Archives of Sexual Behavior, Vol. 33, No. 2, April 2004, pp. 105–116 (© 2004)

A Parent-Report Gender Identity Questionnaire for Children

Laurel L. Johnson, M.A.,[1] Susan J. Bradley, M.D.,[1] Andrea S. Birkenfeld-Adams, Ph.D.,[1,2] Myra A. Radzins Kuksis, Ph.D.,[1,2] Dianne M. Maing, Ph.D.,[1,3] Janet N. Mitchell, Ph.D.,[1,4] and Kenneth J. Zucker, Ph.D.[1,5]

Received June 2, 2003; revision received July 25, 2003; accepted July 25, 2003

This paper reports on the psychometric properties of a 16-item parent-report Gender Identity Questionnaire, originally developed by P. H. Elizabeth and R. Green (1984), to aid in the assessment of children with potential problems in their gender identity development. The questionnaire, which covered aspects of the core phenomenology of gender identity disorder (GID), was completed by parents of gender-referred children ($N = 325$) and controls (siblings, clinic-referred, and nonreferred; $N = 504$), who ranged in age from 2.5–12 years (mean age, 7.6 years). Factor-analysis indicated that a one-factor solution, containing 14 of the 16 items with factor loadings $\geq .30$, best fit the data, accounting for 43.7% of the variance. The gender-referred children had a significantly more deviant total score than did the controls, with a large effect size of 3.70. The GIQ total score had negligible age effects, indicating that the questionnaire has utility for assessing change over time. The gender-referred children who met the complete *DSM* criteria for GID had a significantly more deviant total score than did the children who were subthreshold for GID, although the latter group had a mean score that was closer to the threshold cases than to the controls. With a specificity rate set at 95% for the controls, the sensitivity rate for the probands was 86.8%. It is concluded that this parent-report gender identity questionnaire has excellent psychometric properties and can serve as a useful screening device for front-line clinicians, for whom more extensive, expensive, and time-consuming assessment procedures may be precluded.

KEY WORDS: gender identity disorder; gender identity; gender role; children; assessment.

INTRODUCTION

Both prior to and after the introduction of the gender identity disorder (GID) diagnosis for children in the *DSM* (American Psychiatric Association, 1980), a variety of measurement approaches have been developed to assess sex-typed behavior in children referred clinically for potential problems in their gender identity development. Direct assessment of the child has included observation of sex-typed behavior in free-play tasks, either alone (Doering, Zucker, Bradley, & MacIntyre, 1989; Green & Fuller, 1973; Green, Fuller, Rutley, & Hendler, 1972; Rekers & Yates, 1976; Zucker, Doering, Bradley, & Finegan, 1982) or in the presence of peers (Fridell, 2001) or a parent (Doering, 1981; Rekers, 1975), the Playmate and Play Style Preferences Structured Interview (Fridell, 2001; see also Alexander & Hines, 1994), projective measures, such as the IT Scale for Children (Green, Fuller, & Rutley, 1972), the Draw-a-Person task (Green et al., 1972; Skilbeck, Bates, & Bentler, 1975; Zucker, Finegan, Doering, & Bradley, 1983), and the Rorschach (Tuber & Coates, 1985; Zucker, Lozinski, Bradley, & Doering, 1992), and a factor-analytically derived structured Gender Identity Interview (Zucker et al., 1993).

Assessment via parent-report has been carried out using standardized questionnaires, including the Games

[1]Child and Adolescent Gender Identity Clinic, Child, Youth, and Family Program, Centre for Addiction and Mental Health—Clarke Division, Toronto, Ontario, Canada.
[2]Toronto District School Board, Toronto, Ontario, Canada.
[3]Kinark Child and Family Services, Barrie, Ontario, Canada.
[4]Deceased.
[5]To whom correspondence should be addressed at Child and Adolescent Gender Identity Clinic, Child, Youth, and Family Program, Centre for Addiction and Mental Health—Clarke Division, Toronto, Ontario, Canada M5T 1R8; e-mail: ken_zucker@camh.net.

0004-0002/04/0400-0105/0 © 2004 Plenum Publishing Corporation

Inventory (Bates & Bentler, 1973; Doering, 1981; Klein & Bates, 1980; see also Meyer-Bahlburg, Sandberg, Dolezal, & Yager, 1994; Sandberg & Meyer-Bahlburg, 1994) and the Child Behavior and Attitude Questionnaire (CBAQ; Bates, Bentler, & Thompson, 1973, 1979), as well as responses to an array of interviewer-based questions pertaining to various sex-typed behaviors (Green, 1976, 1987; Zuger & Taylor, 1969).

Although it is generally agreed that none of these measurement approaches, on its own, should be used to replace a clinical interview, usually with parents, to establish the GID diagnosis, they have all shown evidence for discriminant validity when comparing gender-referred children to controls, who have included siblings, clinic-referred, and nonreferred children (for reviews, see Zucker, 1992; Zucker & Bradley, 1995). From a clinical utility standpoint, however, these measurement approaches have several limitations. For example, some of the child measures require standardized stimulus materials not readily available in the typical clinical setting (e.g., in the private practice office) and others are relatively time-consuming (e.g., the administration of the Rorschach and utilization of a complex sex-typed responses coding manual).

With regard to the parent-report measures, such as the CBAQ, one limitation is that it was developed only for boys ages 5–12 years. Subsequently, Meyer-Bahlburg, Sandberg, Yager, Dolezal, and Ehrhardt (1994) developed a revised version with a standardization sample for both boys and girls although the age range was limited to 6–10-year-olds and comparative data on a clinic sample of gender-referred children were not available (see also Meyer-Bahlburg, Feldman, & Ehrhardt, 1985; Sandberg, Meyer-Bahlburg, Ehrhardt, & Yager, 1993). In both the Bates et al. (1973) and Meyer-Bahlburg et al. (1994) reports, another limitation was the absence of information provided by fathers, as mothers were the primary informants. A clear strength of the revised CBAQ was the demonstration of strong normative sex differences, thus permitting the comparison of a clinic-referred sample of children with gender identity problems to scores of both boys and girls.

