed, are reliable indicators of feminine gender identity.

From the above it can be seen that, as suggested by Brown and Tolor (1957), there is a need to test whether the presence of a combination of these phenomena can be a reliable indicator.

An additional aspect of the drawing test that has not yet been investigated in connection with gender identity is the stories children tell about their drawings. The subjects are asked to tell such stories without being given detailed instructions. The stories involve the male and female figures drawn from the subjects' imagination, so that the subjects can easily project their own feelings and preoccupations onto them. Hence, the stories are likely to reflect the subjects' feelings towards each sex. In particular, boys with a feminine sexual identity are likely to portray the male figure less positively in their stories and identify less easily with him (just as they draw the male figure less positively) and more easily with the female figure. This hypothesis has not been tested in the literature, but deserves investigation both in order to help us understand how gender identity is reflected in projective tests and to improve methods of diagnosis.

This is especially necessary since many adults and even some children do not find it easier to express themselves in drawing rather than verbally.

The Rorschach test, which is very different from the DAP, does not expressly require the subject to relate to male and female figures, although some of the cards have been suggested to be symbolic or evocative of one sex or the other.

Most of the research that has attempted to identify problems of gender identity in the Rorschach test has investigated the question of whether there are responses or response patterns typical of male homosexuals (Davids et al., 1956; Goldfried, 1966; Hendlin, 1976; Hooker, 1958; Shafer, 1954; Wheeler, 1949). All the researchers stress the danger of blindly relying on their suggested signs to identify protocols of individual homosexuals because of the high rate of false positives and negatives that would be obtained.

Only the relatively high frequency of feminine sex-typed responses labeled "feminine orientation" by Shafer or Wheeler's sign 20 (feminine clothing) may be relevant to the question of expression of feminine gender identity in the Rorschach by boys.

Several researchers have indeed found that the number of responses of this kind do differentiate between boys with a feminine identity and others. For example, Tuber and Coates (1985) found a higher than normal rate of female, female object, and gender confusion responses among such boys. This finding was replicated by Zucker et al. (1992) in a much better controlled study where the group of gender-confused children (70 boys and 9 girls with an average age of approximately 8 years who had been referred to a clinic specializing in gender disorders) were compared with siblings and with psychiatric and normal controls. Thirty-five percent of the experimental group gave more cross-sex responses than same sex responses, compared to 15% of psychiatric controls and 18% of "normal" controls. The sibling group, however, showed response patterns similar to the experimental group.

Once again, although these differences are significant, the criteria used identify only 35% of the gender-disturbed children and give a 20% false positive rate. Therefore, the relative number of cross-sex to same-sex response in the Rorschach is an inadequate diagnostic tool if used on its own, but would be useful in combination with other diagnostic indicators such as those mentioned above. In addition, Tuber and Coates (1985) have also found evidence of disturbance of object-relations in feminine-identified boys as expressed in the responses involving human and pseudo-human forms (H or (H)). Therefore, examination of the pattern of H responses where the gender is specified may be more revealing; Tuber and Coates' (1985) findings imply that the differences may be more marked among H responses than among gender-related responses in general. In particular, a measure of ease of identification with male and female figures could be obtained by investigating the rates of male and female human responses.

Clearly the focus of such an investigation should be on pure human responses—H—and not quasi-human responses—(H)—since the aim is to assess ease of identification with real human figures.

Other measures of gender identity in the Rorschach test relate to the specific salience of some cards. Card IV is traditionally described as the “father” card because of its supposed tendency to elicit attitudes towards the subject’s father and masculinity per se.

The literature is equivocal as to the validity of the assumptions about the symbolic meaning of Card IV. Aranow and Reznikoff (1976) concluded, in a review of the literature then available, that this card is often associated with the father figure and male authority, and therefore may elicit related feelings; but such interpretations require additional support. In particular, since this card is generally regarded as an unpleasant and frightening stimulus, negative emotional responses to it should not be blindly interpreted as negative feelings towards the father figure. Liaboe and Guy (1985) pointed out the many flaws in the studies they examined in their review of the literature. They concluded that only among disturbed individuals, particularly emotionally disturbed children, does Card IV often elicit associations connected with father-child relatives. This may be because feelings towards one’s parents are particularly salient and conflicted in such individuals and the subject is therefore more responsive to certain stimulus qualities of the card (Hayden, 1981). Much less research has been carried out regarding the association of this card with masculinity; this connection, though widely accepted, can only be postulated.

