DOCUMENT RESUME

ED 156 699 TM 007 134

AUTHOR Campbell, Patricia B.
TITLE A Useful Evaluation of Sex Role Materials: Possibility or Pipe Dream?
PUB DATE Mar 78
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Affective Objectives; Age Differences; Cognitive Objectives; *Curriculum Evaluation; Elementary Secondary Education; *Equal Education; Evaluation Criteria; *Evaluation Methods; *Evaluation Needs; Feasibility Studies; Formative Evaluation; *Instructional Media; Sex Differences; Sex Discrimination; *Sex Role; Sex Stereotypes; Social Attitudes; Student Teachers; Summative Evaluation; Teacher Attitudes; Teacher Educators
IDENTIFIERS Bem Sex Role Inventory; *Project on Sex Stereotyping in Education

ABSTRACT

An evaluation design for educational equity materials was utilized with seven instructional units on sex roles and sex stereotyping for teachers and teacher educators. The instructional units were developed by the Project on Sex Stereotyping in Education in the areas of human growth and development, mathematics education, science education, language arts education, educational history, social studies education, and physical education. Each unit was reviewed by experts in sex stereotyping, subject area, and multi-cultural education. In embarrassment field tests, education students completed questionnaires identifying possible unit improvements. Multiple-choice cognitive tests incorporated alternatives to stereotyped behaviors and myths about males and females. An adaption of the Bem Sex Role Inventory (BSRI), in which subjects rate themselves on characteristics which are given masculinity, femininity, or androgyny scores, was used as an affective instrument for unit evaluations. The units and evaluation instruments were administered by teacher educators. Results indicated that the units significantly increased subjects' knowledge, but did little to affect scores on the BSRI. Inclusion of periodic revision points, and the incorporation of expert and user reviews are among the strengths of the evaluation design. Design weaknesses include the difficulty of obtaining affective measurement results which are both valid and reliable. (Author/JAC)

A Useful Evaluation of Sex Role Materials: Possibility or Pipe Dream?

Patricia B. Campbell
Project on Sex Stereotyping in Education
Women Educators

Paper presented at the annual meeting of the American Educational Research Association, Toronto, Ontario, March, 1978.

A Useful Evaluation of Sex Role Materials: Possibility or Pipe Dream?

Patricia B. Campbell
Project on Sex Stereotyping in Education
Women Educators
P.O. Box 218
Red Bank, New Jersey 07701

This presentation will cover the role that evaluation can play in the development of multi-media instructional units on sex roles and sex stereotyping. An evaluation design for the development of sex role materials, which includes content reviewers, multi-cultural reviewers, embarassment field tests and cognitive, affective and observational validation components will be discussed. The strengths and weaknesses of the design, as found by its application to the seven instructional units developed by the Project on Sex Stereotyping in Education will also be discussed. Some of the design strengths that will be covered, include the inclusion of periodic revision points and the incorporation of a variety of expert and user reviews. Design weaknesses include the difficulty of getting affective measurement results which are both valid and reliable.

A Useful Evaluation of Sex Role Materials: Possibility or Pipe Dream?

In the past few years there has been a tremendous upsurge in the development of educational materials on women's educational equity. The Federal Government (WEEAP Final Report, 1976; WEEAP Final Report, 1977), State Departments of Education (Pennsylvania Department of Education, 1974; Pennsylvania Department of Education, 1973), school systems (Berkeley Unified School District, 1976; Ceres Unified School District, 1977), teacher organizations (Resource Center on Sex Roles, 1974; American Federation of Teachers, 1974), private organizations (Emma Willard Task Force, 1972; Rosenfeld - The Feminist Press, 1976) and even individuals (Rothchild, 1973; Campbell and Thompson, 1974, Campbell, 1975) are developing and producing a wide range of materials whose goal is to promote educational equity for males and females.

Unfortunately many of these materials are not being evaluated. The reasons for this are centered around three major factors - lack of money, lack of confidence in the evaluative process and lack of valid reliable measuring instruments.