This study examined the psychometric properties of another parent-report "behavioral preference" questionnaire, originally developed by Elizabeth and Green (1984) using a sample of 702 MZ and DZ twins who ranged in age from 4–12 years. It contained 17 items: 15 items covered a range of behaviors consistent with aspects of the phenomenology of the GID diagnosis, one item pertained to "romantic interests" in the opposite sex, and the final item was a global rating of bipolar masculinity–femininity. Elizabeth, Meyer-Bahlburg, Janal, Ehrhardt, and Green (1984) showed that, with one exception, all of the items showed strong sex differences. Elizabeth et al. factor-analyzed the questionnaire and found that the majority of items loaded on one factor with factor loadings $\geq .30$, which accounted for most of the common variance.

In this study, we made some minor modifications to the questionnaire, which we have named the Gender Identity Questionnaire, which was then completed by parents of both gender-referred and control children. We examined its psychometric properties in six ways: (1) it was subjected to factor analysis; (2) calculated the mother–father correlation for the total score; (3) compared the total score for the gender-referred probands versus the controls; (4) examined the influence of several demographic variables (sex, age, IQ, and parent's social class and marital status); (5) compared the total scores of the gender-referred children who either met the complete *DSM* criteria for GID or were subthreshold for the diagnosis; and (6) report data on the total score's sensitivity and specificity.

METHOD

Participants

The probands were 325 children (age range, 3–12 years) referred consecutively to, and then assessed in, the Child and Adolescent Gender Identity Clinic, which is housed within the Child Psychiatry Program at the Centre for Addiction and Mental Health—Clarke Division, between May 1985–April 2003. On the basis of clinical interview information derived predominantly from parent interview, 216 (66.5%) of the probands met complete *DSM* criteria for GID and 109 (33.5%) were subthreshold. As noted elsewhere (Zucker & Bradley, 1995), virtually all of the subthreshold cases showed at least some characteristics of the criteria for GID and many would have met the complete criteria when they were younger. In this respect, they were not "false positive" referrals, at least in the extreme sense of the term.

The comparison group consisted of 504 children (age range, 2.5–12 years), containing four subgroups: siblings of the gender-referred probands ($N = 272$), clinical controls ($N = 118$), nonreferred controls ($N = 86$), and miscellaneous cases ($N = 28$; e.g., siblings of adolescents referred for GID or transvestic fetishism). The clinical controls contained some children who were part of a specific study of GID and were pair-matched to a proband on several demographic variables (e.g., sex, age, IQ, and parent's social class and marital status; Birkenfeld-Adams, 1999; Mitchell, 1991), were seen as part of other research studies (Kuksis, 1992; Maing, 1991), or who had been

Table I. Demographic Characteristics of Gender-Referred Children and Controls

Variable		Group: Gender-referred	Group: Controls	t or ^{2}	p
Age (in years)	*M*	7.13	7.85	3.89	<.001
	SD	2.49	2.70		
	Range	3.33–12.99	2.53–12.99		
	N	325	504		
Sex					
Boys	*N* (%)	275 (84.6)	305 (60.5)	53.45	<.001
Girls	*N* (%)	50 (15.4)	199 (39.5)		
Full-scale IQ[a]	*M*	103.88	104.41	<1	*ns*
	SD	16.94	14.97		
	Range	48–143	68–140		
	N	318	177		
Parent's marital status					
Both Parents	*N* (%)	219 (67.4)	296 (58.7)	5.03	.014
Other[b]	*N* (%)	106 (32.6)	208 (41.3)		
Social Class[c]	*M*	42.99	40.75	2.10	.036
	SD	14.92	15.06		
	Range	8–66	8–66		
	N	325	504		

[a]For the gender-referred group, full-scale IQ data were not available for two patients who did not speak English, so Performance IQ was used instead. For the remaining five patients, an IQ score was not available. For the controls, IQ data were systematically collected for only particular subgroups (e.g., as part of matching procedures for other studies); hence, the number of participants with IQ scores was considerably lower than the total sample size.

[b]For marital status, the category "Other" includes the following family constellations: single parent, separated, divorced, widowed, reconstituted (e.g., mother and step-father), living in a group home, etc.

[c]Hollingshead's Four-Factor Index of Social Status (Hollingshead, 1975), absolute range, 8–66.

referred to our clinic for reasons other than gender identity concerns (e.g., sexualized behavior). The clinical controls were heterogeneous with regard to psychiatric diagnoses. The nonreferred controls contained children who were also part of a specific study of GID and, like the clinical controls, were pair-matched to a GID proband on several demographic variables or who were seen as part of other research studies (op. cit.).

The demographic characteristics of the two groups are shown in Table I. It can be seen that the gender-referred children were significantly younger than the controls. Their social class background was significantly higher (although the difference was quite marginal in terms of the mean difference), and significantly more were living with both of their parents. There was no between-group difference in IQ. Thus, for some of the parametric analyses comparing the probands and controls, age, social class, and parent's marital status were covaried.

Measure

At the time of assessment, parents were asked to complete a number of questionnaires, including our revised version of the Elizabeth and Green (1984) Gender Identity Questionnaire. It contained 16 items that covered a range of sex-typed behaviors that corresponded to various features of the core phenomenology of the GID diagnosis. Each item was rated on a 5-point scale for frequency of occurrence (three items also contained a "not applicable" option), with lower scores reflecting more cross-gendered behavior (or less same-gendered behavior). The exact wording of each item is shown in the Appendix. In the current version, Item 13 was written to apply to both boys and girls equally (combining Items 14–15 in the original). Items 14–16 were added to the current version, and Items 16 (about romantic interests) and 17 (the bipolar masculinity–femininity rating) from the original version were omitted.

