Children's Gender-Related Self-Perceptions, Activity Preferences, and Occupational Stereotypes: A Test of Three Models of Gender Constructs[1]

Janet T. Spence[2]
University of Texas at Austin

Sharon K. Hall
University of Houston Clear Lake

Fourth through sixth grade boys (n = 197) and girls (n = 271) were given a simplified form of the Personal Attributes Questionnaire [J. T. Spence and R. L. Helmreich (1978b) The Intermediate Personal Attributes Questionnaire: A Simplified Version for Children and Adults, unpublished manuscript, Department of Psychology, University of Texas at Austin], assessing desirable instrumental and expressive traits; subsets of items from J. P. Boldizar's [(1991) "Assessing Sex Typing and Androgyny in Children: The Children's Sex Role Inventory," Developmental Psychology, Vol. 27, pp. 505–513] children's version of the Bem Sex Role Inventory [S. L. Bem (1974) "The Measurement of Psychological Androgyny," Journal of Consulting and Clinical Psychology, Vol. 42, 155–162]; S. L. Harter's [(1985) Manual for the Self-Perception Profile for Children, Denver: University of Denver] measures of self-esteem; and measures of masculine and feminine activity preferences and prescriptive occupational stereotypes. The children were predominantly white and from middle-class backgrounds. The correlations among the gender-related measures were more congruent with a multifactorial approach to gender than the unifactorial gender schema model or the two-factor model of masculinity and

femininity. Instrumentality, however, was correlated with self-esteem in both genders, a finding most reasonably interpreted in terms of this personality variable per se.

Many contemporary students of gender have not only regarded instrumental and expressive characteristics as central to the understanding of psychological differences between males and females in both childhood and later years, but have treated these trait dimensions as windows to the core of masculinity and femininity. These views are not new but have roots in classic theories of gender. One of the purposes of the present study, which employed children in the middle elementary school grades, was to explore the implications of two models of gender that assume a linkage between instrumentality and expressiveness and broad-gauged gender concepts, and to demonstrate the utility of an alternative, multifactorial approach.

THEORETICAL BACKGROUND

A number of male and female principles that purportedly constitute fundamental psychological distinctions between the genders have been proposed, e.g., orientation to outer vs. inner space (Erikson, 1964), agency vs. communion (Bakan, 1966), individualistic vs. collectivistic (Markus & Kitayama, 1991), and ranking vs. linking (e.g., Sidanius, Cling, & Pratto, 1991). These various theories have in common the implication that males are more instrumental than females—that is, more concerned with mastering the challenges of the external world, whereas females are more expressive and concerned with the well-being of others. These beliefs in gender differences in instrumentality and expressiveness are not restricted to social scientists but have been shown to be widely accepted by society at large.

Assumptions that in the past have been part of these theories are that instrumental and expressive traits are key components of the hypothetical masculinity and femininity concepts, and like the latter, are psychologically incompatible and hence negatively correlated. As noted some time ago by Constantinople (1973), these several presumptions are also illustrated by conventional masculinity–femininity tests in which respondents are given a single score representing their position on a hypothetical masculinity–femininity continuum. The tacit assumption is that observable gender-differentiating phenomena, particularly instrumental and expressive traits, are all indicators of an underlying bipolar factor, masculinity–femininity.

CONTEMPORARY PERSPECTIVES

Several alternatives to these traditional conceptions have been advanced in recent years, spurred on by a change in ideological climate. Most of the research relevant to these theoretical models have utilized one of two novel self-report instrument—the Bem Sex Role Inventory (BSRI; Bem, 1974) or the Personal Attributes Questionnaire (PAQ; Spence, Helmreich, & Stapp, 1974; Spence & Helmreich, 1978a). The items on both of these inventories predominantly describe desirable instrumental traits stereotypically associated with males and desirable expressive traits stereotypically associated with females. In a critical departure from previous practice, the masculine and feminine characteristics are assessed in separate scales. Contrary to the conventional assumption that masculine and feminine characteristics are incompatible, the two scales on both instruments have been found to be essentially uncorrelated.

Bem's Unifactorial Theory of Sex Typing

The initial departure from traditional wisdom is represented by the theoretical propositions advanced by Bem (1974, 1977, 1981, 1985). As suggested by the title of her inventory, Bem presumed that the BSRI assessed the global concepts of sex role identification[3] and masculinity and femininity. (These presumptions, which continue to widely shared, have been extended by many investigators to the PAQ.) Bem has woven her findings with the BSRI into a theory of individual differences in gender-typing or gender schematization that simultaneously postulates that masculinity and femininity are independent [as indicated by the lack of association between scores on the Masculinity (M) and Femininity (F) scales of the BSRI], and retains the conventional notion of a single bipolar continuum of masculinity–femininity essentially based on the difference between M and F scores. However, in contrast to previous investigators who were concerned with those scoring at the extremes, Bem (1974, 1977, 1981) focused attention on the substantial numbers of men and women who fall at the middle of the hypothetical masculinity–femininity distribution, in the case of the BSRI, those whose M and F scores are approximately equal, whatever their absolute level. These individuals she has variously identified as nongender-typed, gender aschematic, or low in gender role identification. Conversely, those with unequal M and F scores in the stereotypic direction she has

---

[3]Bem (1977) initially referred to sex role identification and sex tying rather than to gender role identification and gender typing. In this article, we have substituted gender for sex in recognition of the current practice of reserving the latter for biological phenomena.

identified as gender-typed, gender schematic, or high in gender role identification. In categorical terms (Spence, Helmreich, & Stapp, 1975; Bem, 1977), she has described nongender-typed males and females as falling into two subgroups: androgynous individuals, high in both their M and F scores, and undifferentiated individuals, low in both their M and F scores. Gender-typed individuals are composed of masculine males, high in masculinity and low in femininity as assessed by their BSRI scores, and of feminine females, who exhibit the opposite pattern.[4] She postulates that in comparison to the nongender-typed, males and females who are gender typed in their BSRI scores are more likely to exhibit other age-appropriate characteristics and behaviors stereotypically associated with their gender, to embrace conventional gender ideologies, and to organize information about themselves and the external world using gender schemata. As she has recently stated, "The gendered personality . . . has a readiness to superimpose a gender-based classification on every heterogeneous collection of human possibilities that presents itself" (Bem, 1993, p. 154).

The Two-Factor Model of Masculinity and Femininity

If it is presumed that scores on the gender-related scales of the BSRI and the PAQ are measures of the masculinity and femininity constructs, another interpretation of the finding that scores on the two scales are uncorrelated is that the global masculinity and femininity constructs are themselves unrelated. Some investigators appear to have drawn this conclusion. It follows from this conception that scores on the M scale of the BSRI or PAQ but not scores on the F scale should be correlated with a wide variety of stereotypically masculine characteristics and behaviors, and conversely, that scores on the F scale but not the M scale should be correlated with a wide variety of stereotypically feminine characteristics and behaviors. Although some investigators have explicitly identified the two models in an effort to evaluate the relative merits of the single- and the two-factor approaches (e.g., Edwards & Spence, 1987; Payne, Connor, & Colletti, 1987), others have not distinguished between them, even when their findings appear to support the two-factor conception (e.g., Boldizar, 1991).

Although it might seem that straightforward predictions can be derived from the single-factor and the two-factor models, each has problematic as-

---

[4]This generalization applies only to those whose score discrepancies are in the stereotypic direction (M > F in the case of males and M < F in the case of females). Some men and women exhibit a M–F discrepancy in the nontraditional direction; these feminine men and masculine women can be classified as cross-gender typed. Bem has either ignored these individuals or stated her uncertainty about what to predict.

pects that are mirror images of one another. The unidimensional concep-
tion predicts, for example, that gender-typed persons (as defined by their
BSRI or PAQ scores) are higher in their endorsement of gender ideology
and in their use of gender schema in processing information than nongen-
der-typed persons. It seems implausible, however, to expect that androgy-
nous and undifferentiated men and women (i.e., members of the two
nongender-typed groups) are alike in all of their other gender-related char-
acteristics and behaviors, e.g., their preferences and choices of masculine
and feminine leisure activities, particularly those that may be influenced
by instrumentality or expressiveness per se.

   The two-factor model, on the other hand, yields obvious predictions
about the latter kind of example: scores on the BSRI (or PAQ) M scale
should be positively correlated with masculine preferences whereas scores
on the F scale should be correlated with feminine preferences. The theory's
implications are uncertain, however, for gender-related phenomena that
cannot be divided into independent masculine and feminine categories, for
example, gender-role ideology.