Many materials development efforts are funded only for supplies or not at all, with the developers frequently being people who volunteer their time because they feel there is a need for educational equity materials. Under these circumstances it is somewhat unreasonable to expect a well designed, somewhat expensive evaluation. Even evaluation that could be done inexpensively by the developers themselves is frequently not done because of a feeling that all of the limited resources available should be put into the materials.

Developers disinclination towards evaluation is also because of a lack of confidence in the evaluative process. Evaluators are often viewed as villains whose role is to hurt rather than help. (Granger and Campbell, 1977). Frequently, they are seen as members of an educational establishment that has in the past shown little concern for educational equity.

Another major concern is measurement. Particularly in the affective area, we do not know how to measure the effects that educational equity materials or social change materials in general are having. And without adequate measures even the best evaluation is doomed to failure.

These concerns are legitimate and serious. Although paper and pencil instruments on sex roles and sex bias abound, few deal specifically with education and even fewer have validity and reliability. (Findley et. al.) Also, many professional evaluators are not knowledgable about and sensitive to the issues and concerns of women's educational equity; while many materials developers are not knowledgable about areas like design, measurement and statistics. Even if the evaluator and developer get together, they frequently have no common ground on which to meet.

However, in spite of these difficulties, evaluation must be done on educational equity materials. Potential users need to know if materials are effective, developers need to know how the materials can be improved and, last but not least, federally funded materials developers need to provide the government with evidence that project objectives have been met.

Any evaluation of educational equity materials must include both formative and summative components. In addition, while the evaluation must be theoretically sound, it must also be practical and feasible. The following is one educational equity materials development project's attempt to develop and implement such an evaluation.

The Project on Sex Stereotyping in Education (POSSE) was funded by the Women's Educational Equity Act Program of the Office of Education. Its goal was to develop and validate seven instructional units on sex roles and sex stereotyping for teachers and teacher educators. The multi-mediated units it developed consisted of an audio tape, transparency masters, handouts, a bibliography, an instruction sheet and a type script of the audio tape. The seven content areas that were covered were human growth and development, math education, science education,

language arts education, educational history, social studies education and physical education.

Project objectives were to create units that:

1. provided up-to-date research information on sex stereotyping and its effects;
2. refuted traditional myths about the roles and activities of women and men, boys and girls;
3. increased teachers' awareness of their influence on the development of their students sex roles;
4. provide alternatives to stereotypic behavior for teachers;
5. could be used independently or adapted by course instructors (Campbell, 1976).

An evaluation design was developed that sought to determine if the project met its objectives and what effects, if any, the units had on pre- and in-service teachers.

The formative evaluation component involved the use of reviewers and field tests. Each unit was reviewed for accuracy and completeness by people with expertise in sex stereotyping, unit content area and multi-cultural education. For example, the math unit was reviewed by Dr. Dora Skypek, a math educator from Emory University and by Dr. Gwen Baker, Director of the Multi-cultural Educational Program at the University of Michigan. The review process and the revisions, based on reviewer feedback, insured that accurate, complete and up-to-date information was provided by the unit.

A second component of the formative evaluation was the embarassment field test. Each completed unit was given to a group of education students who upon completing the unit filled out an open-ended field test questionnaire. The students were asked to list the things they liked most, the things they liked least, and the ways they would improve the instructional unit as a whole, the audio tape, the transparencies and the handouts. Students were also interviewed by project staff to determine students' perceptions of the clarity, interest and usability

of the unit. Based on the responses to the questionnaires and the interviews, each unit was revised.

A final component of the formative evaluation was the summative evaluation of earlier units. That is, based on the results of the evaluation of the first several units, changes were made in the formats and development process of the later units. The three step process of review, field test and learning from previous mistakes was found by project staff to be both useful and efficient.

The summative evaluation was centered around a national validation of the units using both cognitive and affective instruments. Each unit was listed on a national sample of pre- and in-service teachers using a pre-test treatment post-test design.

Cognitive tests were developed for each unit using a table of specifications. Test items were in multiple choice and true-false formats and included questions on myths about females and males and on alternatives to stereotypic behaviors. Items for each unit cognitive test were field tested by graduate education students. Based on an item analysis, project director review and the table of specifications, twenty to thirty items were selected for each unit test. Each cognitive test was also analyzed for internal consistency. The results ranged from a low of .59 for the science cognitive test to a high of .88 for the social studies test.