In particular, it would be useful to investigate responses to the W or D7 detail, since these are the areas that give the popular human or human-like responses, which is frequently male (Exner, 1974; classification system: *ibid.*). Furthermore, given Hayden’s findings, it could be of particular value to investigate the responses to Card IV in a clinical group. Marked sex differences among children in responses to this card have been noted: girls aged 10 and above tend more often than boys to relate to the whole image (W) or to D7 (as classified by Exner, 1974) as evocative of a dark and threatening figure; they give more dragon, giant, or monster responses and more often describe figures as sitting on a stump (Ames et al., 1959).

If any pattern of responses can be predicted here, given the sparsity of the literature, it is that the boys with feminine gender identity should, like girls, give more responses in the category of giants, dragons, and monsters than do other boys. Conversely, perhaps “feminine” boys should give more female responses than controls, precisely because of their feminine identification.

Card VII is also traditionally associated with gender identity; it is regarded as the “mother” card and expected to elicit associations with femininity. The literature relating to Card VII is even more equivocal than that relating to Card IV. It appears that, in contrast to Card IV, this card generally elicits positive emotional responses, but there is very little evidence that the card evokes attitudes towards the mother.

The only sex differences in responses that have been noted are that girls give more human responses than do boys (Ames et al., 1959). It does seem that, among adults, this card has the largest preponderance of female percepts (Aranow & Reznikoff, 1976). In spite of the lack of consensus in the literature, many clinicians blindly interpret responses to Card VII as indicative of attitudes towards the mother in particular and femininity in general. Therefore the investigation of the relative frequency of female responses to Card VII in boys with and without a feminine gender identity could be enlightening, but so far there is insufficient evidence to justify any expectation that Card VII can be used as a valid indicator of gender identity. Although female responses are given to other details, the popular female response is to D1 or D2 (Exner, 1974), and an investigation of these D responses would therefore be of most use.

Another possible method of appraisal in the Rorschach is to investigate the response to Card V, where the popular response is almost universally a butterfly or a bat. Among children

it has been found that there is a difference in response pattern to this card between boys and girls. Ames et al. (1952) found that 3 to 10 year-old boys give twice as many bat responses on all cards as girls of these ages; on Card V the boys give more bat responses to the whole image (W) and the girls give more butterfly responses. This is not surprising given that butterflies are generally regarded as delicate and beautiful, qualities traditionally associated with femininity and encouraged in girls, while bats are associated with power and aggression, traditionally more masculine qualities. Thus it would be expected that boys with a feminine gender identity, like girls, would give more butterfly responses to Card V than boys with no gender identity problems.

In summary, there is a collection of signs in the DAP and the Rorschach that have traditionally been used as diagnostic indicators of feminine gender identity among boys and men. Research carried out on the validity of some of these indicators has shown equivocal results. Since these indicators are often used by clinicians, it is important to check their validity.

In addition, while many studies have shown that these signs can be used successfully to distinguish between groups, few tests have been performed to find out if they can be used to distinguish between individuals. The few tests that have been performed have yielded high rates of false positives and negatives. Therefore, the aim of this study was twofold: to investigate the validity of each indicator on an individual, as opposed to a group, level; and to investigate the use of all valid indicators from both the DAP and the Rorschach in combination. The purpose of the latter investigation was to see whether such an approach could reduce the false positive rates without drastically affecting the false negative rates of the test. This methodology, as outlined earlier, makes it possible to locate as many individuals with feminine gender identity as possible by using indicators that match the different habitual modes of response of different individuals. In addition, because repetition of similar response patterns is used as a necessary criterion, false identification of individuals who respond on only one or two of the indicators in a manner indicative of feminine gender identity is thereby prevented. By using a combination of several indicators it may also be possible to assess the severity of the problem in different individuals.

METHOD

Subjects

Nineteen latency aged boys with a feminine gender identity were chosen for the experimental group in this study. They were chosen by two professionals from boys who have been treated at the Bnei Brith Institution for disturbed boys in Jerusalem in the last 11 years. The professionals—the head social worker and the head of the institution, who is a trained psychologist and psychoanalyst—worked at the institution throughout this period and were therefore very familiar with the children. They used their wealth of clinical skills and knowledge to judge each child independently on the basis of both external behavior and psychological conflicts. Only those children about whom both professionals agreed were included in this study. The average age of the experimental group at time of testing was 10 years 2 months (*SD* 8 months) and their average IQ was 96.6 (*SD* 15.5).