RESULTS

Preliminary analyses showed that the data for the different subgroups of controls were comparable, so their data were combined. Table II shows descriptive data for each of the 16 items as a function of group. For each item,

Table II. Frequency Ratings for each Item on the Gender Identity Questionnaire as a Function of Sex and Group

	Group	
	Gender identity disorder ($N = 325$)	Controls ($N = 504$)
Item 1: Favorite playmates	Boys ($N = 270$) Girls ($N = 50$)	Boys ($N = 298$) Girls ($N = 196$)
Same-sex		
Boys	10.0%	69.1%
Girls	10.0%	66.3%
Equal		
Boys	33.0%	27.9%
Girls	34.0%	31.1%
Cross-sex		
Boys	57.1%	3.0%
Girls	56.0%	2.5%
Item 2: Plays with girl-type dolls	Boys ($N = 271$) Girls ($N = 50$)	Boys ($N = 296$) Girls ($N = 197$)
Frequently/favorite toy		
Boys	52.0%	0.7%
Girls	6.0%	55.9%
Once-in-a-while		
Boys	26.6%	10.8%
Girls	20.0%	21.3%
Very rarely/never		
Boys	21.4%	88.5%
Girls	74.0%	22.8%
Item 3: Plays with boy-type dolls	Boys ($N = 271$) Girls ($N = 49$)	Boys ($N = 296$) Girls ($N = 197$)
Frequently/favorite toy		
Boys	15.5%	37.2%
Girls	26.5%	5.6%
Once-in-a-while		
Boys	30.6%	23.0%
Girls	28.6%	17.3%
Very rarely/never		
Boys	53.9%	39.9%
Girls	44.9%	77.2%
Item 4: Experiments with cosmetics and jewelry	Boys ($N = 272$) Girls ($N = 50$)	Boys ($N = 301$) Girls ($N = 197$)
Frequently/favorite activity		
Boys	32.4%	1.0%
Girls	4.0%	57.9%
Once-in-a-while		
Boys	32.7%	7.0%
Girls	22.0%	34.5%
Very rarely/never		
Boys	34.9%	92.0%
Girls	74.0%	7.6%
Item 5: Imitates female characters on TV/movies	Boys ($N = 272$) Girls ($N = 50$)	Boys ($N = 301$) Girls ($N = 197$)
Frequently/favorite activity		
Boys	53.7%	1.3%
Girls	2.0%	26.9%
Once-in-a-while		
Boys	22.4%	5.3%
Girls	8.0%	32.5%
Very rarely/never		
Boys	23.9%	93.4%
Girls	90.0%	40.6%
Item 6: Imitates male characters on TV/movies	Boys ($N = 272$) Girls ($N = 50$)	Boys ($N = 301$) Girls ($N = 197$)
Frequently/favorite activity		
Boys	8.1%	36.2%
Girls	30.0%	0.5%
Once-in-a-while		
Boys	32.4%	34.6%
Girls	32.0%	12.2%
Very rarely/never		
Boys	59.6%	29.2%
Girls	38.0%	87.3%
Item 7: Plays sports with boys (but not girls)	Boys ($N = 272$) Girls ($N = 50$)	Boys ($N = 301$) Girls ($N = 197$)
Frequently/favorite activity		
Boys	18.0%	65.1%
Girls	52.0%	7.1%
Once-in-a-while		
Boys	32.7%	20.6%
Girls	28.0%	21.8%
Very rarely/never		
Boys	49.3%	14.3%
Girls	20.0%	71.0%
Item 8: Plays sports with girls (but not boys)	Boys ($N = 272$) Girls ($N = 50$)	Boys ($N = 293$) Girls ($N = 197$)
Frequently/favorite activity		
Boys	18.7%	5.1%
Girls	18.0%	39.1%
Once-in-a-while		
Boys	29.8%	25.9%
Girls	32.0%	36.5%
Very rarely/never		
Boys	51.5%	68.9%
Girls	50.0%	24.4%
Item 9: Roles in fantasy play	Boys ($N = 247$) Girls ($N = 38$)	Boys ($N = 203$) Girls ($N = 179$)
Same-sex		
Boys	17.4%	92.6%
Girls	15.8%	95.5%
Equal		
Boys	20.6%	7.4%
Girls	15.8%	3.4%
Cross-Sex		
Boys	61.9%	0.0%
Girls	68.4%	1.1%
Item 10: Plays "girl-type" games	Boys ($N = 272$) Girls ($N = 50$)	Boys ($N = 297$) Girls ($N = 197$)
Frequently/favorite activity		
Boys	59.2%	3.3%
Girls	12.0%	76.7%
Once-in-a-while		
Boys	24.6%	15.5%
Girls	24.0%	17.8%
Very rarely/never		
Boys	16.2%	81.2%
Girls	64.0%	5.6%

Table II. (*Continued*)

	Group	
	Gender identity disorder ($N = 325$)	Controls ($N = 504$)
Item 11: Plays "boy-type" games	Boys ($N = 272$) Girls ($N = 50$)	Boys ($N = 298$) Girls ($N = 197$)
Frequently/favorite activity		
Boys	31.2%	84.6%
Girls	78.0%	4.6%
Once-in-a-while		
Boys	39.0%	10.1%
Girls	14.0%	29.4%
Very rarely/never		
Boys	29.7%	5.3%
Girls	8.0%	66.0%
Item 12: Dress-up play	Boys ($N = 240$) Girls ($N = 29$)	Boys ($N = 167$) Girls ($N = 171$)
Same-sex		
Boys	8.3%	86.8%
Girls	10.3%	92.4%
Equally		
Boys	15.8%	10.8%
Girls	6.9%	5.3%
Cross-sex		
Boys	75.9%	2.4%
Girls	82.7%	2.3%
Item 13: States wish to be opposite sex	Boys ($N = 272$) Girls ($N = 50$)	Boys ($N = 301$) Girls ($N = 197$)
Frequently/every day		
Boys	22.1%	0.0%
Girls	44.0%	0.0%
Once-in-a-while		
Boys	36.4%	1.7%
Girls	26.0%	2.0%
Very rarely		
Boys	18.4%	3.3%
Girls	14.0%	5.6%
Never		
Boys	23.2%	95.0%
Girls	16.0%	92.4%
Item 14: States is the opposite sex	Boys ($N = 272$) Girls ($N = 50$)	Boys ($N = 301$) Girls ($N = 197$)
Frequently/every day		
Boys	14.7%	0.0%
Girls	26.0%	0.0%
Once-in-a-while		
Boys	17.3%	0.3%
Girls	20.0%	1.0%
Very rarely		
Boys	16.5%	2.0%
Girls	12.0%	0.5%
Never		
Boys	51.5%	97.7%
Girls	42.0%	98.5%
Item 15: Talks about disliking sexual anatomy	Boys ($N = 272$) Girls ($N = 50$)	Boys ($N = 301$) Girls ($N = 197$)
Frequently/every day		
Boys	3.0%	0.0%
Girls	8.0%	0.5%
Once-in-a-while		
Boys	6.6%	0.0%
Girls	18.0%	1.5%
Very rarely		
Boys	13.6%	1.7%
Girls	10.0%	2.0%
Never		
Boys	76.8%	98.3%
Girls	64.0%	95.9%
Item 16: Talks about liking sexual anatomy	Boys ($N = 272$) Girls ($N = 50$)	Boys ($N = 282$) Girls ($N = 195$)
Frequently/every day		
Boys	2.2%	3.9%
Girls	0.0%	1.0%
Once-in-a-while		
Boys	9.9%	6.4%
Girls	0.0%	6.7%
Very rarely		
Boys	7.4%	6.7%
Girls	6.0%	5.6%
Never		
Boys	80.5%	83.0%
Girls	94.0%	86.7%