Developing an affective test for the units was a difficult task. A search of the literature for appropriate existing instruments proved fruitless, as did an attempt by project staff to construct a valid, reliable measure. As a result, an adaptation of the Bem Sex Role Inventory (BSRI) was used as a pre and post measure. In the BSRI, subjects rate themselves on a series of characteristics. Based on their ratings, three scores are calculated; a masculinity score, a femininity score and an androgyny score. Originally the adaptation was to have subjects rate a boy or a girl on a series of characteristics; however, our own research indicated that the age level of the boy or girl was a major factor in

subject's ratings. Therefore, in six of the seven units, subjects were asked to rate the characteristics of either an elementary school girl, an elementary school boy, a high school girl or a high school boy. (For the seventh unit on educational history, subjects were asked to rate either a male or female educator because, unlike the previous units, its content dealt with adults rather than children.) The BSRI was selected for use because of its high reliability (.93, Bem, 1974) and its criterion validity (Bem, 1975).

Using the adapted BSRI and the cognitive unit tests, the summative evaluation was done with the assistance of an eleven member national task force of teacher educators.

For each unit, participating task force members were sent an evaluation packet that included instructions, copies of the complete unit and testing sets. The testing sets were composed of instructions, pre-affective and cognitive tests and post-cognitive and affective tests.

Task force members were instructed to give the pre-tests, have the subjects go through the unit and give the post-tests. Interviews with the task force members indicated they found no difficulties in giving either the units or the tests.

Each unit was used by at least five people with no previous experience in working with the unit. In each case, without consultant assistance, the people were able to use the unit successfully. In two cases, people were asked to set up the units without any prior instruction and in both cases they were successful.

Originally it was planned to have the post-tests given a week after the students had completed the unit. Unfortunately, it was discovered that this procedure cut the number of available subjects down considerably. Therefore the post-tests were given immediately after the unit was completed. All reported subject scores are on post-tests taken immediately after completing the unit.

The results of the cognitive tests were combined, totaled and analyzed by the school. If a school sample size was under twenty a related t test was used, if the sample size was over twenty, a z test was used. Only one of the analyses (one

school in the human growth and development unit sample) was found to be non-significant. In over 80% of the analyses, the post-test was found to be significantly higher than the pre-test at the .0005 level..

Table 1 gives a summary of the mean pre and post cognitive tests scores in percentages and the Z scores for the seven units.

Table 1
A Summary of Cognitive Test Results

	Pre	Post	Z	N
Human Growth and Development Unit	54	70	Z = 13.57*	322
Math Unit	60.7	73.8	Z = 7.7*	161
Science Unit	54.6	72.1	Z = 11.8*	131
Educational History Unit	48.5	75.6	Z = 22.7*	193
Language Arts Unit	66.8	78.3	Z = 8.85*	153
Social Studies Unit	60	75	Z = 10.79*	155
Physical Education Unit	58.8	73.9	Z = 9.21*	130

* p . .0005

The affective androgyny scores from the science, math, educational history, language arts, social studies and physical education units were analyzed using an analysis of variance with repeated measures with equal cell sizes. Data was randomly eliminated from some cells in order to get equal cell sizes.

The human growth and development data was analyzed using analysis of variance. This data was not able to be analyzed using an analysis of variance with repeated measures because many of the subjects did not follow directions and did not write their subject numbers on both the pre and the post tests. In order to alleviate this problem in the other units, the rest of the validation packets were sent out with subject numbers already stamped on the pre and post tests.

In the human growth and development unit affective test results showed significant F scores on the treatment (pre-post), sex (female-male) and level (elementary-high school) variables. However post hoc analysis found no significant change at

the .05 level in the ratings for elementary and high school boys and girls before and after going through the unit.