The control group was made up of 21 children who, according to the same professionals, did not have problems specifically in the area of gender identity. They were individually matched with the experimental group on age and IQ. The average age of the control group at the time of testing was 10 years 3 months and their average IQ was 97.2 (*SD* 13.20). *t*-tests revealed that there were no significant differences between the groups with respect to age or IQ ($t = .24$ and $.13$ respectively, $p > .20$).

Procedures and Hypotheses

All subjects had been given a battery of psychological tests, including the WISC, DAP, and Rorschach, by a trained psychologist. In all cases except four of the experimental group and three of the control group, the examiners were female. A battery of psychological tests is routinely given to all boys of this institution within the first two years of their stay and again near the end of their stay. Since all the boys were tested at least six months before the beginning of this study, the examiners clearly had no knowledge of the purpose of this study at the time of testing.

The DAP and Rorschach tests compare the two groups of boys, first on each of the indicators discussed above (taken separately), and then on the combination of all of those indicators found to differentiate between the two groups. The indicators investigated were

1. The sex of the first person drawn in the DAP;
2. The value of the female figure relative to the male figure as drawn in the DAP;
3. The sex of object of identification in the stories related in the DAP;
4. The relative number of all cross-gender and same-gender responses in the Rorschach;
5. The number of cross-gender and same-gender responses among human content responses in the Rorschach;
6. D7 and W responses to Card IV; and
7. W responses to Card V.

In addition, the responses to Card VII were investigated, though with no prior expectation as to the results.

The comparisons were made with the aim of testing the hypotheses that it is possible to identify boys with a feminine gender identity using specific signs on the DAP and Rorschach individually and that when these indicators are used in combination the false positive rate is considerably reduced but the false negative rate is barely changed. In contrast to many previous studies, comparisons were made on the basis of the numbers of individuals that tested positive in each group rather than on the basis of group averages. As stated earlier, this approach makes it possible to assess the usefulness of these indicators in discriminating on an individual basis.

Several specific hypotheses can be derived from the foregoing discussion. They are as follows:

1. More boys with feminine gender identity than those in the control group should draw a female figure first in the DAP test.
2. More boys with feminine gender identity than those in the control group should draw the female figure with a greater value than the male figure in the DAP test.
3. Fewer boys with feminine gender identity than those in the control group should identify with the male figure in their stories.
4. More boys with feminine gender identity than those in the control group should give a greater total number of cross-gender than same gender responses in the Rorschach test.
5. More boys with feminine gender identity than those in the control group should give a low number of male responses or a high number of female responses among their human content responses in the Rorschach test.
6. Response patterns to Card IV should be significantly different in the two groups, with more boys with feminine gender identity giving responses relating to the figure as big, dark, and menacing or as females than boys in the control group.

7. More boys with feminine gender identity than those in the control group should give a butterfly response first to Card V of the Rorschach test.
8. By use of the combination of the two tests, the DAP and the Rorschach, that is by comparing the number of boys showing all of the above indicators found to be valid, the number of false positives should be reduced compared to use of any of these indicators individually.

Scoring and Reliability

To check for reliability of the scoring method the tests of five children from each group were scored by two trained persons familiar with these psychological tests but with no prior knowledge of the group to which the subject belonged. An overall interscorer agreement of 96% was obtained across both groups and across all measures except those of the DAP stories. With regard to the latter, all the children's tests were scored by an additional trained person because the scoring criteria used had been developed in this study and had not previously been tested. A reliability of 87% across both groups was obtained (27/31). In the four cases of disagreement, the stories were discussed and a joint decision reached.

Statistical Analysis

Each of the above hypotheses were tested using one-tailed Chi squared tests to examine the contingency between the subjects' test performance and the group to which they belong. Clearly, since the hypotheses do not specify the proportions of subjects expected to respond in the described manners, the expected frequencies were calculated from the marginal totals obtained from the data. In addition, the Chi squared test was used to examine the validity of the combined use of only the new indicators shown in this study to be valid.

RESULTS

Tables 1–10 show the results obtained in this study. The results of each indicator used individually are presented first, followed by the results obtained when the valid indicators were used in combination.

The tests of all subjects were analyzed for all hypotheses except hypothesis 3 (related to the stories of the DAP). Only 17/19 of the experimental group and 14/21 of the control group had told stories about their drawings in the DAP, so that only these 31 protocols were analyzed to test hypothesis 3.

Table 1 shows the number of subjects who drew male and female figures first in the DAP. In both groups the majority of boys drew the male figure first, though more of the boys with feminine gender identity did so. The differences between the groups were, however, not significant (Yates corrected $\chi^2 = .21, p > .20$).