Multifactorial Theory

   A third model, which does not suffer from the internal contradictions
we have just outlined, is beginning to receive attention, not only among
those working with adolescents and adults but also with children (e.g.,
Aube, Norcliffe, & Koestner, 1995; Ashmore, 1990; Biernat, 1991; Downs
& Langlois, 1988; Katz, 1979; Katz & Ksansek, 1994; Marsh & Byrne, 1991;
Orlofsky, 1981; Serbin, Powlishta, & Gulko, 1993; Signorella, 1987, 1992;
Silvern & Katz, 1986; Spence & Helmreich, 1978a; Spence, 1985). Accord-
ing to this position, to which we subscribe, the heterogeneous collection of
self-perceptions, beliefs, and behaviors that distinguish between males and
females do not contribute to a single factor or even to two factors, but
instead are multifactorial. On a descriptive level, the basic proposition of
those who adopt the multifactorial approach is that within each gender,
various classes of gender-related attributes, actions, and beliefs are not nec-
essarily correlated with one another so that knowledge of a person's stand-
ing on one set of gender-related measures cannot automatically be assumed
to yield information about their standing on another set.

   The model that we embrace more specifically proposes that at least
from middle childhood on, self-images of desirable instrumental and ex-
pressive traits are unlikely to be associated with gender-associated charac-
teristics and attitudes unaffected by these personality dimensions per se.
We further propose that the particular set of gender-related characteristics

exhibited by a person of a given age varies from one male or female to the next. Furthermore, gender-related attributes and behaviors are multiply determined and among individuals, differ in their etiology. These propositions, in turn, imply that the equation of self-reports of instrumental and expressive traits with individual differences in global constructs such as masculinity, femininity, and gender schematization is unwarranted.

THE PAQ AND BSRI AND SIMPLER VERSIONS OF THEM

Do the PAQ and BSRI Measure the Same Constructs?

In our discussion so far, we have treated the PAQ and BSRI as essentially interchangeable measures of desirable instrumental and expressive traits. The theoretical question we have posed is whether these inventories measure only these trait dimensions, as our multifactorial approach postulates, or whether they also measure such gender concepts as masculinity and femininity or gender schematization, as presumed by the one- and two-factor models.

Evaluation of these rival positions is somewhat complicated by the presence on the BSRI of other kinds of items in addition to those describing desirable instrumental and expressive traits. Still further complications have been introduced by Frable and Bem (1985; Frable, 1989), who argue that because of the different theories of their developers, the PAQ measures only instrumental and expressive traits, whereas the BSRI also measures the gender constructs specified in Bem's theory: thus only the BSRI and data obtained with it are appropriate to use in evaluating her theory's validity. Claims for the superiority of the BSRI and a children's version of it (Boldizar, 1991) as measures of gender typing, based on their more varied content, have also been advanced by O'Heron and Orlofsky (1990) and by Boldizar (1991).

As will later become apparent, the potential seriousness of these contentions led us to incorporate an investigation of them into the design of the present study. It thus is necessary to describe the properties of the BSRI and PAQ, along with the various versions of these questionnaires, in some detail.

Properties of the PAQ and BSRI and Their Variants

In its present form, the PAQ (Spence & Helmreich, 1978a) contains two major 8-item scales, one consisting exclusively of socially desirable instrumental traits (I scale) and the other consisting exclusively of socially desirable expressive traits (E scale) that stereotypically distinguish between

men and women (Spence et al., 1974, 1975). Factor analyses (e.g., Helm-reich, Spence, & Wilhelm, 1981) indicate that in each gender, each scale is unifactorial and that the I and E scales are essentially uncorrelated.

The BSRI (Bem, 1974) contains two 20-item gender-related scales, la-beled Masculinity (M) and Femininity (F). Items on the M scale predomi-nantly refer to desirable instrumental traits, similar or in several cases identical to those on the PAQ, but it also contains three items of different content, most notably the adjective "masculine," which refers directly to gender. The F scale is even more mixed. Eleven of its 20 items describe desirable expressive traits similar to those found on the PAQ E scale. It also contains other kinds of personality traits, most of which have been judged to be undesirable for both sexes (e.g., Pedhazur & Tetenbaum, 1979) and to have different correlates from the desirable items (e.g., Spence, Helmreich, & Holahan, 1979; Spence, 1993). The F scale also in-cludes the adjective "feminine." (See Spence, 1991, for a classified listing of the items on the PAQ and the BSRI.) Statistical analyses of the BSRI items show that the items contribute to multiple factors in each gender (e.g., Pedhazur & Tetenbaum, 1979; Martin & Ramanaiah, 1988). Of par-ticular importance is the finding that in both men and women, responses to the items "masculine" and "feminine" are negatively correlated.

As consideration of their content suggests, substantial correlations in both sexes have been found between the parallel PAQ and BSRI scales, particularly when only the desirable instrumental and expressive items from the BSRI are included (e.g., Lubinski et al., 1983; Spence, 1993). The desirable I and E items on the two instruments also yield similar results when related to various criterion variables. Taken together, the findings indicate that if the BSRI is a valid a measure of gender concepts whereas the PAQ is not, as has been claimed, this difference is traceable to the BSRI items that do not describe desirable instrumental and expressive attributes.

Several versions of the PAQ appropriate for younger age groups have been developed (Hall & Halberstadt, 1980; Simms, Davis, Foushee, Hola-han, Spence, & Helmreich, 1978; Spence & Helmreich, 1978b). All of these revisions involve quite straightforward translations of the original trait terms into simpler language and item format. With the exception of the items "masculine" and "feminine," the same is true of the Children's Sex Role Inventory (Children's SRI), the children's form of the BSRI devel-oped by Boldizar (1991). In Boldizar's version, the abstract terms "mascu-line" and "feminine" on the BSRI were changed to "I like to do things that boys and men do" and "I like to do things that girls and women do." The significance of these substitutions is amplified at a later point.

PRIOR FINDINGS FROM ADOLESCENTS AND ADULTS

In reviewing the empirical evidence, we treat the PAQ and BSRI as equivalent, returning later to the implications of the BSRI's more diverse content, and to the assertion that only the BSRI measures gender constructs.

Gender Differences

The BSRI and the PAQ have been employed in literally hundreds of studies that have tested samples of males and females ranging from high school students to individuals of late middle age (e.g., to cite only some of the earliest investigations, Abraham, Feldman, & Nash, 1978; Bem, 1974; O'Connor, Mann, & Bardwick, 1978; Spence & Helmreich, 1978a). These studies have consistently demonstrated that males score significantly higher than females on the PAQ I and the BSRI M scales and significantly lower than females on the PAQ E and the BSRI F scales.

Relationships with Self-Esteem

Another consistent finding in the literature is the relationship between the M and I scales of the BSRI and PAQ and measures of self-esteem: in both genders, positive correlations, typically substantial in magnitude, uniformly occur (see, for example, the 1983 Whitley meta-analysis). The correlations with the PAQ E scale tend to be close to zero, and in the case of the BSRI F scale, they tend to be negative (e.g., Antill & Cunningham, 1979).[5]

The spate of investigations aimed at determining relationships between the BSRI and PAQ and self-esteem was initially stimulated by Bem's (1974) suggestion that nongender-typed men and women are mentally healthier than those who are gender typed. However, a number of investigators have concluded from the empirical findings that self-esteem is primarily associated with masculinity in both sexes (Aube, Norcliffe, Craig, & Koestner, 1995; Whitley, 1983). When the data are considered within the context of the general personality literature, we suggest instead that they are more

---

[5]For the sake of completeness, we note that in instances in which measures are used that tap feelings of social competence and interpersonal adjustment and satisfaction, positive correlations are also found with scores on the PAQ E scale, and with undesirable items removed, the BSRI F scale (e.g., Aube, Norcliffe, Craig, & Koestner, 1995; Hunt, 1993; Lubinski, Tellegen, & Butcher, 1983; Spence & Helmreich, 1978a). In the present study, it was sufficient for our theoretical purposes to measure only self-esteem.

reasonably interpreted in terms of the desirable aspects of instrumentality per se rather than the general masculinity construct. This interpretation is strengthened by failures to find correlations between self-esteem and such measures as gender role attitudes (e.g., Spence et al., 1975; Spence & Helmreich, 1978a).

Our contention, then, is that the data on self-esteem (and interpersonal adjustment: see Footnote 5) are irrelevant to questions about the construct validity of the BSRI or the PAQ as measures of masculinity and femininity.

Relationships of the BSRI and PAQ with Other Gender-Related Measures

The critical theoretical question is this, Which of the three theoretical models best describes the relationships between scores on the BSRI and the PAQ scales and other classes of gender-related attributes, behaviors, and beliefs? The bulk of the literature, we suggest, supports the multifactorial model. Thus, for example, studies of adolescents and adults using a variety of gender-related measures (e.g., recreational interests, various kinds of gender stereotypes, gender attitudes, personality inventories including both the BSRI and PAQ) demonstrate minimal relationships among the various measures (e.g., Ashmore, 1990; Deaux & Lewis, 1984; Orlofsky, 1981; Spence & Buckner, 1995; Signorella, 1992; Spence et al., 1975). These latter results provide evidence for the multifactorial approach to gender and even more particularly, suggest that whereas people's perceptions of their instrumental and expressive characteristics appear to be important in their own right, they cannot bear the weight that theories of masculinity and femininity have put upon them.

