The purpose of this study was to improve the reliability of the Sigel Cognitive Style Test. Post hoc analysis of ninety test protocols had indicated that the original thirty-five card test could be shortened to improve the test's reliability. This analysis also showed that males were responding to certain cards differently from females. Consequently, two versions of the test were made for field testing. The one for males consisted of twenty cards, for females sixteen cards. There was an overlap of twelve cards common to both versions. One hundred students in two urban, middle-class high schools were tested with the modified versions. Reliability coefficients were .74 for males and .81 for females, compared to coefficients varying from .51 to .70 for the original test. The implications of the study are that several versions of the test should be constructed for other population segments, providing a more reliable diagnostic educational instrument. (Author)
INTRODUCTION

The concept of "Cognitive Style," is employed in different ways by different investigators. Witkin (7) focuses on the field dependent-field independent aspects of organizing one's environment. Gardner et al. (1) describes "style" in terms of leveling-sharpening. In both these cases, the investigators have identified stylistic behaviors and have also indicated the relevance of these behaviors to such things as personal-social characteristics. These researchers have provided opportunities for the test subject to identify or classify various stimuli, but the categorizational background upon which a person bases his selections or classifications is frequently ignored. The concept of "cognitive style" used in this study considers the kinds of cues a person uses in perceiving similarities, and subsequently in categorizing the various dimensions of his environment. In effect, this investigation deals with the individual's "style of categorization."

METHOD

The Sigel Cognitive Style Test (SCST)\(^1\) consists of thirty-five cards of three pictures each of familiar objects, such as toys, furniture, fruit and human figures. The test
subject's task is to select a given pair from a card and indicate his reason(s) for the selection. His responses are then coded into one of six categories: descriptive-part whole (analytical), descriptive-whole (nonanalytical), relational-contextual, categorical-functional, categorical-class naming and categorical-attribute.

Several investigators have reported reliability coefficients for different versions of this "style of categorization" test. Sigel (6) reported odd-even coefficients ranging from .51 to .61 for the subcategories on one set of tasks. Scott and Sigel (5) indicated that test-retest correlations for the Sigel Cognitive Style Test varied from .45 to .83 for the test's six subcategories, with a coefficient of .71 (N=34) for the overall test. Kagan et al. (2) stated that the corrected split-half reliability coefficient based on three hundred protocols was .94. These results were quite acceptable, however it must be mentioned that only one factor, the analytic attitude (descriptive-part whole) was considered here, whereas those reported by Scott and Sigel (5) were for the six subcategories of the test, including the analytic attitude. When all six of the test's subcategories are taken into consideration, the issue of test reliability becomes quite complex. One can, for instance, obtain high reliability coefficients with a set of cards for the
descriptive-part whole subcategory, yet find very low coefficients for the relational-contextual category. The problem then is to obtain acceptable coefficients for all six subcategories. The task is compounded by the fact that the test subject's responses are open-ended, depending on his preferential scheme of categorization.

An attempt was made to solve the issue of test reliability through an item analysis of the original thirty-five card Sigel Cognitive Style Test (SCST). Post hoc analysis of ninety students' protocols from three urban elementary schools indicated that certain cards were causing most of the problems with the reliability. For example, some cards were receiving the same kinds of responses from nearly all of the test subjects, thus loading that subcategory with frequencies. This situation posed two problems: (1) The discriminating ability of "style" was minimized when most persons reacted in identical ways to the same stimuli and (2) odd-even test reliability was affected when the responses from a card in the "odd" section of the test loaded frequencies on one specific subcategory and the "even" card generated frequencies across all six subcategories. These cards were identified and eliminated, shortening the potential test to twenty cards. At this point, it was still evident that a problem existed with regard to reliability. A post hoc analysis that involved the twenty cards in the shortened
version indicated that the corrected split-half reliability coefficient for one group of girls was .85 for the overall test. On the other hand, for a group of boys, the overall test coefficient was .63. Analysis of the responses showed that the boys were giving a variety of responses for a given card, while the girls were tending to make identical responses. Consideration was then given to identifying those cards that would be appropriate for each of the sexes. This analysis indicated that the test for the boys should consist of twenty cards, the test for the girls of sixteen. There was an overlap of twelve cards common to both versions. At this point, the decision was made to field test these two versions.