Table 1. Sex of First Figure Drawn in the DAP as a Function of Group

	Total	Female (%)	Male (%)
Feminine Gender Identity	19	5 (26.3)	14 (73.7)
Controls	21	8 (38.1)	13 (61.9)

Table 2. *Sex of More Valued Figure Drawn in the DAP as a Function of Group*

	Total	Female (%)	Male (%)
Feminine Gender Identity	19	15 (23.8)	4 [0] (76.2)
Controls	21	8 (23.8)	13 (76.2)

Note.—Figures in square brackets are the number of subjects who drew neither figure as more valued.

Table 2 shows the number and percentage of boys in each group that drew the female or the male figure as more valued in the DAP. Boys who drew neither of the figures as more valued were included with those who drew the male figure as more valued since the hypothesis tested related to greater valuation of the female figure. It is clear from the table that the majority of boys with a feminine gender identity drew the female figure as a greater value while only a minority of the control group did so. Chi squared tests revealed that the differences between the groups were significant (Yates corrected $\chi^2 = 10.03, p < .005$).

Table 3 shows the frequencies and percentages of boys in each group that identified with the female and the male figures in the stories the subjects told about their drawings in the DAP. Boys who did not clearly identify in their stories with either figure were included with those that identified with the female figure since the hypothesis tested relates to inability to identify with the male figure. The majority of boys with feminine gender identity did not identify with the male figure, while the majority of the control group did so. Chi squared tests showed that the differences between the groups were significant (Yates corrected $\chi^2 = 2.49, p < .10$).

Table 4 shows the frequencies and percentages of boys who gave more cross-gender than same-gender responses and more or the same number of same-sex than cross-sex responses in the Rorschach. In both groups those who gave more cross-sex than same-sex responses were in the minority but a greater proportion of the boys with feminine gender identity did so.

These differences were significant (Yates corrected $\chi^2 = 2.14, p < .10$).

Table 5 shows the frequencies and percentages of boys in each group who did and did not show a low number of male human figure responses or a high rate of female human figure responses in the Rorschach. The majority of boys with a feminine gender identity did show such a pattern of responses, while a minority of the control group did so. It should be noted that three out of the four subjects in the experimental group who gave high rates of female responses also gave low rates of male responses, while the one such control subject gave an average level of male responses.

Table 3. *Sex of Identification Figure in the Stories Related in the DAP as a Function of Group*

	Total	Sex of Identification Figure	
		Female (%)	Male (%)
Feminine Gender Identity	17	12 [6] (70.6)	5 (29.4)
Controls	14	5 [3] (35.7)	9 (64.3)

Note.—Figures in square brackets are the number of subjects who identified clearly with neither figure.

Table 4. Relationship of Cross-Gender to Same-Gender Responses in the Rorschach as a Function of Group

	Total	CS > SS (%)	SS > SS (%)
Feminine Gender Identity	19	5 (26.3)	14 [2] (73.7)
Controls	21	1 (4.8)	20 [0] (95.2)

Note.—Figures in square brackets are the number of subjects who gave equal numbers of same- and cross-gender responses.

Chi squared tests revealed that these differences were significant (Yates corrected $\chi^2 = 2.77, p < .05$).

Table 6 shows the frequencies and percentages of boys who gave responses to the W or D7 detail in Card IV in the Rorschach that described the figure as dark and menacing or as female and those who gave other responses.

The two former categories are grouped together in the category of “female-like” responses, since these are patterns more typical of girls than boys (Ames et al., 1959). (No child had difficulty giving a response to this card.)

In both groups just under two thirds of the subjects gave answers in this category, only three of whom (two in the experimental and one in the control group) gave female responses to this card. The differences between the groups were not significant (Yates corrected $\chi^2 = .06, p > .20$).

Table 7 shows the number of subjects who gave bat and butterfly (or dove or other small bird) W responses to Card V of the Rorschach. The one subject who gave a different response was included in the “bat” category since the direction of the hypothesis is for feminine identification, which is symbolized by the butterfly. Over three quarters of the boys with feminine gender identity and just over half of the control group gave “butterfly” responses. Chi squared test revealed that these differences were significant (Yates corrected $\chi^2 = 3.33, p < .05$).

Table 8 shows the number of subjects who gave female responses or had difficulty in responding to Card VII. Just under half of each group gave responses of this type, while only two in each group had difficulty in responding (they gave cloud or stain responses). The differences between the groups were not significant (Yates corrected $\chi^2 = .00, p > .02$).