Seemingly positive evidence in favor of the single- or the two-factor conceptions does appear from time to time. For example, Bem (1981) reported findings that suggested the greater use of gender schema in memory tasks by participants she classified as gender-typed on the basis of their BSRI scores. However, several attempts to replicate these findings using the same or a similar experimental design have failed to find any significant relationships with either the BSRI or PAQ scales, thus failing to support either the single- or the two-factor approaches. (See Spence, 1991, for a review of these experimental studies.)

A study by Frable (1989), which reported findings supporting Bem's position, should be described in more detail because of its methodological relevance to the present study. Frable (1989) argued that the failures of previous studies (e.g., Bem, 1977; Spence et al., 1975; Spence & Helmreich,

1978a) to find a relationship between measures of gender role attitudes and the BSRI and PAQ were jointly due to the use of undisguised measures of role attitudes and the use of the PAQ instead of the BSRI. Using a measure of gender ideology that she designed to be less obvious than commonly used attitude scales, Frable obtained evidence suggesting that gender-typed men and women tended to score higher on her measure than nongender-typed individuals. In a replication and extension of Frable's study, Spence (1993) employed both the BSRI and the PAQ, along with several measures of gender ideology including Frable's. Significant relationships with the role measures were found that were traceable to the effects of the BSRI items "masculine" and "feminine," men's ratings on "masculine" and women's ratings on "feminine" being positively related to endorsement of traditional ideologies. These results raise the possibility that BSRI findings appearing from time to time that seem to favor the single- or two-factor models may, in fact, not be due to the inventory as a whole but primarily due to specific items that do not assess desirable instrumental and expressive traits.

Although the evidence from adolescents and adults generally favor the multifactorial approach, it cannot be assumed that similar results would necessarily obtain in younger age groups. Children's rigidity in their gender stereotypes and their willingness to behave in accord with them only gradually lessen as they grow older (e.g., Signorella et al., 1993). The separation between self-images of "feminine" expressive traits and "masculine" instrumental traits and between these attributes and other gender-related domains may similarly be a developmental achievement.

PRIOR FINDINGS FROM CHILDREN

Gender Differences in Children's Versions of the PAQ and BSRI

In an investigation in which first-grade boys and girls were administered a highly simplified version of the PAQ (Simms et al., 1978), significant gender differences in the expected direction were found on both scales; further, as in older groups, the two sets of self-ratings were uncorrelated. Investigations of children in the middle elementary school grades given the Hall-Halberstadt (1980) Children's PAQ, or Boldizar's (1991) children's version of the BSRI (Children's SRI) have typically reported similar results. (See, e.g., Boldizar, 1991, and Absi-Semaan, Crombie, & Freeman, 1993, for reviews.)

Relationships With Self-Esteem

Several studies of children have also investigated the relationships between versions of the PAQ or the BSRI and measures of self-esteem (e.g., Allgood-Merton & Stockard, 1991; Alpert-Gillis & Connell, 1989; Boldizar, 1991; Hall & Halberstadt, 1980). The results essentially mirror those found with older groups: in both boys and girls, positive relationships are found between self-esteem and the I or M scales of these instruments, whereas the relationships with the E or F scales are typically weaker or absent.

Instrumentality, Expressiveness, and Other Gender-Related Measures

In contrast, studies using a children's version of the PAQ or shortened versions of the BSRI have found these personality measures to be nonsignificantly related to gender role attitudes (Jones & Costin, 1995), judgments about a target child engaged in gender typical or atypical behavior (Lobel, Bempechat, Gewirtz, Shoken-Topaz, & Bach, 1993), and masculine and feminine toy preferences (Silvern & Katz, 1986; Simms et al., 1978). In an exception to these results, Boldizar (1991) reported that significant correlations occurred between toy preferences and scores on her Children's SRI:M scores were significantly correlated with preferences for masculine activities and F scores were significantly correlated with preferences for feminine activities. The pattern of findings, it should be observed, are supportive of a two-factor theory rather than of unifactorial gender schema theory. Except for the Boldizar (1991) investigation, then, the data are consistent with the multifactorial hypothesis.

It should be noted that in all the studies except Boldizar's, the instrument employed was either a children's version of the PAQ or a short version of the BSRI that appeared to include only desirable instrumental and expressive traits. Boldizar's results with masculine and feminine toy preferences may thus have been heavily influenced by specific items tapping characteristics other than desirable instrumental and expressive traits, and not to properties of the M and F scales as a whole. Likely candidates are the two statements "I like to do things that boys and men do" and "I like to do things that girls and women do" that she substituted for the items "masculine" and "feminine" on the BSRI. It is probable that in both boys and girls, the former statement would be correlated with preferences for masculine activities (concrete examples of things that, stereotypically, boys do) and the latter statement would be correlated with preferences for feminine activities (concrete examples of things that, stereotypically, girls do).

One of the purposes of the present study was to explore this possibility by looking at the relationships between children's preferences for masculine and feminine activities not only with instrumental and expressive traits but also with these and other types of items on the Children's SRI M and F scales.

PRESENT STUDY

In the present study children in the fourth through sixth grades were given a battery of self-report instruments that assessed trait self-concepts, prescriptive gender-linked occupational stereotypes, preferences for masculine and feminine activities, and aspects of self-esteem. The trait, preferences, and stereotype measures were chosen to represent three different kinds of gender-related categories; as such, according to a multifactorial perspective, they should have little or no relationship to one another.

The trait self-concept items were made up of the I and E scales of the Intermediate PAQ (INPAQ; Spence & Helmreich, 1978b), a simplified version of the PAQ suitable for children in the middle elementary school grades, and all the items from Boldizar's children's version of the BSRI (Children's SRI) that do not tap desirable instrumental and expressive traits.[6] Thus, the INPAQ I and E items were substituted for the desirable instrumental and expressive items on Boldizar's inventory. Scores on the combined sets of INPAQ items and the subsets of items from Boldizar's measure were expected to provide a reasonable stand-in for her Children's SRI M and F scales as a whole.

Predictions from the Multifactorial Approach

A number of predictions followed jointly from consideration of the background literature and the implications of our multifactorial approach.

Comparison of the Means of Boys and Girls

The means for boys, in comparison to those for girls, were expected to be significantly higher on the INPAQ I scale and on the M items taken from Boldizar's Children's SRI. Boys were also expected to be significantly

---

[6]This procedure was adopted because the amount of time we were permitted by the schools to conduct the study did not allow us to include both sets of instrumental and expressive items in the test battery and we were interested in determining the utility of the INPAQ in research with children.

lower on the INPAQ E scale and on the F items from the Children's SRI. Similarly, boys' preferences for masculine activities were expected to be significantly higher and their preferences for feminine activities were expected to be significantly lower than those for girls. As males generally exhibit conventional gender ideologies to a greater degree than females (e.g., Spence & Helmreich, 1978a), any gender differences in prescriptive occupational stereotypes were expected to reveal greater endorsement by boys. There was no firm basis on which to predict gender differences in self-esteem.

In describing our remaining predictions, we make no reference to the children's gender because we expected similar results from boys and girls.

Correlations Between the INPAQ Scales and Children's SRI Items

Scores on the I and E items were expected to be nonsignificantly correlated. A strict application of the multifactorial view led us to predict that scores on the Boldizar items "I like to do things that girls and women do" and "I like to do things that boys and men do" would not only be unrelated to each other but would also be unrelated to the INPAQ I and E scales.

Finally, previous research (e.g., Aube et al., 1995; Spence et al., 1979) led us to suspect that the undesirable items on the Children's SRI F scale (which we classified as Other F) would be negatively associated with the I items but at the same time, positively associated with the E items.

Relationships with Self-Esteem Scales

Self-esteem was measured by three of Harter's (1985) scales. We expected that the findings would parallel those found in previous studies of both children and older groups, namely that in both genders, scores on the self-esteem scales would be positively correlated with the INPAQ I scale but at best would be minimally correlated with the E scale. On the other hand, we predicted that the largely undesirable Children's SRI items classified as Other F would tend to be negatively correlated with self-esteem.

