ABSTRACT
                Previous research on analogy solutions indicated
that an associative relatedness process was central to the solution
process. One implication of these findings was that differences in
associations to the words comprising the items may produce
differences in test performance and, hence, account for the vast
differentials in performance of persons of various ethnic and
sociocultural groups. This study tested this hypothesis by selecting
two groups of college students of markedly different backgrounds and
with large differences in analogy test scores. Predictability of
analogy item rankings from ranked associative relatedness ratings
within each of the samples fell within the range found in previous
studies. It was not possible, however, to attribute differential test
performances to different associates of the group. (PR)

Associative Relatedness as a Predictor of
Differential Performances on Analogy Items[1]

J. Ronald Gentile
State University of New York at Buffalo

Leon H. Belcher
Texas Southern University

Joel T. Campbell
Educational Testing Service

ABSTRACT

Previous research on analogy solutions indicated that an associa-
tive relatedness process was central to the solution process. One
implication of these findings was that differences in associations to
the words comprising the items may produce differences in test perfor-
mance and, hence, account for the vast differentials in performance
of persons of various ethnic and sociocultural groups. This study tested
this hypothesis by selecting two groups of college students of markedly
different backgrounds and with large differences in analogy test scores.
Predictability of analogy item rankings from ranked associative related-
ness ratings within each of the samples fell within the range found in
previous studies. It was not possible, however, to attribute differ-
ential test performances to different associates of the group.

One of the most difficult challenges to cognitive, as well as to measurement, theorists has been to account for differences in performance on standardized aptitude and achievement tests among various subgroups of the population (such as urban vs. rural, male vs. female, black vs. white, etc.). A primary reason for this difficulty, of course, is that the processes subjects use in solving test items is not well understood, and until a theory of the cognitive processes elicited by test items is developed, it would be difficult, if not impossible, to account for differences in solutions.

Recognizing the importance of studying the processes involved in test item solutions, a program of research was begun to delineate the components of the analogy solution process and thereby provide a theory of problem solving on analogy items (Gentile, 1967 and 1968; Gentile, Kessler and Gentile, 1969). Implicit in these studies was the hope that such a delineation in time would provide an accurate prediction of individual differences in the solution of these items. Analogies of the following form were selected as the items to be studied since they are considered to be measures of reasoning, and since they are quite widely used in standardized tests:

> FIRE: ASHES::
>   1. winter:ice
>   2. tree:leaves
>   3. Christmas:holly
> + 4. event:memories

Although the previously cited studies do not specify the processes of analogical reasoning in any but a pre-theoretical (i.e., incomplete)

3

manner, there was a repeated finding that an associative mechanism

accounted for the major proportion of the variance in the solution pro-

cess. Specifically, rankings of the associative relatedness (the Word

Relatedness Rating Scale, see Gentile and Seibel, 1969) of the four-

word groups which comprise each alternative choice (e.g., FIRE ASHES

WINTER ICE, FIRE ASHES TREE LEAVES, etc.) consistently predicted the

ranked solutions of the analogy items. Supporting a causal interpre-

tation was an experimental study in which primed associates in a pre-

training task were shown to affect subsequent analogy solutions (Gentile

$\underset{\Lambda}{Exp.III}$
et al., 1969). Such data led Gentile et al. to conclude that "the

process that Ss use in solving analogy items is primarily an associative

process" and that the measure of associative relatedness used...appears

to account for from 28% to 50% of the variance in the analogy solutions"

(p. 501).

Assuming that such variance estimates are approximately accurate,

this finding implied a simple explanation for the differential perfor-

mances of different subgroups of the population—namely, that differences

in associations to the words comprising the items are responsible for

the differences in test performance. This suggestion can be inferred

from the work of many investigators (e.g., Bernstein, 1960; Hertzler,

1965; Jensen, 1967; John, 1963;) and is certainly not a new hypothesis.

On the other hand, the analogy studies cited above almost mandated some

test of the hypothesis along the lines of the previous studies. This

paper presents a study designed to yield preliminary data on that question.