Table 9 shows the number of subjects in each group who tested positive on *all* of the indicators that were found to significantly differentiate between the two groups. From the above it can be seen that these indicators were the sex of the more valued figure in the DAP, the sex of the figure of identification in the stories to the DAP, the rate of male and female

Table 5. Rate of Male and Female Human Figure Responses in the Rorschach as a Function of Group

	Total	Low M or High F (%)	High M and High F (%)
Feminine Gender Identity	19	14 [4] (73.7)	5 (26.3)
Controls	21	9 [1] (42.8)	12 (57.2)

Note.—Figures in square brackets are the number of subjects who gave a high rate of female responses.

Table 6. *Rate of Female-Like and Other Responses to Card IV of the Rorschach as a Function of Group*

	Total	Dark, Menacing or Female (%)	Other (%)
Feminine Gender Identity	19	12 [2] (63.2)	7 (36.8)
Controls	21	13 [1] (61.9)	8 (38.1)

Note.—Figures in square brackets are the number of subjects who gave a female response.

human responses in the Rorschach, and the response to Card V in the Rorschach. Although the ratio of cross-sex to same-sex responses did differentiate significantly between the two groups, this indicator was not included in this test because it did not successfully identify a majority of the boys with feminine sexual identity. (For boys who did not tell stories in the DAP the criterion used was that they tested positive on all three of the other valid indicators. Nearly half of the boys with feminine gender identity tested positive on all these indicators, while only one of the control group did so. Chi squared tests revealed that the differences between the groups were significant (Yates corrected $\chi^2 = 7.32, p < .005$).

Table 10 shows the number of boys in each group who tested positive on at least two of the three indicators that were not previously used in studies of this kind. These were the sex identification figure in the DAP stories; the response to Card V in the Rorschach; and the rate of male and female human responses in the Rorschach. For subjects who did not tell stories in the DAP, the criterion used was testing positive on both the other indicators. Over three quarters of the boys with a feminine gender identity tested positive on at least two of these indicators, while less than a third of the control group did so. These differences were significant (Yates corrected $\chi^2 = 9.52, p < .005$).

DISCUSSION

Several of the diagnostic indicators examined in this study have been shown to be valid indicators of feminine gender identity in latency aged boys. That is, it was possible to use each of these indicators to differentiate significantly between the two groups and, furthermore to identify a majority of the boys with feminine gender identity. These indicators were drawing the female figure with a greater value than the male in the DAP; difficulty in identification with the male figure in the stories to the DAP; a low relative rate of male human responses and/or high relative rate of female human responses in the Rorschach; initial butterfly response to Card V in the Rorschach.

Table 7. *Rate of "Butterfly" and "Bat" Responses to Card V of the Rorschach as a Function of Group*

	Total	Butterfly (%)	Bat (%)
Feminine Gender Identity	19	16 (84.2)	3 [1] (15.8)
Controls	21	11 (52.4)	10 [0] (47.6)

Note.—Figures in square brackets are the number of subjects who gave a female response.

Table 8. Rate of Female or No-form Responses and Other Responses to Card VII of the Rorschach as a Function of Group

	Total	Female/ No-form (%)	Other (%)
Feminine Gender Identity	19	8 [2] (42.1)	11 (57.9)
Controls	21	10 [2] (47.6)	11 (52.4)

Note.—Figures in square brackets are the number of subjects who gave no-form response.

The ratio of cross-gender to same-gender responses in the Rorschach also differentiated between the groups, although it only identified a minority of the feminine-identified boys. This last indicator, therefore, though valid, is of less potential use.

Interestingly, two indicators that have traditionally been connected with gender identity were found in this study to be invalid diagnostic indicators, at least for boys of latency age. These were the sex of the first figure drawn in the DAP and the responses to Card IV in the Rorschach. (Responses to Card VII also did not differentiate between the two groups.) This suggests that clinicians should avoid making judgments about gender identity on the basis of these indicators, despite traditionally accepted views.

In general, the rates of successful identification obtained in this study, using the indicators found here to be valid, are a considerable improvement on rates previously obtained for the other indicators such as sex of first figure drawn in the DAP, relative height of figures, relative number of articles of clothing drawn, and relative number of cross-gender to same-gender responses in the Rorschach. Previous studies have shown that these indicators successfully identify at best only about half of all feminine-identified boys tested, while simultaneously giving rise to high false positive rates. Some studies only showed the existence of differences in group averages in performance on certain indicators: the diagnostic value of these indicators on the individual level is therefore dubious. In the present study, the boys were tested on an individual basis and well over half of the feminine-identified boys tested positive on the indicators that were found to be valid. Therefore, the valid indicators used here seem to be of greater potential diagnostic value. Furthermore, for most of these, the false positive rates were lower than those previously obtained for other indicators.