Note. For Items 1, 9, and 12, the same-sex and cross-sex categories combined the response options of *always* and *usually*. For Items 2–8 and 10–11, the response options of *favorite* or *frequently* and *very rarely* or *never* were combined. For Items 13–14, the response options of *every day* and *frequently* were combined. For Items 1, 9, and 12, there was a *Not applicable* option (e.g., "does not play with other children," "does not play these games"). Variation in N across items reflects missing data and/or endorsement of the *Not applicable* option.

it can be seen that the probands had much higher ratings of cross-gender behavior (or lower ratings of same-gender behavior) than did the controls.

Factor Analysis

A principal-axis factor analysis with varimax rotation was performed on the 16 questionnaire items. Several solutions were explored and the one that best fit the data was a one-factor solution, containing 14 of the 16 items with factor loadings $\geq .30$, accounting for 43.7% of the variance (Table III). The two items that did not contain loadings $\geq .30$ were Item 8, "He (she) plays sports with girls (but not boys)," and Item 16, "He (she) talks about liking his (her) sexual anatomy (private parts)." As shown in Table III, the factor loadings obtained in the current sample were very similar to the factor loadings reported by Elizabeth et al. (1984).

Table III. Factor Loadings on the Gender Identity Questionnaire

Item	Factor loading (current study)	Factor loading (Elizabeth et al., 1984)
1	0.77	0.80
2	0.74	0.77
3	0.34	0.30
4	0.71	0.77
5	0.64	0.69
6	0.48	0.59
7	0.62	0.67
8	0.20	0.38
9	0.89	0.92
10	0.83	0.88
11	0.72	0.85
12	0.91	0.94
13	0.81	—
14	0.69	—
15	0.47	—
16	0.02	—

Note. See the Appendix for exact wording of each item.

Proband-Control Comparisons

On the basis of the factor analysis, a mean GIQ total score was calculated by summing the 14 items that loaded $\geq$.30 on the factor and then dividing by 14 (if a rating for an item was missing or the parent endorsed option "f" for any or all of Items 1, 9, or 12, the denominator was modified accordingly). Table IV shows the means and SDs for the GIQ total score as a function of sex and group. A 2 (group) $\times$ 2 (sex) analysis of covariance (ANCOVA) revealed main effects for Group, $F(1, 822) = 1167.9$, $p < .001$, and Sex, $F(1, 882) = 23.4$, $p < .001$, and a Group $\times$ Sex interaction that approached significance, $F(1, 822) = 3.53$, $p = .061$.

The gender-referred group had a significantly higher deviant total score than the controls. Using Cohen's d ($M_1 - M_2/SD_{\text{cont}}$), the effect size (*ES*) for the proband-control comparison was 3.70. The boys had a significantly higher deviant total score than the girls, but this was artifactual because of the disproportionate number of boys with GID. The marginal Group $\times$ Sex interaction showed that both the boys and girls with GID had a significantly higher deviant total score than their same-sex counterparts in the control group (both $ps < .001$); in addition, the girls with GID had a significantly higher deviant total score than the boys with GID ($p < .001$) and the control boys had a significantly lower deviant total score than the control girls ($p = .003$). Across both groups, the mother–father correlation for the GIQ total score was substantial ($r = .90$, $p < .001$).

Table IV. Maternal Ratings on the Gender Identity Questionnaire as a Function of Group (Total Score)

Group	*M*	*SD*	*N*
Gender-referred			
Boys	2.87	0.62	275
Girls	2.62	0.56	50
Combined	2.83	0.62	325
Controls			
Boys	4.25	0.35	305
Girls	4.13	0.36	199
Combined	4.20	0.36	504

Note. Absolute range, 1–5. A "1" is the most deviant or sex-atypical score and a "5" is the most "normal" or sex-typical score. For three gender-referred probands and six controls, only ratings by the father were available. For these nine participants, the father's score was used to replace the missing maternal data.

Demographic Correlates

Across both groups, the correlation between the GIQ total score for maternal ratings was not significantly related to IQ ($r = .02$), and quite weakly related to parents' social class ($r = -.06$, $p = .062$) and marital status ($r = -.07$, $p = .036$). Age was, however, more strongly related to the GIQ total score ($r = .18$, $p < .001$), indicating less deviant scores with increasing age. When the groups were analyzed separately, only age was significantly related to the GIQ total score (probands: $r = .15$, $p < .01$; controls: $r = .10$, $p < .05$).

To explore the relation between age and the GIQ total score in a clinically more meaningful manner, we created 10 age blocks (i.e., 3.00–3.99 years, 4.00–4.99 years, etc.). Table V shows the means and SDs as a function of group and age block. A 2 (group) $\times$ 10 (age block) ANCOVA (with social class and marital status covaried) revealed a significant main effect for Group, $F(1, 807) = 1296.1$, $p < .001$, and a marginal main effect for Age Block, $F(9, 807) = 1.87$, $p = .054$.