Table 2
Human Growth and Development Unit
Affective Test Results #

	Pre	
	Elementary	High School
Boy	- 1.87	- 2.56
Girl	+ 2.25	+ 2.11

	Post	
	Elementary	High School
Boy	- 2.00	- 2.99
Girl	+ 1.92	+ 1.23

Source of Variation	Sum of Squares	D.F.	Mean Square	F.
Pre-Post (P)	31.5119	1	31.5119	5.4375 *
Elementary-High School (E)	64.8951	1	64.8951	11.1978 **
Male-Female (S)	2917.9895	1	2917.9895	503.5076 **
P x E	7.4450	1	7.4450	1.2847
P x S	4.4590	1	4.4590	0.7694
E x S	7.4649	1	7.4649	1.2881
P x E x S	0.5972	1	0.5972	0.1030
Error	3766.9600	650	5.7953	

Scores approaching zero are androgynous.
Scores approaching +4 are feminine.
Scores approaching -4 are masculine.

* $p < .025$
** $p < .001$

In the math unit, affective test results showed significant F scores on the variables, sex, level (elementary-high school) and treatment (pre-post). The interaction of sex, level and treatment was also significant. Post hoc analysis found no significant changes at the .05 level for the ratings for elementary and high school girls. However, elementary and high school boys were rated significantly more masculine after subjects took the unit.

Table 3

Math Unit
Affective Test Results #
N=30 cases per cell

	Pre	
	Elementary	High School
Boy	- .97	- 3.19
Girl	+ 2.77	+ 2.03

	Post	
	Elementary	High School
Boy	- 1.71	- 3.93
Girl	+ 2.80	+ 2.04

Source of Variation	Sum of Squares	D.F.	Mean Square	F.
Elementary-High School (E)	132.46	1	132.46	13.08 **
Male-Female (S)	1417.3	1	1417.3	139.97 **
E x S	32.85	1	32.85	3.24
Error	1174.63	116	10.13	
Pre-post (P)	7.8	1	7.8	4.20 *
E x P	.00305	1	.00305	.0016
S x P	8.81	1	8.81	4.74 *
E x S x P	.00717	1	.00717	.0039
Error	215.5	116	1.86	

Scores approaching zero are androgynous.
Scores approaching +4 are feminine.
Scores approaching -4 are masculine.

* $p < .05$

** $p < .001$

In the science unit, affective test results showed a significant F score on the variable sex. No significant differences on the variables level (elementary-high school) or treatment (pre-post) nor any significant interactions were found.

Table 4

Science Unit
Affective Test Results #
N=29 cases per cell

Pre

	Elementary	High School
Boy	- 0.70	- 2.03
Girl	+ 1.49	+ 1.34

Post

	Elementary	High School
Boy	- .92	- 2.13
Girl	+ 1.54	+ 1.57

Source of Variation	Sum of Squares	D.F.	Mean Square	F.
Elementary-Secondary (E)	25.74	1	25.74	1.92
Female-Male (S)	495.33	1	495.33	37.00 *
E x S	21.26	1	21.26	1.58
Error	1499.26	112	13.34	
Pre-Post (P)	.004	1	.004	.002
P x E	.29	1	.29	.15
P x S	1.31	1	1.31	.70
P x E x S	.01	1	.01	.007
Error	209.83	112	1.87	

Scores approaching zero are androgynous.
Scores approaching +4 are feminine.
Scores approaching -4 are masculine.

* $p < .05$

In the educational history unit, affective test results showed a significant F score on the variable sex. No significant differences on the variable treatment (pre-post) nor any significant interactions between sex and treatment were found.

Table 5

Educational History
Affective Test Results #
N=67 cases per cell

	Pre	Post
Male Educator	- 3.74	- 3.63
Female Educator	+ 0.55	+ 0.38

Source of Variation	Sum of Squares	D.F.	Mean Square	F.
Male-Female (S)	1154.49	1	1154.49	138.89 *
Error	1097.18	132	8.31	
Pre-Post (P)	.073	1	.07	.09
P x S	1.41	1	1.41	1.89
Error	98.89	132	.074	

Scores approaching zero are androgynous.
Scores approaching +4 are feminine.
Scores approaching -4 are masculine.