Our assumption that the relationships between self-esteem and socially desirable instrumental traits and undesirable expressive traits are attributable to these traits per se led us to predict that none of the remaining gender-related measures would be positively associated with self-esteem.[7]

---

[7]The one exception concerns two of the three M items taken from Boldizar's Children's SRI. This pair of items (which we label Other M) are versions of the BSRI items analytic and athletic. Based on a prior investigation with college students (Spence, 1993), we suspected that these two items would be correlated with the INPAQ I items in both genders and like

Correlations Between the Gender-Related Measures

The multifactorial approach to gender is not confined to propositions about personality traits but also bears on the associations among other classes of gender phenomena.

Considering first prescriptive occupational stereotypes, we expected that both boys and girls would rate traditionally masculine occupations as more appropriate for men than women and traditionally feminine occupations as more appropriate for women and that these measures of masculine and feminine stereotypes would be positively correlated. The prediction of theoretical significance, however, is that scores on the stereotype measures would be nonsignificantly related in both boys and girls not only to their I and E scores but also to their scores on the Activity Preferences measure.

We also predicted that Activity Preferences scores would be significantly correlated in both genders with the Boldizar items "I like to do things girls and women do" and "I like to do things boys and men do" because of their similar content. Application of the multifactorial perspective led us to expect, however, that the two Activity measures, masculine and feminine, would not be significantly related to one another. Nor did we expected them to be related to INPAQ I and E scores or to Boldizar's Other F items. We did anticipate, however, that the two Boldizar items classified as Other M (the equivalents of the BSRI items athletic and analytic) would be related to masculine preferences.

Certain of the predictions outlined above, e.g., mean differences between boys and girls, also follow from the unifactorial and two-factor theories. Only their unique predictions are outlined below.

Predictions from the Unifactorial and Two-Factor Theories

The two-factor model leads to the prediction that in both boys and girls, total M scores (and, presumably, each of the components contributing to them) would be significantly related to preferences for masculine (but not feminine) activities and that total F scores would be significantly related to preferences for feminine (but not masculine) activities. Indeed, this is what Boldizar (1991) reported in her study of children given the Children's SRI. Parallel predictions would appear to follow for masculine and feminine prescriptive occupational stereotypes.

The single-factor theory, on the other hand, does anticipate that masculine and feminine occupational stereotypes are correlated. The the-

---

the latter, with self-esteem. We also expected that in both boys and girls, this pair of items might be related to the items assessing preference for masculine activities.

ory further predicts that gender-typed boys, those with high M and low F scores, and gender-typed girls, those with low M scores and high F scores, would be more stereotyped on the occupational measures than nongender-typed boys and girls, those whose scores on the M and F scales are more nearly equal, whatever their absolute level. In statistical terms, this theory predicts a complex interaction between M and F scores and the criterion variables, an interaction that differs for males and for females. What the theory would predict about the preferences for masculine and feminine activities of those falling into the various gender-typing groups is less clear, but presumably they would be the same as for the occupational stereotypes.

METHOD

Participants

The participants were 197 boys and 271 girls ranging in age from approximately 9 to 12 years who were enrolled in fourth-, fifth-, or sixth-grade classes in schools in Galveston County, Texas. The school district serves an area whose residents are almost exclusively white and in socio-economic status are largely middle to upper middle class. In each grade, the children who took part were all those whose parents had returned a signed consent form, were present on the day their class was tested, and agreed to take part. Participation rates varied across classrooms (but not across grades or gender) from close to 60% to close to 100%. The variability appeared to be a joint function of the teachers' stress that the children in their class should take consent forms home and bring signed forms back to the teacher, and the absentee rate on the particular day on which testing took place. No child with a signed consent form refused to take part.

Measures

Self-Report Trait Measure

The trait measure consisted of 28 intermingled items from the Intermediate PAQ (INPAQ; Spence & Helmreich, 1978b) and selected items from the Children's Sex Role Inventory (Children's SRI; Boldizar, 1991). Each item was accompanied by a 5-point Likert scale whose extremes were labeled *Very Much Like Me* and *Not at All Like Me*. Item scores ranged

from 0 to 4 and were coded such that high scores indicated the presence of the masculine or feminine characteristic.

   Intermediate Personal Attributes Questionnaire. As in the PAQ, the IN-PAQ consists of an 8-item instrumental (I) scale and an 8-item expressive (E) scale. Each item appears in the form of a statement directly or indirectly describing the presence (or the absence) of the instrumental trait in a parallel PAQ I scale item or of the expressive trait in a parallel E scale item. The items on the two INPAQ scales are shown in Table I. In college students, the INPAQ has been shown to be highly correlated with the original PAQ and to have very similar psychometric properties to it, including similar means and standard deviations for each scale for each sex (Spence & Helmreich, 1978b).

   In the present sample, the Cronbach alphas for the E scale were .80 and .71 for boys and girls, respectively. Parallel figures for the I scale were .46 and .43. These disappointing outcomes for the I scale are to be contrasted with the substantially higher alphas found with the same items in older groups and also with the longer scale on the Hall-Halberstatdt Chil-

Table I. Items on the Instrumental (I) and Expressive (E) Scales on the INPAQ[a]

Expressive scale
   I am very emotional. (That means my feelings get stirred up easily.)
   I really like to do things for other people.
   I am very gentle.
   I am very helpful to other people.
   I don't pay much attention to how other people are feeling. (R)
   I am very kind to other people.
   I try to understand how others are feeling.
   I am a very warm, friendly person.

Instrumental scale
   I am able to do tough things by myself if I have to and I don't need other
      people to help me or tell me what to do.
   I am very busy and active.
   I enjoy trying to win games and contests.
   When I have to decide about something important, it's hard for me to
      make up my mind. (R)
   I feel sure I can do most of the things I try.
   I am better at doing things than other people.
   When I'm in a tough spot, I get very bothered and don't know what to
   do. (R)
   I give up very easily. (R)

---

[a]R denotes items that are reverse scored.

dren's PAQ.[8] These findings indicate that the INPAQ I scale is not reliable enough in its present form for assessing individual children in this age group. It may nonetheless be useful for theoretical purposes when group data are being analyzed. This utility can be demonstrated—and in fact, was demonstrated by the present results—by showing that the pattern of relationships into which I scores enter are congruent with the previous literature and with theoretical predictions.

Children's Sex Role Inventory Items. On the BSRI's 20-item M and F scales, 17 items can be classified as describing desirable instrumental attributes and 11 items can be classified as describing desirable expressive attributes (Spence, 1991). The INPAQ I and E scales were substituted for Boldizar's versions of these items on the Children's SRI. As might be expected, items on the latter are similar in content to statements on the INPAQ. For example. "I am very gentle" and "I enjoy trying to win games and contests" (items from the INPAQ E and I scales, respectively) are highly similar to "I am a gentle person" and "When I play games I really like to win" (items from the Children's SRI F and M scales, respectively).

All other items from the Children's SRI, that is, the 3 M items and the 9 F items that do not describe socially desirable instrumental and expressive traits, were included in the present study. (The complete set of Children's SRI items may be found in Boldizar, 1991.) Two of the three items taken from the M scale of Boldizar's measure corresponded to the BSRI M items athletic and analytical (here labeled Other M), and the third is the statement "I like to do things that boys and men do" (here labeled Masculine) which replaced the adjective "masculine" on the BSRI.

The F scale items taken from Boldizar are the statement "I like to do things girls and women do" (here labeled Feminine) intended to replace the adjective "feminine" on the BSRI, and statements corresponding to the BSRI items gullible, childlike, does not use harsh language, yielding, soft-spoken, loves children, shy, and flatterable (here labeled Other F). The number of Other F items made it reasonable to determine their reliability. Cronbach alphas turned out to be .63 and .47 for boys and girls, respectively. Although these items were not intended by Bem or by Boldizar to constitute a separate subscale, their alphas were sufficient for us to treat them as such for purposes of data analysis.

---

[8]Doubling the length of the 8-item I scale would make the number of items approximately equal to the number of items on the Hall-Halberstadt Children's PAQ and also to the number of instrumental items on Boldizar's M scale. Application of the Spearman–Brown prophecy formula in which the number of I items was assumed to be 16 increased the alphas to the low .60's for both boys and girls.

Activity Preferences

The measure consisted of a list of 24 activities, most adopted from Yekel, Bigler, and Liben (1991). Each item was rated on a 5-point scale ranging from *Dislike a Lot* to *Like a Lot*. The activities were equally divided into those stereotypically associated with boys (masculine), with girls (feminine), and equally with both sexes (neutral). We selected the items within the categories to reflect a range of types of activities. Based on pilot work, gender items were paired in terms of the kind of activity listed (e.g., a masculine and a feminine physically active game or sport, quiet game, and TV program). This was done in an attempt to maximize preferences based on gender stereotypes per se, unconfounded by differences between the masculine and feminine lists in other properties. Post hoc analyses confirmed assignments to gender category, significant gender differences in the predicted direction emerging for every gender-related item. Cronbach alphas for the masculine items were .70 for both sexes and for the feminine items were .74 and .67 for boys and girls, respectively.

