Assessing Recognition of Sexist Language: Development and Use of the Gender-Specific Language Scale

Mark R. McMinn

Paul E. Williams

Lisa Graham McMinn

Assessing Recognition of Sexist Language: Development and Use of the Gender-Specific Language Scale[1]

Mark R. McMinn[2]
Wheaton College

Paul E. Williams
George Fox College

Lisa C. McMinn
Portland State University

Teaching effective writing in the social sciences includes teaching recognition of sexist language. The development and teaching uses of the Gender-Specific Language Scale (GSLS), an instrument designed to assess recognition of sexist language, are described. Three experiments with predominantly European-American male and female students provide support for the reliability and validity of the GSLS, and suggest that it measures a different construct than an essay questionnaire used in previous studies of sexist language. Implications for teaching are discussed.

If our perceptions of the world are shaped by the language we use (Whorf, 1956), then our ideas about gender are influenced by ways we refer to human actors in particular roles. Several researchers have demonstrated that sexist language affects the response of those hearing or reading (Benoit & Shell, 1985; Briere & Lanktree, 1983; Dayhoff, 1983; McMinn, Lindsay, Hannum, & Troyer, 1990), which suggests that identifying and correcting

[2]To whom correspondence should addressed at Department of Psychology, Wheaton College, Wheaton, IL 60187.

sexist language is an important part of any educational program that attempts to move students beyond sexist stereotypes. Accordingly, since 1982 all American Psychological Association journals have required nonsexist language for submitted manuscripts. There is a growing social awareness of sexism, and the social scientist who writes with gender-specific language is often viewed as insensitive or out-of-date.

Given the increasing concern about sexist language and the importance of the topic, it is surprising that there are no standardized instruments available to measure the use or recognition of sexist language. Even teachers who conscientiously teach students to recognize sexist language have no systematic way of measuring the effectiveness of their teaching strategies. To date, studies measuring sexist language have consistently used individually designed instruments that have no demonstrated reliability or validity.

In their studies of the effects of sexist language, McMinn and his colleagues have used an essay response to open-ended questions to evaluate college students' tendencies to use sexist language (McMinn & Foster, 1991; McMinn et al., 1990; McMinn, Troyer, Hannum, & Foster, 1991). One such question was, "A business executive discovers a long-time employee has been stealing from the company. What should the executive do first?" Students' responses were rated by independent judges for sexist language. Most often, students using sexist language assumed a male business executive.

There are several disadvantages to using an open-ended essay to evaluate sexist language production. First, some students are more concise than others. Those responding concisely, such as those who wrote "Talk to the employee," may have avoided sexist language by the brevity of the response rather than by sensitivity to gender-specific assumptions. Thus, additional error variance is introduced by the varying lengths of responses on the essay questions. Second, McMinn's earlier studies showed the production of sexist language to be only mildly affected or not at all affected by either a didactic presentation or a computerized presentation on the perils of sexist language, suggesting that either college students are slow to learn about sexist language or the essay instrument has not effectively detected the changes. Third, it seems likely that recognition of sexist language and production of sexist language are slightly different phenomena. The scale described here measures recognition of gender-specific language rather than production of sexist language, which the essay questions measured.

One of the difficulties in measuring sexist language is the lack of precision in defining sexist language. At the core of the problem is the distinction between sexist attitudes and language use. If a news broadcaster reports that "the university basketball team won tonight, but the women's team lost by a narrow margin," the primary problem is the underlying sexist

assumption that the men's basketball team is primary whereas the women's team is secondary. Does this then count as sexist language? Actually, the broadcaster is using precise language given his or her underlying assumptions. Is this a problem of sexist assumptions, sexist language, or both? Another example can be seen in the editorial writer who proclaims, "if a nurse is concerned about bad hours, she ought to consider working for a temporary agency." In this case the writer may not be revealing sexist assumptions—there is no sense in which male is being valued over female, or vice versa. In fact, the writer may have never experienced a male nurse. Nonetheless, the language is gender specific in referring to a career that is not. In this case, there may be no sexist assumptions, but there is gender-specific language.