Duncan's multiple range tests showed that the 3–4-, 5–6-, and 6–7-year-olds all had a significantly higher deviant total score than did the 10–11- and 11–12-year-olds (all $ps < .05$) and the 4–5-year-olds had a significantly higher deviant total score than all age groups 7 years and older (all $ps < .05$). It should, however, be noted that the largest mean difference between two age blocks (.608) was substantially smaller than the mean differences between the probands and controls across all age blocks (see Table V). Effect sizes for the group differences were calculated for each age block and ranged from 3.03–4.77.

Table V. Maternal Ratings on the Gender Identity Questionnaire as a Function of Group and Age Block

Group	*M*	*SD*	*N*
Gender-referred			
3–4 years	2.73	0.73	19
4–5 years	2.72	0.56	59
5–6 years	2.81	0.55	53
6–7 years	2.80	0.61	53
7–8 years	2.84	0.62	29
8–9 years	2.77	0.64	33
9–10 years	2.98	0.68	34
10–11 years	2.82	0.61	13
11–12 years	2.92	0.78	13
12–13 years	3.15	0.58	19
Controls			
3–4 years	4.01	0.40	42
4–5 years	4.21	0.39	47
5–6 years	4.20	0.37	52
6–7 years	4.19	0.33	70
7–8 years	4.24	0.40	51
8–9 years	4.19	0.32	63
9–10 years	4.25	0.40	54
10–11 years	4.30	0.31	40
11–12 years	4.26	0.33	51
12–13 years	4.15	0.33	34

Note. Absolute range, 1–5.

Diagnostic Comparisons

For the gender-referred group alone, Table VI shows the demographic variables and the GIQ total score as a function of diagnostic status, that is, whether or not the proband met the complete *DSM* criteria for GID. At the time of assessment, the probands who met the complete *DSM* criteria for GID were significantly younger than those who were subthreshold, $t(323) = 8.49$, $p < .001$. They were also more likely to be living in a two-parent family, $^2(1) = 5.03$, $p < .025$) and tended to have a higher social class background, $t(323) = 1.93$, $p = .054$. The two diagnostic subgroups did not differ significantly with regard to sex composition and IQ. Accordingly, in comparing the diagnostic subgroups on the GIQ total score, we covaried age and parent's social class and marital status.

A 2 (diagnostic group) $\times$ 2 (sex) ANCOVA revealed main effects for Diagnostic Group, $F(1, 318) = 57.0$, $p < .001$, and Sex, $F(1, 318) = 7.1$, $p < .01$. Probands who met the complete *DSM* criteria for GID had a significantly higher deviant total score than those who were subthreshold and, as noted earlier, the girls had a significantly higher deviant total score than the boys. Collapsed across sex, the ES, using the SD of the subthreshold group, was 1.37, which was substantially smaller than the proband-control comparison (see earlier).

Table VII shows the means and SDs as a function of diagnostic group and age block for the gender-referred group alone. Because the number of participants in some cells was small, we combined the 3–4- and 4–5-year-olds and 10–11-, 11–12-, and 12–13-year-olds. A 2 (diagnostic group) $\times$ 7 (age block) ANOVA revealed only a main effect for Diagnostic Group, $F(1, 311) = 114.1$, $p < .001$.

Table VI. Demographic Variables and Maternal Ratings on the Gender Identity Questionnaire as a Function of Diagnostic Status for Gender Identity Disorder

Variable		Complete *DSM* criteria for gender identity disorder: Yes	No
Age (in years)	*M*	6.37	8.62
	SD	2.06	2.59
	N	216	109
Sex			
Boys	*N* (%)	182 (84.3)	93 (85.3)
Girls	*N* (%)	34 (15.7)	16 (14.7)
Full-scale IQ	*M*	104.5	102.66
	SD	16.17	18.4
	N	212	106
Parents' social class	*M*	44.12	40.75
	SD	14.83	14.9
	N	216	109
Parents' marital status			
Two parent	*N* (%)	155 (71.8)	64 (58.7)
Other	*N* (%)	61 (28.2)	45 (41.3)
GIQ total score	*M*	2.59	3.29
	SD	0.53	0.51
	N	216	109

Table VII. Maternal Ratings on the Gender Identity Questionnaire as a Function of Diagnostic Group and Age Block (Gender-Referred Group Alone)

Diagnostic group	*M*	*SD*	*N*
Gender-referred: Criteria met for GID			
3–5 years	2.56	0.49	65
5–6 years	2.73	0.51	46
6–7 years	2.62	0.52	41
7–8 years	2.54	0.55	16
8–9 years	2.53	0.61	20
9–10 years	2.49	0.66	11
10–13 years	2.52	0.60	17
Gender-referred: Subthreshold for GID			
3–5 years	3.54	0.43	13
5–6 years	3.39	0.52	7
6–7 years	3.42	0.46	12
7–8 years	3.20	0.50	13
8–9 years	3.14	0.50	13
9–10 years	3.22	0.57	23
10–13 years	3.27	0.51	28

Note. Absolute range, 1–5.

Table VIII shows the ES for three contrasts blocked by age: the GID threshold versus subthreshold cases; the GID threshold cases versus the controls, and the GID subthreshold cases versus the controls. Of course, the GID threshold cases had much larger ES than the GID subthreshold cases when compared to the controls. Of greater interest is the comparison of the GID threshold versus subthreshold cases and the GID subthreshold cases versus the controls. With the exception of the 3–5-year-age block, the ES were substantially larger for the GID subthreshold cases versus the controls when compared to the GID threshold versus subthreshold cases. It is of note that the largest ES for the subthreshold-control cases was for the 8–9-year-old group, which corresponds to the mean age at assessment (referral) of the subthreshold cases (see Table VI).

Sensitivity and Specificity

If the GIQ were to be used as a screening questionnaire, it would be important to have a high threshold for specificity, that is, to have a low rate of false positives for controls. Accordingly, we set the specificity rate at 95% ($\leq$3.54). This yielded a sensitivity rate of 86.8% for the gender-referred group (for the threshold cases alone, the sensitivity was 96.3%; for the subthreshold cases alone, the sensitivity was 67.9%). When we used a more stringent specificity rate of 99% ($\leq$3.14), the sensitivity rate fell to 69.5% (for the threshold cases alone, 84.3%; for the subthreshold cases, 40.4%).