Prescriptive Occupational Stereotypes

Sixteen occupations, taken from Yekel et al. (1991) and from previous pilot work, were listed. Because prescriptive rather than descriptive stereotypes were being elicited, the instructions emphasized that the children were to rate who *should* do each job and that there were no right or wrong answers. Each item was rated on a 5-point scale with the labels *Only Men, Mostly Men, Both Men and Women, Mostly Women,* and *Only Women.* Six of the occupations were masculine according to both descriptive and prescriptive stereotypes, 6 were stereotypically feminine, and 4 were gender neutral. Post hoc analyses of the means of each item were determined for both boys and girls and in each case confirmed the assignment of item to category. Alphas for the Masculine occupations were .62, and .61 and for the Feminine occupations were .57 and .58 for boys and girls, respectively.

Harter Self-Perception Profile for Children (SPPCS)

The Harter SPPCS (Harter, 1985) contains six 6-item scales tapping different aspects of self-esteem. Because of time limitations, only three of these scales were administered. Two of the scales, Scholastic Competence and Social Acceptance, were chosen because they represented different ar-

eas of functioning of major significance in children's lives, whereas the third, Global Self-Worth, was chosen as a measure of overall self-esteem that transcended specifics.

All items are set up in a format in which two contrary statements appear (e.g., "Some kids are often unhappy with themselves" and "Other kids are pretty pleased with themselves"); below each of the statements are the two response phrases, *A lot like me* and *Sort of like me*. The child is instructed first to choose one of the two statements and then to indicate his or her choice of the two responses by circling one of the alternatives. Cronbach alphas ranged from .77 to .82.

Survey Booklet

The measures were administered in a printed booklet. On the first page, the children were asked to supply identifying information that included their gender, birth date, grade, and teacher. The objective questionnaires then appeared, each starting on a separate page. The labels in the booklet given to each measure and the order in which they appeared are as follows: *About Me: How I Am* (self-report trait items), *About Me: What I Like to Do* (activity preferences), *Who Should Do These Jobs?* (occupational stereotypes), and *More About Me* (Harter scales). All responses were entered directly onto the booklet. The booklets were destroyed after the data were entered into the computer by code number.

Procedure

The children who participated in each class were tested during a single session conducted by several female undergraduates. Each session was approximately 55–60 minutes long.

After giving preliminary instructions and instructions for the first measure in the survey booklet, the person in charge read each item aloud, pausing for the children to respond. As the children worked, her assistants unobtrusively checked the children's understanding of what they were supposed to do. After all the children had completed the first measure, the second measure was explained and the items read aloud. This procedure was repeated until all measures were completed. The children were thanked for their participation and then returned to their regular schoolwork.

RESULTS

For analyses involving comparison of means, analyses of variance (ANOVAs) that took grade as well as gender into account were first conducted. These analyses, supplemented by inspection of the means, indicated that the direction of the differences between boys and girls were consistent across grades and that there were no systematic trends across grade levels. Informal inspection of the correlations between measures also revealed no discrepancies across grades worthy of note. For this reason, the data from the several grades were combined and the results reported here were based on the total group of each gender.

The means of the measures for boys and for girls and the results of the ANOVAs comparing the genders are reported in Table II. The correlations between the measures for boys and for girls are reported in Table III. Because of the large number of correlations that were computed, only r's with p values of .01 or less were considered significant and included in Table III.

Our presentation of the results, below, correspond to the sets of predictions from the multifactorial approach, presented earlier.

Comparison of the Means of Boys and Girls

Inspection of Table II indicates that the pattern of significant and nonsignificant differences between the means of boys and girls on the various measures confirmed our predictions as well as the results of prior studies. Thus, in both genders, significant differences between means in the expected direction were found for the INPAQ I and E scales and also for each of the subsets of items drawn from the Children's SRI. Also as expected, boys preferred masculine activities significantly more and feminine activities significantly less than girls. A 2 (gender) × 2 (masculine or feminine activities) between-subject within-subject ANOVA indicated that the margin by which boys preferred masculine activities over feminine ones was significantly ($p < .001$) larger than that by which girls preferred feminine over masculine activities. Prior studies (e.g., Schau, Kahn, Diepold, & Cherry, 1980) have also demonstrated that boys are more gender-typed than girls in their interests.

It can be seen in Table II that the ratings of both boys and girls were in the stereotyped direction for both masculine and feminine occupations and in Table III that the correlations between the masculine and feminine categories were moderately high. Significant gender differences occurred for both masculine and feminine occupations: girls endorsed both stereo-

Table II. Scale Means for Boys (n = 197) and Girls (n = 273) and ANOVA for Sex[a]

		No. Items	Means		ANOVA	
			Boys	Girls	F	p
INPAQ	I	8	20.75	19.18	20.62	.0000
			(3.99)	(3.54)		
	E	8	20.33	23.90	54.30	.0000
			(6.24)	(4.42)		
CSRI	OM	2	6.57	4.31	3.77	.0001
			(1.71)	(1.91)		
	OF	8	17.33	20.15	45.16	.0001
			(5.56)	(4.77)		
	Masc	1	3.28	2.15	272.69	.0000
			(0.49)	(1.30)		
	Fem	1	1.28	3.14	871.06	.0000
			(0.69)	(0.98)		
Harter	SC	6	17.65	17.31	<1	ns
			(4.67)	(4.41)		
	SA	6	17.42	17.75	<1	ns
			(4.36)	(4.49)		
	GL	6	19.32	18.96	<1	ns
			(4.22)	(4.23)		
Act	M	8	33.73	23.50	349.49	.0000
			(5.40)	(6.13)		
	F	8	13.34	24.73	402.60	.0000
			(5.06)	(6.62)		
	Nu	8	31.68	33.35	16.07	.0001
			(4.82)	(4.17)		
Occup	M	6	13.29	14.35	19.17	.0000
			(2.92)	(2.39)		
	F	6	23.57	22.95	6.08	.01
			(2.85)	(2.59)		
	Nu	4	11.81	11.29	<1	ns
			(1.51)	(1.08)		

[a]INPAQ: Intermediate PAQ; I: instrumental; E: expressive; CSRI: Children's SRI: OM: Other Masculine; OF: Other Feminine; Masc: CSRI equivalent to BSRI "masculine"; Fem: CSRI equivalent to BSRI "feminine"; Harter SC: Scholastic Competence; SA: Social Acceptance, GL: Global Self-Worth, Total: total of subscales; Act: Activity Preferences; M: masculine; F: feminine; Nu: neutral; Occup: Occupational Stereotypes. On the Stereotype measure, high scores indicated a choice of women over men. Standard deviations are in parentheses.

types to a significantly lesser degree than boys, an outcome congruent with the general finding that males are more bound to traditional gender ideologies than females (e.g., Spence & Helmreich, 1978a).

Finally, the means of boys and girls did not differ on the three Harter self-esteem scales, thus confirming the majority of previous findings.

Table III. Significant Correlations ($p < .01$) Between the Scales for Boys ($n = 197$) and Girls ($n = 273$)[a]

Scale	INPAQ		CRSI				Harter			Activities			Occup		
	I	E	OM	OF	Masc	Fem	SC	SA	GL	M	F	Nu	M	F	Nu
INPAQ															
I	—		31	-29	25	-16	26	30	27		37	20			
E		—	28	44	20	21									
CSRI															
OM	33	27	—		-16	21	24	31	18	26	23	16	26		
OF	18	61	20	—	26			-20		41				-18	
Masc			34	26	—					25					
Fem	-18		21			—	41	57			40	22	25		
Harter															
SC	42		27			41	—	47	54				19		
SA	35		26	-20		57	47	—	55		-28		-22		
Global	35		32	-19			52		—				23		
Act															
M		28	25	21	25					—		51	30		
F			23	20		29					—	32	32	-17	
Nu		38	16	36		22				51	32	—	19	-19	
Occ															
M	-18		26			25	19	-22	23	22	19	19	—		
F											22	-19		—	
Nu													-51		—

[a]Decimals are omitted. Girls are above the diagonal and boys are below. INPAQ I = Instrumental, E = Expressive; CSRI = Children's SRI; OM = other masculine, OF = other feminine; Masc = CSRI equivalent to BSRI "masculine"; Fem = CSRI equivalent to BSRI "feminine"; Harter SC = Scholastic Competence, SA = Social Acceptance, GL = Global Self-Worth; Act = Activities, M = masculine, F = feminine, Nu = neutral; Occ = Occupational stereotypes. M = masculine, F = feminine, Nu = neutral. On the stereotype measure, high scores indicate choice of women over men. For $df = 271$, $p_{.01} = .15$, $p_{.001} = .22$. For $df = 195$, $p_{.01} = .18$, $p_{.001} = .24$.