METHOD

General Overview. The general hypothesis under test throughout the

several previous studies was that the preferences of alternative choices

of analogy items could be predicted from knowledge of the relatedness

of the words comprising those alternative choices. To measure this,

subjects are required to rank the analogy alternative choices from best

to worst (1 to 4 or 1 to 5 for four- or five-choice items, respectively).

The rankings of the alternative choices obtained from this group by

summing across individuals within items constitute the criterion to be

predicted. These subjects are thus referred to as the criterion group.

Another group of subjects receive the Word Relatedness Rating

Scale (WRRS) in which they are asked to rate on a nine-point scale (from

Extremely-Easy-to-Relate to Very-Difficult-to-Relate) "how difficult it

would be for you to write sentences showing how the words above the rating

scale are related or go together." The words to be rated for relatedness

are the four word groupings which comprise each alternative choice in.

an analogy item. In the sample item, for instance, two of the sets of

words to be rated would be FIRE ASHES WINTER ICE and FIRE ASHES TREE

LEAVES. These ratings on each set of four words are summed across

subjects within each analogy item, and these summed ratings are ranked

from highest to lowest relatedness (1 to 4 or 1 to 5 for four- and

five- choice items, respectively). These group rankings constitute the

predicted rankings for each item, and this group of subjects is called

the predictor group.

Using this general procedure, support for the hypothesis that

different associates are responsible for different analogy solutions

would be obtained if each of the following conditions held. First,

two groups of subjects must be sampled from populations very different

in experiential backgrounds (and presumably therefore in associative re-

sponses to words) and whose performances on analogy items are markedly

different. Second, within each sample, WRRS ratings should reliably

predict analogy solutions within the range of previous studies. Third,
there should be zero order between-sample correlations on the WRRS ratings
and on the analogy item solutions.

Subjects.

Subjects were 98 students from Lycoming College (LC), Williamsport,
Pa. and 104 students from Texas Southern University (TSU) Houston, Texas.
Gross comparisons of the two schools on the only available data indicated
that the socio-economic situations were sufficiently different to jus-
tify the assumption that the two samples for this study were drawn from
different populations. For example, in 1968-69 LC had an enrollment which
included 1.2% blacks, the total estimated cost of an academic year
(including tuition, room and board, books, and fees) was $2700-3200, and
the mean family income of students on aid programs (about $\frac{1}{2}$ of the student
population) was $8797 per year.[2] In contrast, TSU had a 1968-69 enroll-
ment which was 98% black, had a total estimated academic year cost of
$1460, and had 65% of the Freshmen coming from families earning less
than $6000 per year.

The LC students serving as subjects were both underclassmen and
upperclassmen enrolled in four classes: elementary statistics, educ-
cational psychology, developmental psychology, and advanced sociology
(reported in Gentile et al; 1969, Exp. IV). The TSU subjects were all
Freshmen enrolled in a reading and study skills course.

Analogy Items.

Fifteen analogy items, selected from the semi-secure item files of
Educational Testing Service, and previously used by Gentile et al.
(1969) in Exp. IV were used in this study. These items cannot be printed
here because of their semi-secure classification, but they are similar
in form to the sample item in the introduction of this paper, with the

exception that the present items have five alternative choices. One
random order of items was used for all subjects.

Procedure.

In each of the regularly scheduled classes referred to above,
subjects were assigned to the criterion group (LC: N=50; TSU: N=56) or
to the predictor group (LC: N=48; TSU: N=48) by a procedure which in-
sured only that subjects with adjacent seats were in different groups.

As noted above, the criterion group solved the fifteen analogy
items concurrently with the predictor group's rating the word sets on
the WRRS. In addition to the 75 sets of words to be rated by the
predictor group (five in each of the 15 analogy items), six trios of
word pairs selected from the Palermo-Jenkins (1964) norms and ten sets
of word triplets selected from the Nunnally-Koplin (1967) norms were
included for validity data. The Palermo-Jenkins word pairs, which had
been used in the studies referred to previously, were selected from
the norms so that subjects rated a given stimulus word three times:
paired with a high, medium, and low frequency response (e.g., CHAIR
TABLE, CHAIR SEAT, and CHAIR KITCHEN, respectively, were one trio of
word pairs). In addition, trios of word triplets were randomly selected
from the Nunnally-Koplin Joint Meaningfulness table, according to the
same rule, to give a stimulus pair combined with a high, medium, and
low response word (e.g., CORN FIELD FARM, CORN FIELD CROP, and CORN
FIELD ROOTS were one trio of triplets to be rated).