This can also be seen in the Publication Manual of the American Psychological Association (APA, 1983), where a distinction is made between problems of evaluation and problems of designation. Evaluation errors refer to the sexist assumptions that underlie, but may be revealed in, language. For example, "Man and wife" reveals a evaluation error because of the implication that a woman's identity is in relation to her spouse and a man's identity is not. Designation errors refer to imprecise language that may perpetuate sexism in culture. For example, many have used generic masculine pronouns to refer to humankind. Although the writer may not have sexist intentions, such language is not perceived to be generic by those hearing and reading the words (Cole, Hill, & Dayley, 1983; Gastil, 1990; Hamilton, 1988; Schneider & Hacker, 1973; Wilson & Ng, 1988). This is similar to Pearson's (1985) distinction between man-linked words (designation errors) and sexist cultural practices (evaluation errors).

Through a series of studies with introductory college students, we developed a scale to measure recognition of sexist language. Initially our efforts were to measure both evaluation errors and designation errors by designing two scales for each construct. The initial scale consisted of 24 items. However, the scales measuring evaluation errors lacked good psychometric properties, and so the scale has evolved to a 12-item Gender-Specific Language Scale (GSLS) with the intent of measuring only designation errors. This evolution is described here.

The scale consists of sentences with a variety of grammatical, spelling, and punctuation problems. Six of the sentences also include gender-specific language. Each sentence with gender-specific language is matched with a similar sentence not containing gender-specific language. In addition to the sexist designation problem in half of the sentences, the number of errors in each sentence ranges from zero to two. Respondents are instructed to identify the problems in each sentence, including problems with grammar, spelling, punctuation, and discriminatory language. The three studies de-

scribed here were designed to refine item selection and assess the reliability and validity of the GSLS. It should be noted that the scale presented here is specific to Standard American English, and is not intended to be cross-culturally valid.

STUDY 1

The purpose of the first experiment was to refine a preliminary draft of the scale and assess its reliability. The scale items were initially derived from principles and examples in the publication manual of the American Psychological Association (American Psychological Association, 1983). The original scale consisted of 12 items with sexist language and 12 matched items without sexist language. The 12 sexist language items were broken into four subscales, two that generally reflect evaluation errors and two that generally reflect designation errors: pronouns (designation), generic male (designation), stereotypic assumptions (evaluation), and parallelism (evaluation). Each subscale consisted of 3 items and had 1 item with no errors except for the sexist language, 1 item with one error plus the sexist language, and 1 item with two errors plus the sexist language. The initial item pool, without the grammar and spelling errors, is listed in Appendix A.

The Pronouns subscale had items that used gender specific pronouns inappropriately. For example, "Much has been written about the effect that a child's position among his siblings has on his intellectual development." The correct response on this item is to identify the word "his" as discriminatory. Although gender-specific pronouns may reflect underlying evaluation biases, they more directly demonstrate a problem of designation—a learned, formerly approved use of language that has persisted despite recent changes in language use.

The Generic Male subscale had items that assumed a masculine referent to include all humans. For example, "The use of experiments in psychology presupposes the mechanistic nature of man." The correct response on this item is to identify the word "man" as discriminatory. As with the Pronoun subscale, when these items are not correctly identified it may reflect either an evaluation or a designation problem. Because using a generic masculine referent was previously acceptable in the English language, errors on this scale are assumed to be designation problems that have not yet been unlearned.

The Stereotypic Assumption subscale had items that implied stereotypic gender roles. For example, "The client's husband lets her teach part-time," is discriminatory because it implies the husband is in charge of the

wife's work schedule. A more appropriate way of stating this is, "The client reported that her husband lets her teach part-time." Items on this scale were intended to reflect evaluation errors because they probe for underlying attitudes about gender and gender roles.

The Lack of Parallelism subscale had items that referred to men and women in disparate ways. For example, "The surgical team included a lady doctor and a male nurse," is discriminatory because the parallel word for male is female and the parallel word for lady is gentleman. This scale was also intended to measure evaluation problems.

Items from the various subscales, and their matched counterpart items, were randomly distributed throughout the 24-item scale. The instructions given the respondents are written at the top of the scale, listed in its current form in Appendix B.

Method

Participants. Participants were recruited from an introductory sociology class at Portland State University. Of the 129 participants, 57 were men, 65 were women, and 7 did not report their gender. Although ethnicity information was not requested, the sample was drawn from a representative general education class at a university with a population of 1.1% American Indian, 7.7% Asian American, 2.7% African American, 2.5% Latino, 81.2% European American, and 4.8% international students ("1992 Enrollment," 1994). Participation was voluntary.

Procedure. Participants were given the questionnaire and allowed time in class to identify the grammatical, spelling, punctuation, and discriminatory language errors. Scales were collected the same class period, and participants were debriefed immediately.