We also examined the relation between specificity and sensitivity for the gender-referred group alone by comparing the threshold versus subthreshold cases. On the one hand, an argument could be made that it would be important to not make a false positive diagnosis based on a screening questionnaire for the subthreshold cases. On the other hand, it could be argued that the clinical needs of the subthreshold cases may not be that substantially different from the threshold cases, at least in certain domains of functioning (say, e.g., peer relations). Thus, it would be important to minimize the number of false negatives. With a 95% specificity rate ($\leq$2.50), sensitivity was only 45.4%. With an 80% specificity rate ($\leq$2.83), sensitivity rose to 65.7%, and, at a 70% specificity rate ($\leq$3.57), sensitivity was 97.7%.

DISCUSSION

The results of this study showed that the GIQ had excellent psychometric properties. Although the number of items contained in the questionnaire was relatively small compared to other measures, such as the CBAQ, they were intended to cover, in a relatively global manner, most of the core behavioral features that currently comprise the *DSM* diagnosis of GID. The factor analysis indicated that a one-factor solution provided the best fit to the data. It should be noted that the majority of the items on the factor pertained to gender role behaviors (e.g., peer affiliation preference, roles in fantasy play, cross-dressing) and only two items related to gender identity per se (i.e., the verbalized wish to be of the opposite sex or verbalized statements that one "is" a member of the opposite sex). One other item that loaded on the factor pertained to verbalized statements

Table VIII. Proband-Control Effect Sizes as a Function of Diagnostic Group and Age Block

	Age block						
Contrast	3–5	5–6	6–7	7–8	8–9	9–10	10–13
GID+ vs. GID−	2.27	1.26	1.73	1.32	1.22	1.28	1.47
GID+ vs. Control	3.92	3.97	4.75	4.25	5.18	4.40	5.37
GID− vs. Control	1.44	2.18	2.33	2.60	3.28	2.57	3.03

Note. GID + = met complete *DSM* criteria for Gender Identity Disorder. GID− = subthreshold for complete *DSM* criteria for Gender Identity Disorder.

of anatomic dysphoria. Given the conceptual distinction that is often made between gender identity and gender role (for definitions, see Zucker & Bradley, 1995), one could have predicted, for example, a two-factor solution. It is likely that a one-factor solution emerged as the best fit because of the high correlation between cross-gender identity and cross-gender role in a sample that included a large percentage of children with potential problems in their gender identity development.

Perhaps because most of the items covered in the questionnaire pertained to observable behaviors (e.g., sex of playmates preference, toy interests, etc.), the mother–father correlation was much higher than is typically found on standardized parent-report questionnaires pertaining to general behavioral problems, which cover a much wider range of behaviors, some of which are less easily observed (Achenbach, McConaughy, & Howell, 1987; Duhig, Renk, Epstein, & Phares, 2000).

As can be seen in Table III, most of the factor loadings were greater than .60. The weakest factor loading that was greater than .30 was for Item 3 (.34), which pertained to playing with "boy-type" dolls, with "G.I. Joe" and "Ken" given as examples. It is possible that this relatively weak loading was the result of examples that were either outdated (in the case of "G.I. Joe") or perceived by parents as really more a companion to "Barbie" (i.e., the "Ken" doll). Thus, a revision to the GIQ should consider a modification to the wording of the item, such as "He (She) plays with "boy-type" dolls or action heroes, with more contemporary examples (e.g., the Power Rangers, Spiderman, etc.).

Item 15, which pertained to verbalized anatomic dysphoria, was a relatively crude attempt to index this component of the *DSM* criteria. Although the mothers judged its presence to be more common among the probands than the controls (see Table II), it should be noted that the majority rated it as "never" occurring (76.8% of the boy probands and 64.0% of the girl probands compared to 98.3% of the boy controls and 95.9% of the girl controls). Although this may reflect accurately the relatively low salience of anatomic dysphoria among prepubertal children with GID, it is clear that more work is required in order to develop a more comprehensive approach to its assessment. For this, it would probably be useful to generate a more systematic structured interview schedule with both parents and children regarding various indicators of possible anatomic dysphoria (cf. Lothstein, 1992).

The effect size for the proband-control comparison for the GIQ total score was quite large, indicating how markedly different children with GID are with regard to the core phenomenology compared to control children. More importantly, the total score appears to have strong sensitivity using a conventional specificity rate of 95%.

As shown in Table V, the relative absence of age effects (blocked by year) suggests that, by age 3–4, sex-typed behavior is organized in a fairly coherent manner and remains stable over time at least to the end of what is typically considered "childhood" in Western culture. From a treatment standpoint, the relative absence of age effects suggests that the GIQ may be a reasonable measure to assess change over time (e.g., in relation to specific therapeutic interventions to alter GID).

The GIQ was able to distinguish, within the group of gender-referred probands alone, those who met the complete *DSM* criteria for GID versus those who did not. This provides some evidence for the validity of the clinician diagnostic ratings. Of course, one has to be cautious in making strong conclusions about this because the clinician ratings were based primarily on parent-report in the more unstructured assessment interview and the GIQ was completed by the same informants. Thus, the relation between the two measures had method variance only. Elsewhere, however, we have shown that the clinician diagnostic ratings were significantly related to independent measures of sex-typed behavior of the probands, providing some evidence therefore for the validity of the diagnostic subgrouping (i.e., threshold versus subthreshold cases) (see, e.g., Zucker & Bradley, 1995; Zucker, Finegan, Doering, & Bradley, 1984). Future studies should be able to examine this relation for the GIQ as well.

It should be noted that critics have expressed concern that the *DSM* criteria may not adequately differentiate children with GID from those children who merely show a pattern of "gender nonconforming" behavior but who are not "truly" GID (e.g., Corbett, 1996; Haldeman, 2000; Richardson, 1996). Although the results of this study showed that the *DSM* criteria were successful in distinguishing threshold versus subthreshold cases on the GIQ, sensitivity was lower when comparing these two subgroups. This is not a particularly surprising finding. Given that the *DSM* system is a categorical one, it is to be expected that there is going to be more overlap when comparing threshold versus subthreshold cases as opposed to comparing probands versus controls. Whether or not revisions to the *DSM* criteria can result in a sharper distinction between threshold and subthreshold cases using independent measures is an empirical issue that requires further exploration.