All of these sets of words - 75 four-word sets from the analogy
items, 30 three-word sets from the Nunnally-Koplin norms, and 18 two-
word sets from the Palermo-Jenkins norms - were placed in the rating
booklet in one random order for all subjects. Because of two misspelled
words in the booklet used at LC, one analogy item and one Nunnally-

Koplin set had to be discarded from the analysis, leaving 14 analogy

Items and 9 Nunnally-Koplin sets for the LC analyses only.

RESULTS

As noted earlier, a comparison necessary before the test of the
hypothesis of interest is that of whether the two samples differ in
analogy solutions. Therefore, the proportion of correct solutions
(as Keyed by Educational Testing Service) was obtained for each subject.
The mean proportion of correct responses for the LC sample was .56
(S.D.=.16), while for the TSU sample it was .25 (S.D.=.12), yielding
a t_{96} = 12.775, p <.001.

The predictability for each of the 15 analogy items for the
within-sample, as well as the between-sample, comparisons were all
calculated by the procedure in Table I, which presents the TSU data on
predictability of the analogy item solutions (the criterion) from the
WRRS ratings (the predictor). Table I shows these data cast into a
Friedman Two-Way ANOVA by Ranks table (Siegel, 1956) in which the
heading row is the analogy rankings of the TSU criterion group and
the body of the table presents the TSU predictor group rankings for
each item.

INSERT TABLE I ABOUT HERE

This analysis yielded a χ^2_r = 26.28 (p. <.001, df=4), which indicates
a highly significant relationship between the predicted (by the WRRS
ratings) and actual (from analogy solutions) rankings for the 15 analogy
items for the TSU subjects. The extent of this relationship was esti-
mated by calculating the mean rank-order correlation of these 15 items,
which was .65.

Similar procedures were used to estimate the interrelationships
between the various combinations of LC and TSU predictor and criterion

groups, and the mean correlations are presented in Table 2. Included
in Table 2 as the correlation of variables 3 and 4 is the .65 rho
discussed in the last paragraph. The comparable data for the LC group
is the correlation of variables 1 and 2, which was .53. These results
indicate that the WRRS ratings do predict significantly the analogy
solutions within both the LC and TSU groups. Interpretation of the
remaining correlations in Table 2 are deferred to the discussion.

INSERT TABLE 2 ABOUT HERE

Similar analyses were conducted on the normative pairs of Palermo-
Jenkins and trios of Nunnally-Koplin (presented in Table 3).

INSERT TABLE 3 ABOUT HERE

DISCUSSION

The mean rank-order correlations presented in Table 3 replicate
the previous relations between the WRRS and normative measures of
associative relatedness (Gentile and Seibel, 1969; Gentile et al.,
1969), and add further support to the argument that within both the
LC and TSU students sampled, associative relatedness reliably predicts
analogy solutions. Thus far, then, the results of this study satisfy
the first two conditions necessary (see the General Overview) for the
hypothesis of different associates being responsible for different
solutions to analogy items. The final condition to be satisfied is to
obtain zero-order correlations between-samples on the WRRS ratings of
words in analogy items[3] and on analogy solutions.

Returning to Table 2, however, presents difficulties for the
hypothesis, since the mean correlation between the LC and TSU students
on the WRRS ratings was .72. This indicates good agreement between the
samples on the relatedness of word groups. Further, although there was
a very large difference in proportion of analogies solved correctly
(i.e., ranked 1), there was a significant tendency for both samples to
rank the alternative choices of each analogy item in a similar manner,
as indicated by the mean rho of .49. These two correlations indicate
that there is substantially more agreement between the TSU and LC
students in their ratings of associative relatedness of words than was
expected (supporting the results of a normative study by Belcher and
Campbell, 1968).