The 12 items containing sexist language were then scored by one of the authors. Each item was scored separately, and entered in an SPSS data file for reliability analyses.

Results

Internal consistency was measured with coefficient alpha. The four subscales combined produced an internal consistency of .77. Three of the four subscales had modest reliability (Pronouns *alpha* = .67; Generic Male *alpha* = .75; Stereotypic Assumptions *alpha* = .59), reflecting the brevity of each subscale. The Lack of Parallelism subscale had poor internal consistency (*alpha* = .42), and so diminished the overall reliability of the scale.

The internal consistency after the Lack of Parallelism subscale was omitted increased to .81.

In addition to the poor reliability of the Lack of Parallelism scale, two other observations contributed to its eventual deletion from the scale. First, the correlations among subscales (see Table I) demonstrated a weak relationship between the Lack of Parallelism scale and the other scales. Moreover, inspection of the interitem correlation matrix (see Table II) indicated that the scales items on the Lack of Parallelism subscale correlated poorly with other items.

Thus, 6 items (3 from the Lack of Parallelism subscale and 3 matched items) were deleted from our original scale, leaving a scale of 18 items.

Discussion

The modest reliability of the individual subscales was due in part to the limited number of items in each subscale. Internal consistency of the overall scale increased to an acceptable level, especially after the Lack of Parallelism subscale items were omitted.

Interestingly, the subscales with the lowest reliabilities were those designed to measure problems of evaluation. The two scales designed to measure problems of designation had relatively high reliability. This probably reflects the greater difficulty in measuring underlying assumptions than in measuring language-use behaviors.

This experiment provides initial support for the reliability of the GSLS, especially for the subscales designed to measure designation errors. Additional studies were necessary to demonstrate validity of the scale and com-

Table I. Correlation Among the Subscales of the GSLS[a]

	Pronouns	Generic Male	Stereotypic Assumptions	Lack of Parallelism
Pronouns	1.0			
Generic Male	.61[b]	1.0		
Stereotypic Assumptions	.76[b]	.62[b]	1.0	
Lack of Parallelism	.22	.14	.34[b]	1.0
Total	.80[b]	.77[b]	.88[b]	.59[b]

[a]All correlations are Pearson product-moment correlations.
[b]One tailed $p < .01$.

Table II. Interitem Correlation Matrix in Study 1[a]

Item	P1	P2	P3	G1	G2	G3	S1	S2	S3	L1	L2	L3
P1	1.0											
P2	.47	1.0										
P3	.30	.47	1.0									
G1	.32	.42	.43	1.0								
G2	.35	.34	.11	.39	1.0							
G3	.36	.35	.10	.39	.71	1.0						
S1	.41	.43	.28	.20	.23	.19	1.0					
S2	.33	.45	.24	.28	.36	.32	.26	1.0				
S3	.09	.26	.09	.04	.38	.36	.26	.50	1.0			
L1	.19	.09	.19	.08	.05	.06	.34	.07	.15	1.0		
L2	.19	.17	.10	.01	.07	.00	.13	.22	.05	.23	1.0	
L3	.06	.12	.10	.09	.08	.16	.29	.11	.16	.22	.16	1.0

[a]Item numbers refer to the items listed in Appendix A. "P" refers to the Pronouns subscale, "G" to the Generic Male subscale, "S" to the Stereotypic Assumptions subscale, and "L" to the Lack of Parallelism subscale. All correlations are Pearson product-moment correlations.

pare results on the GSLS with scores on the essay evaluation we used in earlier studies.

STUDY 2

The purpose of the second study was to investigate the content validity of the GSLS and each of its items. If the subscales measure recognition of sexist assumptions and language, then instruction about sexist language should alter one's performance on the scale.

Method

Participants. Participants were recruited from an introductory psychology class at George Fox College. Of the 35 participants, 21 were women, 13 were men, and 1 did not report gender. Although ethnicity information was not requested, the sample was drawn from a representative general education class at a college with a population of 1.9% American Indian, 2.2% Asian American, 1.1% African American, 1.3% Latino, 90% European American, and 3.5% international students ("1992 Enrollment," 1994). Participation was voluntary.