From a clinical utility standpoint, the GIQ appears to have reasonable practical applications. For an individual child, a clinician can utilize this questionnaire by calculating the mean total score and see to what extent it approximates the mean for probands in a particular age range and to what extent it departs from the mean of the controls in the same age range. Such an application may be helpful in formulating the therapeutic needs of an

individual child for the office-based practitioner for whom more extensive diagnostic assessment tools are not readily available.

APPENDIX: GENDER IDENTITY QUESTIONNAIRE (BOY VERSION; SCORING WEIGHTS IN PARENTHESES)

Instructions: Please answer the following behavioral statements as they currently characterize your child's behavior. For each question, circle the response that most accurately describes your child.

1. His favorite playmates are
   a. always boys (5)
   b. usually boys (4)
   c. boys and girls equally (3)
   d. usually girls (2)
   e. always girls (1)
   f. does not play with other children
2. He plays with girl-type dolls, such as "Barbie"
   a. as a favorite toy (1)
   b. frequently (2)
   c. once-in-a-while (3)
   d. rarely (4)
   e. never (5)
3. He plays with boy-type dolls, such as "G.I. Joe" or "Ken"
   a. as a favorite toy (5)
   b. frequently (4)
   c. once-in-a-while (3)
   d. rarely (2)
   e. never (1)
4. He experiments with cosmetics (makeup) and jewelry
   a. as a favorite activity (1)
   b. frequently (2)
   c. once-in-a-while (3)
   d. rarely (4)
   e. never (5)
5. He imitates female characters seen on TV or in the movies
   a. as a favorite activity (1)
   b. frequently (2)
   c. once-in-a-while (3)
   d. rarely (4)
   e. never (5)
6. He imitates male characters seen on TV or in the movies
   a. as a favorite activity (5)
   b. frequently (4)
   c. once-in-a-while (3)
   d. rarely (2)
   e. never (1)
7. He plays sports with boys (but not girls)
   a. as a favorite activity (5)
   b. frequently (4)
   c. once-in-a-while (3)
   d. rarely (2)
   e. never (1)
8. He plays sports with girls (but not boys)
   a. as a favorite activity (1)
   b. frequently (2)
   c. once-in-a-while (3)
   d. rarely (4)
   e. never (5)
9. In playing "mother/father," "house," or "school" games, he takes the role of
   a. a girl or woman at all times (1)
   b. usually a girl or woman (2)
   c. half the time a girl or woman and half the time a boy or man (3)
   d. usually a boy or man (4)
   e. a boy or man at all times (5)
   f. does not play these games
10. He plays "girl-type" games (as compared to "boy-type" games)
    a. as a favorite activity (1)
    b. frequently (2)
    c. once-in-a-while (3)
    d. rarely (4)
    e. never (5)
11. He plays "boy-type" games (as compared to "girl-type" games)
    a. as a favorite activity (5)
    b. frequently (4)
    c. once-in-a-while (3)
    d. rarely (2)
    e. never (1)
12. In dress-up games, he likes to dress up
    a. a girl or woman at all times (1)
    b. usually a girl or woman (2)
    c. half the time a girl or woman and half the time a boy or man (3)
    d. usually a boy or man (4)
    e. a boy or man at all times (5)
    f. does not play these games
13. He states the wish to be a girl or a woman
    a. every day (1)
    b. frequently (2)
    c. once-in-a-while (3)
    d. rarely (4)
    e. never (5)

14. He states that he is a girl or a woman
    a. every day (1)
    b. frequently (2)
    c. once-in-a-while (3)
    d. rarely (4)
    e. never (5)
15. He talks about not liking his sexual anatomy (private parts)
    a. every day (1)
    b. frequently (2)
    c. once-in-a-while (3)
    d. rarely (4)
    e. never (5)
16. He talks about liking his sexual anatomy (private parts)
    a. every day (5)
    b. frequently (4)
    c. once-in-a-while (3)
    d. rarely (2)
    e. never (1)

Note. With the appropriate pronoun changes, the Girl Version is identical to the Boy Version.

ACKNOWLEDGMENT

Dr. Mitchell died on March 16, 2000 and we dedicate this article to her memory.

REFERENCES

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213–232.

Alexander, G. M., & Hines, M. (1994). Gender labels and play styles: Their relative contribution to children's selection of playmates. *Child Development, 65,* 869–879.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

Bates, J. E., & Bentler, P. M. (1973). Play activities of normal and effeminate boys. *Developmental Psychology, 9,* 20–27.

Bates, J. E., Bentler, P. M., & Thompson, S. K. (1973). Measurement of deviant gender development in boys. *Child Development, 44,* 591–598.

Bates, J. E., Bentler, P. M., & Thompson, S. K. (1979). Gender-deviant boys compared with normal and clinical control boys. *Journal of Abnormal Child Psychology, 7,* 243–259.

Birkenfeld-Adams, A. S. (1999). *Quality of attachment in young boys with gender identity disorder: A comparison to clinic and nonreferred control boys*. Unpublished doctoral dissertation, York University, Downsview, Ontario.

Corbett, K. (1996). Homosexual boyhood: Notes on girlyboys. *Gender and Psychoanalysis, 1,* 429–461.

Doering, R. W. (1981). *Parental reinforcement of gender-typed behaviors in boys with atypical gender identity*. Unpublished doctoral dissertation, University of Toronto.

Doering, R. W., Zucker, K. J., Bradley, S. J., & MacIntyre, R. B. (1989). Effects of neutral toys on sex-typed play in children with gender identity disorder. *Journal of Abnormal Child Psychology, 17,* 563–574.

Duhig, A. M., Renk, K., Epstein, M. K., & Phares, V. (2000). Interparental agreement on internalizing, externalizing, and total behavior problems: A meta-analysis. *Clinical Psychology: Science and Practice, 7,* 435–453.