Because of practical limitations, two questions were considered for this investigation: (1) could the test be shortened from thirty-five to perhaps twenty cards with an acceptable level of reliability? and (2) should a different set of cards be used for each of the sexes in order to achieve higher reliability coefficients? Since urban high school students were available, that segment of the population was selected for this round of field testing of the two modified versions of the test.

DATA SOURCES

Forms HSMCM and HSMCF of the Sigel Cognitive Style Test were administered to four classes in two urban high schools.
Thirty-seven males took Form HSMCM and sixty-three females took Form HSMCF. Form HSMCM consisted of twenty cards and Form HSMCF of sixteen cards.

RESULTS

The corrected split-half correlation coefficients for males (Form HSMCM) were as follows:

- Descriptive-Part whole (analytical) .87
- Descriptive-Whole (nonanalytical) .63
- Relational-Contextual .75
- Categorical-Functional .59
- Categorical-Class naming .73
- Categorical-Attribute .79

The overall test correlation coefficient for this form was .74. The results for the corrected split-half correlation coefficients for females (Form HSMCF) were:

- Descriptive-Part whole (analytical) .92
- Descriptive-Whole (nonanalytical) .81
- Relational-Contextual .79
- Categorical-Functional .68
- Categorical-Class naming .80
- Categorical-Attribute .80

The overall test correlation coefficient for Form HSMCF was .81

DISCUSSION

The results of this investigation indicated that two shortened versions of a style of categorization test could be constructed. These two versions of the test have moderate
reliability for urban high school male and female populations. There are several points that should be discussed. On the positive side, the students who took the shortened versions of the SCST were able to complete the test in 30-40 minutes. Previous experience with high-school-age students and adults had indicated that the original thirty-five card test tended to become tedious, lasting for an hour or more. Since the shortened versions demonstrated moderate reliability, the fact that they could be given within a single class period was a plus factor.

On the problem side, it seems that, with a different version of this test for each population segment, norms will have to be established and made available for persons interested in studying data in this area of evaluation. Once the norms have been established, the procedure for comparing data from two different segments will be complex. Although each group is taking a "style of categorization" test, comparison of the data obtained from each test may be difficult since one will deal with two tests, not one. The direction to be taken is not clear at this juncture. One route might be to administer these two versions of the test used in this study to subjects at different age levels and by socio-economic groups. In this way, the plausibility of using one test for each sex could be determined. Longitudinal studies involving one sex would then be possible.
Comparison of "styles of categorization" between sexes would be difficult but not impossible. Another route would be to develop a single version for use with all groups. This might be the most rewarding since both longitudinal and cross-sectional studies could be done once norms have been established. But is this route possible? Our feeling at this time is that the task would be very difficult with the presently available test material.

CONCLUSION/IMPLICATIONS

In general terms, increasing the length of a test will also increase the reliability. The procedure used in this study was the reverse. The test was shortened to twenty cards in the case of Form HSMCM and to sixteen cards for Form HSMCF. The explanation for this divergence is that apparently the thirty-five-card version was "out of balance." For example, some cards were receiving the same kinds of responses from most test subjects regardless of sex, thus loading one of the subcategories with frequencies. When nearly every person uses the same categorization cues for identical stimuli, not only is the test reliability affected but the concept of each person's "style" has little meaning. For both these reasons, these cards were identified during the item analysis phase and eliminated from the form being considered.
The complexity of the item-analysis phase was increased by the fact that a given card might stimulate the male segment of the test population to give a variety of responses, but the females might tend to give identical responses to that same card. This points up the need for the development of a specific version of the test for different segments of the total population.

Development of a "style of categorization" test that is reliable should be pursued. The beginning effort in this study has indicated that moderately reliable versions of the test are possible, although they are restricted for use by a given population segment. The feasibility of having a reliable test for all segments of the population should be determined. Since styles of categorization are related to science concept achievement (4), reading ability (3), and the dynamics of personality (6), it is thought that consideration must be given to the measurement of a student's "style" as potentially an important part of the school program.
REFERENCES


FOOTNOTE

1. For further information on the SCST, contact I. E. Sigel, University of New York, Buffalo, New York.