From a comparison of the false positive and negative rates, and the significance levels obtained for these indicators, it is clear that some of the indicators are better than others at characterizing feminine-identified boys. The most valid and reliable indicator according to this measure was that of the relative value of the female figure drawn in the DAP, followed by the responses to Card V of the Rorschach and the rates of male and female human responses in the Rorschach, and lastly the sex of the identification figure in the DAP stories. Moreover, further

Table 9. Frequency of Boys Testing Positive on All and Only on Some of the Valid Indicators as a Function of Group

	Total	Positive on All (%)	Positive (%)
Feminine Gender Identity	19	9 (47.4)	10 (52.6)
Controls	21	1 (4.8)	20 (95.2)

Table 10. *Frequency of Boys Testing Positive on at Least Two of the Valid Indicators New to This Study as a Function of Group*

	Total	Positive on All (%)	Positive on Some (%)
Feminine Gender Identity	19	16 (84.2)	3 (15.8)
Controls	21	6 (28.6)	15 (71.4)

investigation showed that assessment of performance on these three valid indicators in combination revealed that the false positive and negative rates were considerably lower than the rates associated with each individual indicator. Over three quarters of the feminine-identified boys tested positive on at least two of these indicators while only just over a quarter of the controls did so. These results are comparable to those obtained for the sex of the more valued figure drawn in the DAP, itself a multidimensional indicator. Therefore, not only are these three indicators apparently individually valid as diagnostic indicators of feminine gender identity in boys, but when used in combination they give a low false negative and positive rates as the DAP indicator. The implication is that they should also be used, especially in combination, as diagnostic indicators of feminine gender identity in late latency boys in future studies and clinical practice.

The additional advantage of these new indicators is twofold. First, they are not based on the subjects' drawings but on their verbal responses to the Rorschach and the narration of stories to their drawings. Thus they make it possible to diagnose subjects who cannot or do not express their inner world easily via drawing. Second, two of these three indicators assess difficulty in identification with the male figure and not only preference for the female figure. These indicators therefore can be used to diagnose "feminine" boys who have particular difficulties with masculine, rather than feminine, identification.

The above findings are also important for the support they lend to the hypothesis that diagnostic indicators should preferably be used in combination. This hypothesis was further supported by the number of subjects who gave positive responses on all the valid indicators found in this study. Only one of the control group (less than 5%) did so, while almost half of the feminine-identified group of boys did. This level of false positive rates (less than 5%) is much lower than the rates obtained for the individual indicators (between 23% and 52%), which is significant if these indicators are to be used in clinical situations. Although the rate of successful identification was reduced by a third (from about 75% to 50%) by using the combination of indicators, the false positive rate was reduced to a much greater extent (by over three quarters), which makes this approach of great potential diagnostic use. The ratio of true positive to false positive rates (47% to 4.8%—about 10:1) is much better than in all the studies discussed earlier (with the exception of Abraham's [1976, 1979] study, which used a nonclinical control group and problematic diagnostic criteria). (For instance, in Zucker et al.'s [1983] and Green et al.'s [1972] studies of the rate of drawing the female figure first, which were the only ones of these studies where the true positive rate reached 50% for boys in this age group, the false positive rates were about 15% for "normal" controls and about 25% for "clinical" controls. This gives a true positive to false positive rate of only 2:1.) Even the false positive and false negative rates obtained in the present study for subjects who tested positive on some but not all of the diagnostic indicators (as in the sex of the more valued figure in the DAP, or the combined use of the other three valid indicators just discussed) compare very favorably to previous studies. In these cases, while the false positive rates were similar to those in previous studies (around 25%), the

true positive rates are much greater (around 75%, giving a true positive to false positive ratio of 3:1 as compared to 2:1).

The assessment of performance with a combination of indicators can be done in different ways. On the one hand, if one wants to maximize the rate of characterizing subjects with feminine gender identity (while keeping the false positive rate relatively low), then the criterion that should be used is that of testing positive on most of the indicators. It seems that in this case the use of a combination of indicators takes into account the fact that individuals express gender identity in different ways and via different media. It is therefore successful in characterizing more subjects. On the other hand, if one wants to minimize the number of false positives, then the criterion should be testing positive on all the indicators. This makes it possible to eliminate subjects who do not have a feminine gender identity, but nevertheless (for various reasons) respond similarly on some (but not most) of the indicators. Whichever of these approaches is used, the ratio of true positives to false positives is much better than when indicators are used individually.