* $p < .001$

In the language arts unit, affective test results showed significant F scores on the variables sex, level (elementary-high school) and treatment (pre-post) and on the interaction of sex, level and treatment. Post hoc analysis found no significant changes at the .05 level for the ratings for elementary and high school girls and for high school boys. However elementary boys were rated significantly more masculine after subjects took the unit.

Table 6

Language Arts Unit
Affective Test Results #
N=34 cases per cell

Pre

	Elementary	High School
Boy	- 1.06	- 2.26
Girl	+ 2.24	+ 1.39

Post

	Elementary	High School
Boy	- 2.04	- 2.83
Girl	- 2.71	+ 0.9

Source of Variation	Sum of Squares	D.F.	Mean Square	F.
Elementary-Secondary (E)	91.69	1	91.69	9.54 *
Male-Female (S)	1009.64	1	1009.64	105.05 **
E x S	1.89	1	1.89	.19
Error	1268.58	132	9.61	
Pre-Post (P)	10.26	1	10.26	4.24 *
P x E	1.31	1	1.31	.54
P x S	9.81	1	9.81	4.06 *
P x E x S	8.08	1	8.08	3.34

Scores approaching zero are androgynous.
Scores approaching +4 are feminine.
Scores approaching -4 are masculine.

* $p < .05$

** $p < .001$

In the social studies unit, affective test results showed a significant F score on the variable sex. No significant differences on the variables level (elementary-high school) or treatment (pre-post) nor any significant interactions were found.

Table 7

Social Studies Unit
Affective Test Results #
N=25 cases per cell

Pre

	Elementary	High School
Boy	- 1.45	- 1.81
Girl	+ 2.40	+ 1.51

Post

	Elementary	High School
Boy	- 1.51	- 2.2
Girl	+ 1.95	+ 1.27

Source of Variation	Sum of Squares	D.F.	Mean Square	F.
Elementary-Secondary (E)	21.60	1	21.60	2.17
Male-Female (S)	620.49	1	620.49	62.56 *
E x S	0.90	1	0.90	0.09
Error	952.02	96	9.91	
Pre-Post (P)	4.035	1	4.035	3.25
P x E	0.04	1	0.04	0.03
P x S	0.18	1	0.18	0.14
P x E x S	0.93	1	0.93	0.75
Error	118.99	96	1.23	

Scores approaching zero are androgynous
Scores approaching +4 are feminine.
Scores approaching -4 are masculine.

* $p < .001$

In the physical education unit, affective test results showed significant F scores on the variables sex, level (elementary-high school) and treatment (pre-post). No significant interactions were found. Post hoc analysis found no significant changes at the .05 level for the ratings for elementary and high school girls and for elementary boys. However, high school boys were rated significantly more masculine after subjects took the unit.

Table 8

Physical Education Unit
Affective Test Results #
N=28 cases per cell

	Pre	
	Elementary	High School
Boy	- .89	- 2.40
Girl	+ 2.17	+ 1.18

	Post	
	Elementary	High School
Boy	- 1.14	- 3.04
Girl	+ 1.98	+ 1.00

Source of Variation	Sum of Squares	D.F.	Mean Square	F.
Elementary-Secondary (E)	100.94	1	100.94	12.26 *
Male-Female (S)	667.37	1	667.37	81.06 **
E x S	7.14	1	7.14	.86
Error	889.07	108	8.23	
Pre-Post (P)	5.46	1	5.46	5.73 *
P x E	.49	1	.49	.51
P x S	.96	1	.96	1.01
P x E x S	.58	1	.58	.61
Error	102.82	108	.95	

Scores approaching zero are androgynous.
Scores approaching +4 are feminine.
Scores approaching -4 are masculine.

* $p < .05$

** $p < .001$

Overall there were few pre-post differences in the affective results. The four significant differences that were found indicated a tendency to rate boys more masculine on the post-test than on the pre-test.

These results must be viewed with caution for several reasons. First, a number of subjects seemed offended by the affective questionnaire. About 10% of

the subjects did not complete the affective tests while less than 1% of the subjects did not complete the cognitive tests. Several people wrote comments indicating dissatisfaction with the affective test. For example, one person wrote "The tape tells us we shouldn't stereotype but then you give us a test that forces us to stereotype." Other people completely filled out the test with the neutral response.