Correlations Between the INPAQ and Children's SRI Scales

As may be seen in Table III, INPAQ I and E scores failed to be significantly correlated, as has been found in prior investigations of both children and adults. However, I scores were positively related in both genders to the pair of Children's SRI items classified as Other M (items corresponding to the BSRI items analytic and athletic) and were negatively related to the largely undesirable traits classified as Other F. The latter, however, were significantly correlated with INPAQ E scores. These results, which mirrored those found with college students and the BSRI (Spence, 1993), confirmed our predictions.

Of special interest were the correlations involving the pair of novel items on the Children's SRI, namely "I like to do things that girls and women do" and "I like to do things that boys and men do." As we predicted, these items (here labeled Children's SRI Masculine and Feminine) were nonsignificantly related to one another. Stringent application of the multifactorial model also led us to predict that the Masculine and Feminine items would be nonsignificantly related to the other sets of personality items, most importantly to the I and E items on the INPAQ. This prediction was not uniformly confirmed. Thus in both genders, INPAQ I scores were positively related to the Masculine item, "I like to do things that boys and men do," and negatively related to the Feminine item, "I like to do things girls and women do." However, in the case of boys and the Masculine item, the correlation did not reach the .01 level. In addition, in both genders, the Children's SRI Other F items (and in the case of girls, the INPAQ E items) were significantly related to the Feminine item "I like to do things girls and women do." In girls, there was also a significant negative relationship between the Other F items and the Masculine item.

Although these significant findings could be interpreted in terms of broad gender concepts, it also seems plausible that children believe activities associated with males are active, exciting, and adventuresome, and those associated with females are more passive. (As others have noted, e.g., Serbin et al., 1993, these perceptions have a basis in reality.) Boys and girls high in instrumental characteristics may therefore tend to be attracted to the former and have less interest in the latter. Similarly, those high in the characteristics tapped by the Other F items may be more attracted to those they perceive as less active and challenging. These speculations gain some support when the data from the activity preferences measure are examined. As is discussed below, few significant correlations in the predicted direction occurred, none involving the I scale items.

Relationships with Self-Esteem Scales

As in previous studies of self-esteem, I scores (and Other M scores) were significantly related in both sexes to the Harter self-esteem scales whereas relationships with E scores were not. Congruent with our previous results for undesirable feminine traits in older groups (Spence et al., 1979), significantly negative correlations were found between the Other F items and the Harter Social Acceptance (SA) scale.

We have proposed that these relationships with self-esteem are due to the effects of desirable and undesirable instrumental and expressive personality traits in and of themselves; self-esteem should therefore not be related to other kinds of gender-related categories. With several isolated exceptions, our results confirm this expectation. None of the significant correlations that emerged, it should be noted, were found in both genders or with all of the self-esteem measures. In light of the large number of correlations that were computed, it is probably wise not to make too much of these outcomes in the absence of replication.

Correlations Between the Gender-Related Measures

Activity Preferences

Looking first at the masculine and feminine Activity Preferences scores, as we had predicted, the two categories of preferences were not correlated with one another in either gender. However, both masculine and feminine preferences were significantly related to neutral preferences. As discussed below, these findings appear to be primarily determined by scores on two of the personality measures, the INPAQ E scale and the Children's SRI items classified as Other F.

Turning to the personality measures, we first consider the Children's SRI Masculine and Feminine items ("I like to do things that boys and men do" and "I like to do things that girls and women do"). We predicted that because of their overlapping content, in both boys and girls masculine Activity Preferences would be significantly related to the Children's SRI Masculine item (but not the Feminine item) and similarly, feminine Activity Preferences would be significantly related to the Children's SRI Feminine item (but not the Masculine item). These predictions were confirmed. Thus our suspicion was verified that these items were at least partially responsible for the Boldizar (1991) findings that her M and F scales were related to children's preferences for masculine and feminine toys, respectively.

As for the other personality items, INPAQ I scores in both genders were significantly correlated neither with masculine preferences nor with feminine preferences, as expected by our multifactorial approach. However, several unexpected correlations were found between Activity Preferences scores and the E and Other F scores. In boys, both sets of scores were positively related to all three preferences measures, masculine, feminine, and neutral (but nonsignificantly in the case of E and the feminine measure), suggesting that a generalized acquiescence bias was involved. Further supporting this suggestion, the highest correlation with E and with Other F was with neutral preferences. Similar results were found in girls: both E scores and Other F scores were significantly correlated with feminine preferences and to a somewhat lesser extent, with neutral preferences. In an effort to reduce the positivity bias, a series of partial correlations were computed for each gender. Partial r's were determined between each of the two personality measures (E and Other F) and each of the two preference scores (masculine and feminine) with neutral scores controlled and then between each of the personality measures and neutral preferences with masculine and feminine preferences controlled. In all instances, the correlations dropped to nonsignificance (p's $> .01$). The intrusion of an unexpected response bias into the results makes it difficult to draw conclusions about the meaning of these specific findings.

When considered as a whole, however, the activity preferences data were more supportive of the multifactorial perspective than the alternative models.

Prescriptive Occupational Stereotypes

As already noted, both boys and girls exhibited prescriptive stereotypes about the occupations appropriate for men and for women; furthermore, the scores for masculine and feminine stereotypes were moderately correlated. (Stereotype scores were all coded in the same direction, high scores indicating more appropriateness for women.)

We anticipated that the stereotype measures would be unrelated to preferences for masculine and feminine activities and also to the various personality measures. By and large, this is what occurred but once more, several unexpected correlations emerged. In both genders, those who scored low on the feminine Activity Preferences measure were more likely to exhibit stereotypic beliefs about masculine occupations than those who scored high on the feminine Activity Preferences measure. In boys, two additional findings occurred: high I scores were associated with stronger stereotypes about masculine occupations, whereas high scores on the Chil-

dren's SRI Feminine item ("I like to do things that girls and women do") were associated with weaker stereotypes on this measure.

Additional Analyses

Tests of the single-factor and the two-factor models are most appropriately based on the total M scores (INPAQ I + Children's SRI Other M items and the Masculine item) and total F scores (INPAQ + Children's SRI Other F items and the Feminine item).

Correlations were first determined for each gender between the masculine and the feminine preference measures and total M and total F scores. The two-factor approach leads to the expectation that for both boys and girls, significant correlations would be found between masculine preferences and M scores and between feminine preferences and total F scores. The former prediction was not upheld, the correlations for neither gender reaching significance. On the other hand, the latter prediction was confirmed, Total F scores being significantly related to Feminine preferences (r's of .17 and .31 for boys and girls, respectively, p's < .01). Thus, there was partial support for the two-factor theory (as well as partial replication of Boldizar's, 1991, toy preference results). Inspection of the correlations for the individual components of the total scores indicate that all three components making up the total F score were positively correlated in both sexes with feminine preferences. As may be seen in Table III, for girls the E items made the strongest contribution. For boys, the correlation with the E items was positive but nonsignificant ($r = .12$), the significant effect being more influenced by the set of Other F items.

The two-factor model would seem to lead to the somewhat implausible prediction that total M scores would be related to masculine (but not feminine) occupational stereotypes and total F scores would be related to feminine (but not masculine) occupational stereotypes. None of the correlations, however, was significant.

In contrast, Bem's unifactorial theory implies that masculine and feminine occupational stereotypes should be correlated; as we also predicted, this turned out to be the case. The theory also predicts that interactions would be found between total M and F scores and each of the stereotype measures such that gender-typed boys (high M and low F) and gender-typed girls (high F and low M) would exhibit more stereotyping than those whose M and F scores were more nearly equal. To test this hypothesis, a series of regressions analyses were conducted for each gender, separately for the masculine and the feminine stereotypes. In one set of these analyses, the interaction term (M × F) was entered first, followed by terms for M

and for F. In the second set, M and F were entered first, followed by the M × F interaction term. In the event that an interaction turned out to be significant, it could be decomposed to determine whether it corresponded to Bem's predictions. However, in no instance did the interaction term even approach significance (p's > .05). For this reason, the results of these analyses are not tabled.

It was unclear what the unifactorial theory would predict about masculine and feminine activity preferences. However, the same kind of regression analyses as described above were conducted with the preferences data; again none of the interactions terms approached significance.

DISCUSSION

Evidence for the Multifactorial Approach

The BSRI and PAQ, along with children's versions of these instruments, continue to be used frequently in gender research, as measures of masculinity and femininity, gender role orientation, or other similar concepts. The theoretical presumptions guiding these studies reflect what we have identified as the single-factor or the two-factor models.