Procedure. First, participants were randomly divided into two groups. The experimental group was given a three-page worksheet titled, "Learning

to Recognize Sexist Language." The worksheet described different types of sexist language, gave examples of sexist language, and gave sample sentences for participants to edit. None of the sentences in the worksheet were duplicated from the GSLS. The control group received a worksheet of similar length and structure titled, "Learning Rules of Standard English Grammar." The control worksheet addressed proper use of verbs, and did not address sexist language. Participants in both groups then completed the 18-item GSLS, which was later scored by one of the authors.

Results

Table III shows the number of each sample that correctly identified the sexist language in each scale item containing sexist language. Chi-square analyses demonstrate that the experimental intervention produced a significant difference for each scale item with sexist language.

The internal consistency (coefficient alpha) for the overall GSLS was .95, higher for this sample than for the first. This increase in internal consistency suggests that the worksheets participants completed before completing the scale may have increased the consistency of their responses to the various items.

Table III. Numbers of Respondents Correctly Identifying Sexist Language in Study 2[a]

Item no.	Experimental	Control	Chi-square
P1	14	2	15.4^d
P2	17	1	27.5^d
P3	17	4	18.3^d
G1	17	2	24.1^d
G2	13	3	10.5^c
G3	14	2	15.4^d
S1	11	4	5.0^b
S2	17	2	24.1^d
S3	14	0	22.0^d

[a]All scores represent the number of respondents correctly identifying sexist language. There were 18 participants in the experimental group and 17 in the control group. Item numbers refer to the items listed in Appendix A.
[b]$p < .05$.
[c]$p < .01$.
[d]$p < .001$.

Discussion

The GSLS appears to be measuring what it purports to measure. Participants receiving instruction in sexist language more consistently identified sexist language in each of the scale items containing sexist language than did control participants.

It could be argued that those reading the passage on sexist language were merely more sensitized to the directions on the subsequent administration of the scale than those in the control group, and therefore this is not an indication of the scale's validity. However, sensitivity is the relevant issue in correcting designation errors in sexist language. Those who do not have underlying sexist assumptions should be able to recognize sexist language once they have been sensitized to the problems with sexist language. Thus, these data provide evidence of the ecological validity of the instrument.

Two of the three items on the Stereotypic Assumption subscale were problematic. First, Item S1 showed the weakest discrimination of all items between the experimental and control groups. Second, there was no variance on Item S3 among the control group, suggesting it may not be measuring the same domain as the other items. Also, Item S3 showed weak intercorrelations with many items in Study 1 (see Table II). Because of these problematic items and the fact that the stereotypic assumptions subscale had lower internal consistency than the other remaining subscales in both Study 1 and Study 2, the Stereotypic Assumptions subscale was dropped from the final version of the test (see Appendix B). Once both of the subscales designed to measure evaluation problems were excluded from the scale, the test was renamed from its original title, "Sexist Language Scale," to "Gender-Specific Language Scale." That is, the remaining items on the test are designed to detect recognition of designation errors: gender-specific language.

STUDY 3

The purpose of Study 3 was to compare scores on the Gender-Specific Language Scale with scores on the essay questions we have used in previous research. As suggested previously, it may be that recognizing and producing sexist language are different activities that are not highly correlated. Those who avoid sexist language in an essay question may be insensitive to recognizing gender-specific language despite their avoidance of it in writing.

If the Gender-Specific Language Scale and the essay questionnaire measure similar constructs, a strong negative correlation between scores

on the two scales would be expected. That is, those who score high on the Gender-Specific Language Scale are sensitive to the problem of sexist language, and should score low on the essay questionnaire by not producing sexist sentences.

Method

Participants. Participants were recruited from undergraduate classes at George Fox College ($n = 33$) and Portland State University ($n = 108$). Of the 141 participants, 44 were men, 96 were women, and 1 did not report gender. Ethnicity was not requested of participants, but the ethnic diversity of the two schools was reported previously in descriptions of Studies 1 and 2. Participation was voluntary.

Procedure. Participants completed both the 18-item version Gender-Specific Language Scale (the decision to drop the Stereotypic Assumptions subscale had not yet been made) and the essay questions used for previous research. The three essay questions were as follows:

1. A business executive discovers a long-time employee has been stealing from the company. What should the executive do first?
2. A nurse discovers a hospital patient has been given blood contaminated with the AIDS virus. What should the nurse do first?
3. A professor discovers a student has cheated on an exam. What should the professor do first?