Furthermore, this approach might make it possible to assess the relative severity of gender disturbance in individuals by examining the relative number of indicators on which each individual tests positive. It is likely that if subjects test positive on all of a variety of measures, their disturbance is severe and distorts many of their modes of responding. Similarly, testing positive on fewer indicators probably indicates a less extensive disturbance. Clearly, this thesis should be tested in future research.

Thus, it appears that the use of a combination of diagnostic indicators has great potential importance for the assessment of gender identity and indeed any other psychological disturbance. On a clinical level, this approach is often intuitively taken; diagnoses are made on the basis of accumulated evidence from several psychological tests. However, in experimental work, researchers often search for single diagnostic indicators—an approach that has almost certainly had a negative effect on clinical diagnosis using projective tests. The findings of the present study are therefore of special relevance to the methodology to be used in future research and clinical diagnosis as well as the nature of the particular diagnostic indicators used.

Unlike many previous studies, the comparisons made here were with a control group of boys with similar, very severe, emotional problems; the differences between the groups are therefore clearly due to differences in gender identity rather than in general psychological health. Conversely, it could be argued that because the study was carried out using only severely emotionally disturbed boys as subjects, the indicators are only valid for this group. However, previous studies, such as those of Zucker et al. (1983, 1992), have shown that differences between children with gender problems and clinical controls are less marked than when the comparisons are made with “normal” children. Therefore, it is likely that even lower false positive rates than those obtained in this study would be found if the children with gender problems were compared with a non-clinical control group using the same indicators. This implies that these indicators are probably valid in non-clinical individuals and may have even greater diagnostic accuracy than in a clinical group such as that from which the subjects of this study were chosen. Nevertheless, it is clearly necessary to investigate the validity of the indicators, and in particular of this last hypothesis, in a group of non-clinical boys.

REFERENCES

- ABRAHAM, A. (1976). *Les identifications de l'enfant a travers son dessin*. Ed. Privat. Paris.
- ABRAHAM, A. (1979). A syndrome of inversed identification in the Machover (D.A.P.) test. *Megamot*, 24(4), 543–563 (in Hebrew).
- ABRAHAM, A. (1989). *The visible and the hidden in human figure drawing*. Tel Aviv: Reshafim (in Hebrew).