In contrast, we have endorsed an alternative perspective that basically states that gender-differentiating phenomena are not only multidimensional (as is readily apparent and not a matter of dispute) but multifactorial, the different factors having variable and complex relationships with one another. We have more specifically proposed that although instrumental and expressive characteristics have important implications for both genders, these trait clusters tend not to be correlated with other gender-related characteristics and behaviors simply because they are gender related. Thus, we assert, personality measures such as the BSRI and PAQ do not tap broad gender constructs such as masculinity and femininity. Our reading of the literature based on samples of college age and older given the BSRI and PAQ suggests that the preponderance of evidence supports this view (e.g., Ashmore, 1990; Spence, 1991).

Fewer studies have been conducted with children. The major purpose of the present study was to investigate the possibility that at least in children in the middle elementary school grades (i.e., children old enough to have acquired a firm sense of gender identity and a knowledge of societal stereotypes), the multifactorial approach to gender phenomena has more utility than models assuming that these phenomena can be subsumed under a single or at most, two superordinate constructs.

The predictions based on the multifactorial perspective that we put forward, namely that INPAQ I and E scales would be unrelated to children's preferences for masculine and feminine activities and to prescriptive occupational stereotypes, were largely confirmed whereas those based on the one- and two-factor models were not. Although our focus has been on instrumental and expressive attributes, application of the multifactorial perspective led us to propose that other classes of gender-relevant attitudes and behaviors also tend to be unrelated. Our data from the activity preferences and occupational stereotypes measures are congruent with these expectations. The results of a recent investigation by Bigler, Liben, Lobliner, and Yekel (1996) in which children were asked to rate both themselves and males and females in general in multiple domains also support the multifactorial view. When these data are combined with those from older groups, we conclude that at least from middle childhood on, various classes of gender-related phenomena tend to be independent.

It should be acknowledged, however, that our results were not perfectly clearcut: although they seldom were found in both boys and girls, instances occurred of significant correlations across domains that we had not predicted. The possibility that these scattered findings (which were not uniformly consistent with either the single-factor or the two-factor models) are due only to chance cannot be dismissed. On the other hand, at least some of them may turn out to be replicable and reflect complexities that we have yet to appreciate. We should point out in this connection that it is not our view that gender-related measures are never related to one another. The evidence already indicates that sometimes they are (e.g., trait stereotypes have been shown to be correlated with gender role attitudes, Spence et al., 1975; in the current study, stereotypes about masculine and feminine occupations are also related). The multifactorial approach is not yet a well-developed model and states only that categories of gender-related attributes, beliefs, and behaviors typically contribute to separate factors whose relationships to other factors are variable in magnitude, even though often close to zero, and are often complex.

Significance of the Findings with Self-Esteem

In confirmation of prior results with children (Hall & Halberstadt, 1980; Boldizar, 1991) as well as with adolescents and adults (Whitley, 1983), we found that the self-esteem measures were significantly correlated with the INPAQ I scale but not the E scale. We have proposed that these findings are more validly interpreted in terms of desirable instrumental and expressive attributes per se than in terms of gender constructs. This con-

tention implies that the relationship between the self-esteem measures and the INPAQ scales would not be duplicated with the activity preferences and stereotype measures. This expectation was also confirmed. We thus conclude that the findings obtained with measures of self-esteem and other measures of psychological adjustment yield information about the implication of instrumental and expressive attributes for psychological functioning but are irrelevant to questions about the BSRI and PAQ as measures of the theoretical concepts of masculinity, femininity, and gender typing.

The Confusion of Personality and Gender Concepts

Our interpretation of the self-esteem data illustrate a point that has plagued studies using the BSRI and PAQ: the frequent failure to distinguish between outcomes obtained with these measures that might be attributable to instrumentality and expressiveness per se from those that might be attributable to gender constructs. The problem arises when the actual content of the scales on the two questionnaires is ignored and it is automatically assumed that they are measures of hypothetical gender concepts. In some cases, investigators at least tacitly base their hypotheses on the implications of instrumentality and expressive characteristics for other attributes and behaviors but nonetheless interpret their findings in terms of gender role identification, gender schematization, or masculinity and femininity. In other cases, investigators have employed novel tasks for which little rationale has been provided and conclude solely on the basis of significant relationships with the BSRI or PAQ that something has been learned about abstract gender concepts.[9] Adding to the confusion, investigators often fail to distinguish between the single-factor and two-factor theories and which of these quite different models they endorse and their significant results presumably support. Awareness of the multifactorial perspective at least has the virtue of creating awareness that the PAQ and BSRI are valid measures of instrumentality and expressiveness and that studies should not be designed nor data interpreted on uncritical acceptance of the assumption that these instruments assess something beyond these trait clusters.

[9]The converse of this argument, with which we agree, has been put forward by Bem (1985), namely, that the substantial number of investigators reporting nonsignificant results that they claim fail to support her views employ situations or tasks that have dubious relevance to her theory (or often, we should add, to any other theory).

The BSRI vs. the PAQ

Is there any evidence to support the claim (e.g., Boldizar, 1991; Frable & Bem, 1985; Frable, 1989; O'Heron & Orlofsky, 1990) that the BSRI and the Children's SRI have validity as measures of global gender constructs whereas the various versions of the PAQ measure only desirable I and E traits? Put another way, does the fact that the BSRI is more mixed in content than the PAQ and has a multifactorial structure lead the BSRI to being superior to the PAQ (if not qualitatively different) as a measure of gender-typing constructs? An affirmative answer would have significant implications for the conduct of future gender research as well as the interpretation of extant data involving these instruments.

Data from the present study, combined with the results from previous studies of both children and older groups (e.g., Lobel et al., 1993; Spence, 1993), suggest that this is not the case. As our results illustrate, the different types of items going to make up the M and F scales of the BSRI and of the Children's SRI typically do not act in concert. Rather the several components of these scales are not uniformly related to other variables, including measures of other kinds of gender-related phenomena, and when significant relationships are found, the overall results do not lend themselves to any simple proposition about masculinity or femininity or gender schema. (From the point of view of global gender concepts, the most interesting and relevant items are the pair of adjectives "masculine" and "feminine" found on the original BSRI.) The use of overall M and F scores is thus likely to obscure the contribution that each component makes to the specific criterion under investigation. Investigators who elect to use the BSRI as a measure of gender constructs not only have the obligation to make their theoretical assumptions explicit but (as with any self-report measure with a complex factorial structure) also should analyze their data to reflect the influence of the components of the M and F scales as well as the total scores.

Conclusion

Our data from school-aged children, added to the results of previous studies of both children and adults, suggest that the instrumental and expressive trait dimensions tapped by the PAQ and BSRI do not contribute uniquely to our understanding of gender and gender roles; they are but one class of gender-related variables. More generally, the bulk of the evidence supports the multifactorial perspective rather than either the unifactorial and the two-factors models. A major implication of this conclusion

is that masculinity and femininity and related constructs such as gender role identity and gender schematization, as they traditionally have been conceived, lack scientific merit. Acceptance of the multifactorial approach should help lead the way to the development of more complex, dynamic theories of gender. Fortunately, some beginning steps in this direction are already being made (e.g., Biernat, 1991; Serbin et al., 1993; Spence, 1985).

REFERENCES

Abraham, B., Feldman, S. S., & Nash, S. C. (1978). Sex-role self-concept and sex-role attitudes: Enduring personality characteristics or adaptations to changing life situations? *Developmental Psychology, 14,* 393-400.

Absi-Semaan, N., Crombie, G., & Freeman, C. (1993). Masculinity and femininity in middle childhood: Developments and factor analyses. *Sex Roles, 28,* 187-206.

Allgood-Merton, B., & Stockard, J. (1991). Sex role identity and self-esteem: A comparison of children and adolescents. *Sex Roles, 25,* 129-139.

Alpert-Gillis, L. J., & Connell, J. P. (1989). Gender and sex-role influences on children's self-esteem. *Journal of Personality, 57,* 98-114.

Antill, J. K., & Cunningham, J. D. (1979). Self-esteem as a function of masculinity in both sexes. *Journal of Consulting and Clinical Psychology, 47,* 783-785.

Ashmore, R. D. (1990). Sex, gender, and the individual. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 486-526). New York: Guilford.

Aube. J., Norcliffe, H., & Koestner, R. (1995). Physical characteristics and the multifactorial approach to gender characteristics. *Social Behavior and Personality, 23,* 69-82.

Aube, J., Norcliffe, H., Craig, J.-A., & Koestner, R. (1995). Gender characteristics and adjustment-related outcomes: Questioning the masculinity model. *Personality and Social Psychology Bulletin, 21,* 284-295.

Bakan, D. (1966). *The duality of human existence.* Chicago: Rand McNally.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42,* 155-162.