Responses to each of the three essay questions were scored for the number of times sexist language was used. Most commonly, sexist language is introduced by assuming the business executive is male or the nurse is female. Occasionally participants assume the college professor is male. In previous studies (McMinn et al., 1990), scorers of the essay test have achieved an interrater reliability of 1.00. Thus, only one skilled rater was used to score the essay responses in this study.

Results

The overall score on the Gender-Specific Language Scale was the number of gender-specific language items correctly identified, and ranged from zero to six. The overall score on the essay questionnaire was the total number of times sexist language was used in responding to the three essay questions, and ranged from zero to seven.

The internal consistency of the essay questionnaire was low (*alpha* = .43). This is not surprising given the distinct nature of each of the essay questions—they are not intended to be homogeneous, and some items have been shown to be more sensitive to detecting sexist language than others (McMinn, Lindsay, Hannum, & Troyer, 1990).

For the overall sample, the two measures of sexist language showed a negative correlation ($r = -.225$, $p < .01$). The correlation of the Gender-Specific Language Scale and the essay questionnaire among George Fox students was similar to that of the overall sample ($r = -.177$, ns). The magnitude of the correlation among Portland State students was also similar ($r = -.244$, $p < .05$). Because there was not a significant difference in the correlations between the samples ($q_s = .07$, $q_c = .34$; see Cohen, 1988), the samples were combined for subsequent analyses. Because the low internal consistency of the essay measure attenuates the correlation between the two scales, a correlation coefficient corrected for measurement error was computed, resulting in a corrected correlation coefficient of –0.42 (see Hunter, Schmidt, & Jackson, 1982).

Discussion

Because both the essay questionnaire and the GSLS are measuring constructs that are somewhat related, we would expect them to be correlated. In fact, they are. The low magnitude of correlation suggests that only a modest amount of variance, approximately 18%, is shared between the two measures. Thus, the essay questionnaire appears to be measuring a construct that overlaps with, but is quite different from, what the GSLS measures. The primary difference appears to be that the GSLS measures recognition of sexist language whereas the essay questionnaire measures production of sexist language. Those who fail to recognize sexist language may not produce it in their writing despite their lack of awareness. Conversely, some people may recognize sexist language in someone else's writing, but not in their own.

USES FOR THE GSLS

There are at least two potential uses for the Gender-Specific Language Scale. First, it can be a useful learning tool for college students. For example, administering the scale to an introductory social science class is a useful way of sensitizing students to the problem of sexist language. After administering the scale and having students score their re-

sponses, there is a natural opportunity to discuss the problems with gender-specific language and grammatically acceptable alternatives for restructuring sentences containing gender-specific language. Alternatively, the scale could be used after a discussion of sexist language to assess the effectiveness of the presentation.

An additional teaching tool is a Macintosh computer program developed and tested by McMinn and Foster (1991). The program is a computer-aided technique that helps students recognize and correct sexist language. Students learn grammatically correct alternatives to noninclusive language. Completing the program is an appropriate follow-up assignment to an in-class discussion of sexist language.

Second, the GSLS provides the first empirically evaluated tool available to measure gender-specific language. We hope the GSLS may stimulate more research on effective ways to teach students to recognize sexist language and avoid it in their writing. The results of McMinn's earlier studies (McMinn et al., 1990, 1991) have demonstrated little or no effect of presenting students with information about sexist language on their production of sexist language. These studies, however, have used the essay questionnaire to measure sexist language production rather than the GSLS. Replication studies using the GSLS will be forthcoming. We may find that didactic presentations are more useful in helping students recognize sexist language than they were in preventing production of sexist language.

Although the GSLS appears to have adequate psychometric properties, this was accomplished by removing items relating to underlying sexist perceptions and attitudes that may shape language (evaluation problems). Apparently, the easier task is to measure the recognition of gender-specific words (designation problems). Although this is an important start, more work is needed to further clarify and operationalize the evaluation errors that contribute to sexist language.

Because of the homogenous nature of our experimental groups, it will be important to continue evaluating the GSLS. For example, we do not know if the GSLS will provide the same type of information among more ethnically diverse groups or among middle-aged and elderly adults.

A scale to assess recognition of sexist language is an important starting point. The long-term goals are to discover teaching methods that go beyond recognition to preventing production of sexist language and, more importantly, to identify and dispel the underlying evaluation errors that perpetuate sexism.

APPENDIX A

Initial Sexist Item Pool

Pronouns

    P1. Much has been written about the effect that a child's position among his siblings has on his intellectual development.