Elizabeth, P. H., & Green, R. (1984). Childhood sex-role behaviors: Similarities and differences in twins. *Acta Geneticae Medicae et Gemellologiae, 33,* 173–179.

Elizabeth, P. H., Meyer-Bahlburg, H. F. L., Janal, M. N., Ehrhardt, A. A., & Green, R. (1984, September). *Do human twins show a "free-martin" like phenomenon*? Poster session presented at the meeting of the International Academy of Sex Research, Cambridge, England.

Fridell, S. R. (2001). *Sex-typed play behavior and peer relationships in boys with gender identity disorder*. Unpublished doctoral dissertation, Ontario Institute for Studies in Education of the University of Toronto.

Green, R. (1976). One-hundred ten feminine and masculine boys: Behavioral contrasts and demographic similarities. *Archives of Sexual Behavior, 5,* 425–446.

Green, R. (1987). *The "sissy boy syndrome" and the development of homosexuality*. New Haven, CT: Yale University Press.

Green, R., & Fuller, M. (1973). Family doll play and female identity in preadolescent males. *American Journal of Orthopsychiatry, 43,* 123–127.

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.

Green, R., Fuller, M., Rutley, B., & Hendler, J. (1972). Playroom toy preferences of fifteen masculine and fifteen feminine boys. *Behaviour Research and Therapy, 3,* 425–429.

Haldeman, D. C. (2000). Gender atypical youth: Clinical and social issues. *School Psychology Review, 29,* 192–200.

Hollingshead, A. B. (1975). *Four-factor index of social status*. Unpublished manuscript, Department of Sociology, Yale University, New Haven, CT.

Klein, A. R., & Bates, J. E. (1980). Gender-typing of game choices and qualities of boys' play behavior. *Journal of Abnormal Child Psychology, 8,* 201–212.

Kuksis, M. R. (1992). *Intercorrelations among risk, vulnerability, and protective factors in child psychopathology*. Unpublished doctoral dissertation, York University, Downsview, Ontario.

Lothstein, L. M. (1992). Clinical management of gender dysphoria in young boys: Genital mutilation and *DSM* IV implications. *Journal of Psychology and Human Sexuality, 5*(4), 87–106.

Maing, D. M. (1991). *Patterns of psychopathology in sexually abused girls*. Unpublished doctoral dissertation, University of Windsor.

Meyer-Bahlburg, H. F. L., Feldman, J. F., & Ehrhardt, A. A. (1985). Questionnaires for the assessment of atypical gender role behavior: A methodological study. *Journal of the American Academy of Child Psychiatry, 24,* 695–701.

Meyer-Bahlburg, H. F. L., Sandberg, D. E., Dolezal, C. L., & Yager, T. J. (1994). Gender-related assessment of childhood play. *Journal of Abnormal Child Psychology, 22,* 643–660.

Meyer-Bahlburg, H. F. L., Sandberg, D. E., Yager, T. J., Dolezal, C. L., & Ehrhardt, A. A. (1994). Questionnaire scales for the assessment of atypical gender development in girls and boys. *Journal of Psychology and Human Sexuality, 6*(4), 19–39.

Mitchell, J. N. (1991). *Maternal influences on gender identity disorder in boys: Searching for specificity*. Unpublished doctoral dissertation, York University, Downsview, Ontario.

Rekers, G. A. (1975). Stimulus control over sex-typed play in cross-gender identified boys. *Journal of Experimental Child Psychology*, *20*, 136–148.

Rekers, G. A., & Yates, C. E. (1976). Sex-typed play in feminoid boys versus normal boys and girls. *Journal of Abnormal Child Psychology*, *4*, 1–8.

Richardson, J. (1996). Setting limits on gender health. *Harvard Review of Psychiatry*, *4*, 49–53.

Sandberg, D. E., & Meyer-Bahlburg, H. F. L. (1994). Variability in middle childhood play behavior: Effects of gender, age, and family background. *Archives of Sexual Behavior*, *23*, 645–663.

Sandberg, D. E., Meyer-Bahlburg, H. F. L., Ehrhardt, A. A., & Yager, T. J. (1993). The prevalence of gender-atypical behavior in elementary school children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *32*, 306–314.

Skilbeck, W. M., Bates, J. E., & Bentler, P. M. (1975). Human figure drawings of gender-problem and school-problem boys. *Journal of Abnormal Child Psychology*, *3*, 191–199.

Tuber, S., & Coates, S. (1985). Interpersonal phenomena in the Rorschachs of extremely feminine boys. *Psychoanalytic Psychology*, *2*, 251–265.

Zucker, K. J. (1992). Gender identity disorder. In S. R. Hooper, G. W. Hynd, & R. E. Mattison (Eds.), *Child psychopathology: Diagnostic criteria and clinical assessment* (pp. 305–342). Hillsdale, NJ: Erlbaum.

Zucker, K. J., & Bradley, S. J. (1995). *Gender identity disorder and psychosexual problems in children and adolescents*. New York: Guilford Press.

Zucker, K. J., Bradley, S. J., Lowry Sullivan, C. B., Kuksis, M., Birkenfeld-Adams, A., & Mitchell, J. N. (1993). A gender identity interview for children. *Journal of Personality Assessment*, *61*, 443–456.

Zucker, K. J., Doering, R. W., Bradley, S. J., & Finegan, J. K. (1982). Sex-typed play in gender-disturbed children: A comparison to sibling and psychiatric controls. *Archives of Sexual Behavior*, *11*, 309–321.

Zucker, K. J., Finegan, J. K., Doering, R. W., & Bradley, S. J. (1983). Human figure drawings of gender-problem children: A comparison to sibling, psychiatric, and normal controls. *Journal of Abnormal Child Psychology*, *11*, 287–298.

Zucker, K. J., Finegan, J. K., Doering, R. W., & Bradley, S. J. (1984). Two subgroups of gender-problem children. *Archives of Sexual Behavior*, *13*, 27–39.

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*, 295–310.

Zuger, B., & Taylor, P. (1969). Effeminate behavior present in boys from early childhood. II. Comparison with similar symptoms in non-effeminate boys. *Pediatrics*, *44*, 375–380.