Bem, S. L. (1977). On the utility of alternate procedures for assessing psychological androgyny. *Journal of Personality and Social Psychology, 31,* 634-643.

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88,* 354-364.

Bem, S. L. (1985). Androgyny and gender schema theory: A conceptual and empirical integration. In T. B. Sonderegger (Ed.), *Psychology and gender: Nebraska Symposium on Motivation.* Lincoln: University of Nebraska Press.

Bem, S. L. (1993). *The lenses of gender: Transforming the debate on sexual inequality.* New Haven, CT: Yale University Press.

Biernat, M. A. (1991). A multicomponent, developmental analysis of sex-typing. *Sex Roles, 24,* 567-586.

Bigler, R. S., Liben, L. S., Lobliner, D. B., & Yekel, C. A. (1996). *The structure of gender schemata: Relations among multifactorial measures in children and adults.* Manuscript submitted for publication.

Boldizar, J. P. (1991). Assessing sex typing and androgyny in children: The Children's Sex Role Inventory. *Developmental Psychology, 27,* 505-513.

Constantinople, A. (1973). Masculinity–femininity: An exception to the famous dictum? *Psychological Bulletin, 80,* 389-407.

Deaux, K., & Lewis, L. (1984). Structure of gender stereotypes: Interrelationships among components and gender label. *Journal of Personality and Social Psychology, 499,* 991-1004.

Downs, C. A., & Langlois, J. H. (1988). Sex typing: Construct and measurement issues. *Sex Roles, 18,* 87-100.

Edwards, V. J., & Spence, J. T. (1987). Gender-related traits, stereotypes, and schemata. *Journal of Personality and Social Psychology, 53,* 146-154.

Erikson, E. (1964). Inner and outer space: Reflection on womanhood. *Daedalus, 93,* 1-15.

Frable, D. E. S. (1989). Sex typing and gender ideology: Two facets of the individual's gender psychology that go together. *Journal of Personality and Social Psychology, 56,* 95-108.

Frable, D. E. S., & Bem, S. L. (1985). If you are not gender aschematic, all members of the opposite sex look alike. *Journal of Personality and Social Psychology, 49,* 459-468.

Hall, J. A., & Halberstadt, A. G. (1980). Masculinity and femininity in children: Development of the Children's Personal Attributes Questionnaire, *Developmental Psychology, 16,* 270-280.

Harter, S. L. (1985). *Manual for the Self-Perception Profile for Children.* Denver: University of Denver.

Helmreich, R. L., Spence, J. T., & Wilhelm, J. A. (1981). A psychometric analysis of the Personal Attributes Questionnaire. *Sex Roles, 7,* 1097-1108.

Jones, D. C., & Costin, S. E. (1995). Friendship quality during preadolescence and adolescence: The contribution of relationship orientations, instrumentality, and expressivity. *Merrill Palmer Quarterly, 41,* 517-535.

Katz, P. A. (1979). The development of female identity. *Sex Roles, 5,* 155-178.

Katz, P. A., & Ksansnek. K. A. (1994). Developmental aspects of gender role flexibility and traditionality in middle childhood and adolescence. *Developmental Psychology, 30,* 272-282.

Lobel, T. E., Bempechat. J., Gewirtz, J. C., Shoken-Topaz, T., & Bach, E. (1993). The role of gender-related information and self-endorsement of traits in preadolescents' inferences and judgments. *Child Development, 64,* 1285-1294.

Lubinski, D., Tellegen, A., & Butcher, J. N. (1983). Masculinity, femininity, and androgyny viewed and assessed as distinct concepts. *Journal of Personality and Social Psychology, 44,* 428-439.

Markus, H. R., & Kitayama, S. (1991). Culture and self: Implications for cognition, emotion, and motivation. *Psychological Review, 98,* 224-253.

Marsh, H. W., & Byrne, B. M. (1991). Differentiated additive androgyny model: Relations between masculinity, femininity and multiple dimensions of the self-concept. *Journal of Personality and Social Psychology, 61,* 811-828.

Martin, H., & Ramanaiah, N. (1988). Confirmatory factor analysis of the Bem Sex-Role Inventory. *Psychological Reports, 62,* 343-350.

O'Connor, K., Mann, D., & Bardwick, J. (1978). Androgyny and self-esteem in the upper-middle-class: A replication of Spence. *Journal of Counseling and Clinical Psychology, 46,* 1168-1169.

O'Heron, C. A., & Orlofsky, J. L. (1990). Stereotypic and nonstereotypic sex role trait and behavior orientations, gender identity, and psychological adjustment. *Journal of Personality and Social Psychology, 58,* 134-143.

Orlofsky, J. L. (1981). Relationship between sex role attitudes and personality traits and the Sex Role Behavior Scale—I: A new measure of masculine and feminine role behaviors and interests. *Journal of Personality and Social Psychology, 40,* 927-940.

Payne, T. J., Connor, J. M., & Colletti, G. (1987). Gender-based schematic processing: An empirical investigation and reevaluation. *Journal of Personality and Social Psychology, 52,* 937-945.

Pedhazur, E. J., & Tetenbaum. T. J. (1979). Bem Sex Role Inventory: A theoretical and methodological critique. *Journal of Personality and Social Psychology, 37,* 996-1016.

Schau, C. G., Kahn, L., & Diepold, J. H. (1980). The relationship of parental expectations and preschool children's verbal sex tying to their sex-typed play behavior. *Child Development, 54,* 266-270.

Serbin, L. A., Powlishta, K. K., & Gulko, J. (1993). The development of sex typing in middle childhood. *Monographs of the Society in Research in Child Development, 58,* No. 2, Serial No. 232, 1-97.

Sidanius, J., Cling, B. J., & Pratto, F. (1991). Ranking and linking as a function of sex and gender role. *Journal of Social Issues, 47,* 131-147.

Signorella, M. L. (1987). Gender schemata: Individual differences and context effects. *New Directions for Child Development, 38,* 23-37.

Signorella, M. L. (1992). Remembering gender-related information. *Sex Roles, 27,* 143-156.

Signorella, M. L., Bigler, R. S., & Liben, L. S. (1993). Developmental differences in children's gender schemata about others: A meta-analytic review. *Developmental Review, 13,* 147-183.

Silvern, L. E., & Katz, P. A. (1986). Gender roles and adjustment in elementary school children: A multidimensional approach. *Sex Roles, 14,* 181-202.

Simms, R. E., Davis, M. H., Foushee, H. C., Holahan, C. K., Spence, J. T., & Helmreich. R. L. (1978). *Psychological masculinity and femininity in children and their relationships to trait stereotypes and toy preferences.* Paper presented at annual meeting of Southwestern Psychological Association.

Spence, J. T. (1985). Gender identity and its implications for the concepts of masculinity and femininity. In T. B. Sonderegger (Ed.), *Psychology and gender: Nebraska Symposium on Motivation.* Lincoln: University of Nebraska Press.

Spence, J. T. (1991). Do the BSRI and PAQ measure the same or different concepts? *Psychology of Women Quarterly, 15,* 141-165.

Spence, J. T. (1993). Gender-related traits and gender ideology: Evidence for a multi-factorial theory. *Journal of Personality and Social Psychology, 64,* 624-635.

Spence, J. T., & Buckner, C. (1995). The enigma of masculinity and femininity: Gender identity and its contribution to gender-differentiating phenomena. In P. Kalbfleisch & M. Cody (Eds.), *Gender, power, and communication in human relationships.* Hillsdale, NJ: Erlbaum.

Spence, J. T., & Helmreich, R. L. (1978a). *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents.* Austin: University of Texas Press.

Spence, J. T., & Helmreich, R. L. (1978b). *The Intermediate Personal Attributes Questionnaire: A simplified version for children and adults.* Unpublished manuscript, Department of Psychology, University of Texas at Austin.

Spence, J. T., Helmreich, R. L., & Holahan, C. K. (1979). Negative and positive components of psychological masculinity and femininity. *Journal of Personality and Social Psychology, 37,* 1673- 1682.

Spence, J. T., Helmreich, R., & Stapp, J. (1974). The Personal Attributes Questionnaire: A measure of sex-role stereotypes and masculinity–femininity. *JSAS Catalog of Selected Documents in Psychology, 4,* 43-44.

Spence, J. T., Helmreich, R. L., & Stapp, J. (1975). Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology, 32,* 29-39.

Whitley, B. E. (1983). Sex role orientation and self esteem: A critical meta-analytical review. *Journal of Personality and Social Psychology, 44,* 765-785.

Yekel, C. A., Bigler, R. S., & Liben, L. S. (1991). *Children's gender schemata: Occupation, activity, and trait attributions for self and others.* Paper presented at the Biennial Meeting of the Society for Research in Children Development, Seattle, WA.