    P2. First the individual becomes aroused by violations of personal space, and then he attributes the cause of this arousal to other people in his environment.

    P3. Each person's alertness was measured by the difference between his obtained relaxation score and his obtained arousal score.

Generic Male

    G1. The use of experiments in psychology presupposes the mechanistic nature of man.

    G2. The chairman of the board presided over the meeting.

    G3. The mailman was never late, no matter how bad the weather.

Sterotypic Assumptions

    S1. The client's husband lets her teach part-time.

    S2. Research scientists often neglect their wives and children.

    S3. Many housewives find their work as stressful as the work of business executives.

Lack of Parallelism

    L1. The surgical team included a lady doctor and a male nurse.

    L2. The college basketball team was undefeated and ranked third in the nation, but the women's team had the worst record in the league.

    L3. Ten college men and seven coeds went out for pizza.

APPENDIX B

Written Language Quiz

Read each of the following statements carefully and circle every problem you find, including problems with grammar, spelling, punctuation, and discriminatory language. Work quickly as you will have only five to ten minutes to complete this task.

Example: The job was demanding, resulting in numerus frustration's.
1. Each persons' alertness was measured by the diference between his obtained relaxation score and his obtained arousal score.
2. The use of experiments in psychology presupposes the mechanistic nature of man.
3. The busness executive's learned about domestic tasks from the homemakers.
4. When making an important decision one must first determine how other's will be affected and if the outcome is worth the cost.
5. The chairman of the board precided over the meeting.
6. The mailman wasn't never late, no matter how bad the whether.
7. She said she would ask her husband if she could go on the weekend trip with us.
8. The supervisor talked individually with the employees who were to be layed off.
9. The fire fighters' maintained composure when comfronted by the large dog.
10. First the individual becomes aroused by violations of personal space and then he attributes the cause of this arosal to other people in his environment.
11. Evolutionary theory proposes that the human species is evolving through a process of survival of the fittest.
12. Much has been written about the effect that a child's position among his siblings has on his intellectual development.

REFERENCES

1992 enrollment by race at 3,300 institutions of higher education. (1994, February 23). *The Chronicle of Higher Education,* pp. A31-A40.

American Psychological Association. (1983). *Publication manual of the American Psychological Association* (3rd ed.). Washington, DC: Author.

Benoit, S. S., & Shell, W. (1985). Can sex-biased communication affect business students' career decisions? *Delta Pi Epsilon Journal, 28,* 123-129.

Briere, J., & Lanktree, C. (1983). Sex-role related effects of sex bias in language. *Sex Roles, 9,* 625-632.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cole, C. M., Hill, F. A., & Dayley, L. J. (1983). Do masculine pronouns used generically lead to thoughts of men? *Sex Roles, 9,* 737-750.

Dayhoff, S. A. (1983). Sexist language and person perception: Evaluation of candidates from newspaper articles. *Sex Roles, 9,* 527-539.

Gastil, J. (1990). Generic pronouns and sexist language: The oxymoronic character of masculine generics. *Sex Roles, 23,* 629-643.

Hamilton, M. C. (1988). Using masculine generics: Does generic "he" increase male bias in the user's imagery? *Sex Roles, 19,* 785-799.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies.* Beverly Hills, CA: Sage.

McMinn, M. R., & Foster, J. D. (1991). A computer program to teach nonsexist language. *Teaching of Psychology, 18,* 115-117.

McMinn, M. R., Lindsay, S., Hannum, L. E., & Troyer, P. K. (1990). Does sexist language reflect personal characteristics? *Sex Roles, 23,* 389-396.

McMinn, M. R., Troyer, P. K., Hannum, L. E., & Foster, J. D. (1991). Teaching nonsexist language to college students. *Journal of Experimental Education, 59,* 153-161.

Schneider, J. W., & Hacker, S. L. (1973). Sex role imagery and use of the generic "man" in introductory texts: A case in the sociology of sociology. *American Sociologist, 8,* 12-18.

Pearson, J. C. (1985). *Gender and communication.* Dubuque, IA: Wm. C. Brown Publishers.

Whorf, B. L. (1956). *Language, thought and reality: Selected writings of Benlamin Lee Whorf* (J. B. Carroll, Ed.). Cambridge, MA: MIT Press.

Wilson, E., & Ng, S.H. (1988). Sex bias in visual images evoked by generics: A New Zealand study. *Sex Roles, 18,* 159-169.