- AMES, L.B., LEARNED, J., METRAUX, R.W., & WALKER, N. (1952). *Child Rorschach responses: Developmental trends from two to ten years*. New York: Hoeber Inc.
- AMES, L.B., METRAUX, R.W., & WALKER, N. (1959). *Adolescent Rorschach responses: Developmental trends from ten to sixteen years*. New York: Hoeber Inc.
- ARANOW, E., & REZNIKOFF, M. (1976). *Rorschach content interpretation*. New York: Grune and Stratton, Inc.
- BAKER, A.J., MATTHIS, J.K., & POWERS, C.A. (1953). Drawing characteristics of male homosexuals. *Journal of Clinical Psychology, 9*, 185–188.
- BEM, S.E. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*, 155–162.
- BILLER, H.B. (1974). *Paternal deprivation*. Lexington, MA (DC): Heath.
- BROWN, D.G. (1956). Sex role preference in young children. *Psychological Monographs, 70*, no. 14, Whole No. 421.
- BROWN, D.G., & TOLOR, A. (1957). Human figure drawings as indicators of sexual identification and inversion. *Journal of Perceptual and Motor Skills, 7*, 199–211.
- DAVIDS, A., JOELSON, M., & MCARTHUR, C. (1956). Rorschach and TAT indices of homosexuality in overt homosexuals, neurotics and normal males. *Journal of Abnormal Psychology, 53*, 161–172.
- DE MARTINO, M.F. (1954). Human figure drawings by mentally retarded males. *Journal of Clinical Psychology, 10*, 241–244.
- EXNER, J.E. (1974). *The Rorschach: A comprehensive system*. Vol. I. New York: Wiley.
- GOLDFRIED, M.R. (1966). On the diagnosis of homosexuality. *Journal of Consulting Psychology, 30*, 338–349.
- GRANICK, S., & SMITH, L.J. (1953). Sex sequence in the Draw-A-Person test and its relation to the MMPI masculinity-femininity scale. *Journal of Consulting and Clinical Psychology, 17*, 71–73.
- GREEN, R., FULLER, M., & RUTLEY, B. (1972). It = scale for children and Draw a Person test: 30 feminine vs. 25 masculine boys. *Journal of Personality Assessment, 36*, 349–352.
- HAMMER, E.F. (1954). Relationship between diagnosis of psychosexual pathology and the sex of the first drawn person. *Journal of Clinical Psychology, 10*, 168–170.
- HAYDEN, B.C. (1981). Rorschach Cards IV and VII revisited. *Journal of Personality Assessment, 45*, 226–229.
- HENDLIN, S.J. (1976). Homosexuality in the Rorschach. A new look at old signs. *Journal of Homosexuality, (1)*, 303–312.
- HOOKE, E. (1958). Male homosexuality in the Rorschach. *Journal of Projective Techniques, 22*, 33–54.
- KURTZBERG, R.L., CAVIOR, N., & LIPTON, D.S. (1966). Sex drawn first and sex drawn larger by opiate addict and non-addict inmates on the Draw-A-Person test. *Journal of Projective Techniques and Personality Assessment, 10*, 55–58.
- LEFKOWITZ, M.M. (1962). Some relationships between sex role preference of children and other parent and child variables. *Psychological Reports, 10*, 45–53.
- LEVY, S. (1959). Figure drawing as a projective technique. In L. Abt & L. Bellak (Eds.), *Projective psychology*, (pp. 257–298). New York.
- LIABOE, G.P., & GUY, J.D. (1985). The Rorschach “father” and “mother” cards: An evaluation of the research. *Journal of Personality Assessment, 49(1)*, 1–5.
- MACHOVER, K. (1949). *Personality projection in the drawing of the human figure*. Springfield IL: Thomas.
- MEYER-BAHLBURG, H.F.L., FELDMAN, J.F., & EHRHART, A.A. (1985). Questionnaires for the assessment of atypical gender role behavior. A methodological study. *Journal of American Academy of Child Psychiatry, 24(6)*, 695–701.
- REKERS, G.A. (1981). Psychosexual and gender problems. In E.J. Mash & L.G. Tendor (Eds.), *Behavioral assessment of child disorders* (pp. 483–526). New York: Guilford.

- REKERS, G.A., & MOREY, S.M. (1989). The relationship of maternal report of feminine behaviors and extroversion to clinician's rating of gender disturbance. *Perceptual and Motor Skills*, 69, 387–394.
- SHAFER, R. (1954). Psychoanalytic interpretation in Rorschach testing. New York: Grune and Stratton.
- SKILBECK, W.M., BATES, J.F., & BENTLER, P.M. (1975). Human figure drawings of gender problem and school problem boys. *Journal of Abnormal Child Psychology*, 3(3), 191–199.
- SPENCE, J.T., HELMREICH, R., & STAPP, J. (1979). The personality Attributes Questionnaire: A measure of sex role stereotypes and masculinity–femininity. *JSAS Catalog Sel Doc Psychology*, 4:43.
- TUBER, S., & COATES, S. (1985). Interpersonal phenomena in the Rorschachs of extremely feminine boys. *Psychoanalytic Psychology*, 2, 251–265.
- TYSON, P. (1982). A developmental line of gender identity, gender role and choice of love object. *Journal of the American Psychoanalytic Association*, 30(1): 61–86.
- WHEELER, R.M. (1949). An analysis of Rorschach indices of male homosexuality. *Journal of Projective Techniques*, 13, 97–126.
- WYSOCKI, A.C., & WYSOCKI, B.A. (1977). Human figure drawings of sex offenders. *Journal of Clinical Psychology*, 33(1).
- ZUCKER, K.J. (1985). Cross-gender identified children. In B.W. Steiner (Ed.), *Gender dysphoria: Development, research management*, (pp. 75–174). New York: Plenum.
- ZUCKER, K.J., FINEGAN, J.K., DOERING, R.W., & BRADLEY, S.J. (1983). Human figure drawings of gender problem children: A comparison to psychiatric and normal controls. *Journal of Abnormal and Child Psychology*, 11(2), 287–298.
- ZUCKER, K.J., LOZINSKI, J.A., BRADLEY, S.J., & DOERING, R.W. (1992). Sex typed responses in the Rorschach protocols of children with gender identity disorder. *Journal of Personality Assessment*, 58(2), 295–310.