It may be that the math, language arts and physical education units (those units in which a significant difference was found) included information about the characteristics of boy students that made the subjects view boys as possessing a greater degree of masculine characteristics. It may also be that, as one subject wrote "This test has nothing to do with the goals of the tape.", and the results tell us little about the tapes.

Also recent research indicates that changes in teachers stereotypic classroom behavior may not be reflected on their scores on the BSRI. Redd (1976) found that attending several sessions on sex role stereotyping in education significantly reduced teachers stereotypic classroom behavior but did not affect their scores on the BSRI.

In summary, the units significantly increased subjects' knowledge but did little to affect subjects' scores on the BSRI.

For the 1977-78 year, POSSE has retained many of the evaluative components but is incorporating a classroom observation component along with a paper and pencil affective measure. However, the cost and difficulty of conducting observations preclude the widespread use of this method. More research and development work is needed on ways of measuring the impact of social change materials in the affective domain. Until that work is done, a useful evaluation of sex role materials must remain more a pipe dream than a possibility.

Bibliography

Bem, Daryl J. and Sandra Bem, Training the Woman to Know Her Place. Pennsylvania Department of Education, Box 911, Harrisburg, Pennsylvania, 1973.

Bem, Sandra, Sex-role adaptability: One consequence of psychological androgyny. Journal of Personality and Social Psychology, 1975, 31, 634-643.

Bem, Sandra, The measurement of psychological androgyny. Journal of Consulting and Clinical Psychology, 1974, 42, 155-162.

Berkeley Unified School District, The House that Jack and Jill Built. Available from Berkeley Unified School District, Berkeley, CA, 1976.

Campbell, Patricia B. and Charles Thompson, "Little Girls' Place, Little Boys' World", Project 70 Audio Services, Atlanta, 1974.

Campbell, Patricia B., "From Tomboy to Femme Fatale", Project 70 Audio Services, Atlanta, 1975.

Campbell, Patricia B., "Sex Role Awareness Materials for Teachers and Teacher Educators", A proposal funded by the United States Office of Education, 1976.

Ceres Unified School District, Be What You Want to Be. Available from Ceres Unified School District, P.O. Box 307, 6th and Lawrence Streets, Ceres, California 95307.

Emma Willard Task Force on Education, Sexism in Education. Available from Emma Willard Task Force on Education, University Station Box 14229, Minneapolis, Minnesota 55414, 1971.

Findley, Carmen, Beverly Parks and Kathleen Williams, Measures of Educational Equity for Women, American Institute for Research, Palo Alto, 1977.

Granger, Robert C. and Patricia B. Campbell, "The School Psychologist as Program Evaluator", Journal of School Psychology, Vol.15, No.2, 1977, pp.174-183.

Pennsylvania Department of Education, Images of Women, Available from Pennsylvania Department of Education, Box 911, Harrisburg, Pennsylvania, 1973.

Pennsylvania Department of Education, Self Study Guide to Sexism in Schools. Available from Pennsylvania Department of Education, Box 911, Harrisburg, Pennsylvania, 1974.

Redd, A. Loretta, "The Influence of a Sex Role Instructional Unit on the Modification of Attitudes and Behaviors of Elementary School Teachers", University Microfilms, 1977.

Resource Center on Sex Roles, Today's Changing Roles: An Approach to Non-Sexist Teaching, National Foundation for the Improvement of Education, 1974.

Rosenfelt, D.S.; ed., Strong Women, An Annotated Bibliography of Literature for the High School Classroom, The Feminist Press, Box 334, Old Westbury, N.Y., 1976.

Rothchild, Nina, Sexism in Schools, A Handbook for Action. Available from author, 14 Hickory Street, Mahtomedi, Minnesota 55115, 1973.

Stern, Marjorie, ed., Changing Sexist Practices in the Classroom. American Federation of Teachers, AFL-CIO, 1012-14th Street N.W., Washington, D.C. 20005.

Women's Educational Equity Act First Annual Report, United States Office of Education, Washington, 1976.

Women's Educational Equity Act Second Annual Report, United States Office of Education, Washington, 1977.