The final two correlations to be discussed are extremely interesting
and provocative. The LC WRRS ratings correlate .69 with the TSU analogy
solutions, which is as high as the WRRS ratings of the TSU predictor

group correlated with the analogy solutions of the TSU criterion group:
.65. In contrast, the TSU WRRS ratings show a zero-order correlation
(rho=.25) with the LC analogy solutions. This finding suggests that
the analogy solution process of the TSU students is much more of an
associative process than is the process used by the LC students.

In other words, analogy items appear to be solved by some mechanism
whereby the words comprising each alternative choice are compared along
the dimension of associative relatedness. However, the solution process
of those subjects who score poorly on analogy items includes more of
this associative relatedness mechanism than does the process of those
subjects who serve well (an interpretation suggested by Willner, 1964).
The something extra that the high scorer's do remains to be specified,
concurrently with an elucidation of the remaining dimensions of the
processes required to solve analogy items.

REFERENCES

BELCHER, LEON H. and CAMPBELL, JOEL T., "An Exploratory Study of Word
   Associations of Negro College Students," Psychological Reports 23:119-134;
   1968.

BERNSTEIN, BASIL. "Language and Social Class." British Journal of Sociology
   11:271-276; September 1960.

GENTILE, J. RONALD. "Reasoning, Relationships, and Relatedness of Words:
   A Study of the Effects of Sociocultural Level and Definitions on
   the Solution of Analogy Items," Research Bulletin 67-36, Princeton,
   N. J.: Educational Testing Service, 1967, pp. 1-97.

GENTILE, J. RONALD. "Sociocultural Level and Knowledge of Definitions in
   the Solution of Analogy Items," American Educational Research Journal
   5:626-638; November 1968.

GENTILE, J. RONALD and SEIBEL, ROBERT. "A Rating Scale Measure of Word
   Relatedness," Journal of Verbal Learning and Verbal Behavior 8.252-256;
   April 1969.

GENTILE, J. RONALD, KESSLER, DELORES K. and GENTILE, PATRICIA K. "Process
   of Solving Analogy Items," Journal of Educational Psychology 60:494-502;
   December 1969.

HERTZLER, JOYCE O. A Sociology of Language. New York: Random House, Inc.,
   1965, pp. 368-370.

JENSEN, ARTHUR R. "Social Class and Verbal Learning," In DeCecco, J. P. (Ed.),
   The Psychology f Language, Thought, and Instruction. New York: Holt,
   Rinehart and Winston, Inc., 1967, pp. 103-117.

JOHN, VERA P. "The Intellectual Development of Slum Children: Some Pre-
   liminary Findings." American Journal of Orthopsychiatry 33:813-822;
   October 1963.

NUNNALLY, JUM C. and KOPLIN, JAMES H. "Table H-1, Joint Meaningfulness:
Free Association Norms for 120 Word Pairs," In, The Effects of
Word-Relatedness on Learning (Project No. 2573 Final Report) Nashville:
Vanderbilt University, 1967, pp. H-1 to H-26.

PALERMO, DAVID S. and JENKINS, JAMES J. Word Association Norms: Grade
School Through College. Minneapolis: University of Minnesota Press,
1964, pp. 1-358.

SIEGEL, SYDNEY. Nonparametric Statistics for the Behavioral Sciences. New
York: McGraw-Hill, 1956, pp. 166-172.

Willner, Allen. "An Experimental Analysis of Analogical Reasoning,"
Psychological Reports 15:479-494; Monograph Supplement 3, 1964.

1. Appreciation is extended to Mr. John J. Fremer, Jr. of Educational
   Testing Service for his aid in obtaining the items used in this study;
   to Mr. John G. Hancock, Dr. David J. Loomis, Mr. Lee B. Ross, and
   their students for permission to use their class time to collect these
   data; to Mr. Hunter M. Breland for transcription of data; and to
   Miss Delores Kessler Kennedy and Mrs. Patricia K. Gentile for their
   help during several stages of this study.

2. Presumably the family income of students not receiving aid was higher,
   although data were not available.

3. There is no necessary contradiction between expecting zero-order corre-
   lations between samples on the WRRS ratings of words appearing in
   analogy items and large positive correlations between samples on the
   WRRS ratings of words appearing in the norms, because the latter are
   high frequency words in the language, while the former cover the range
   from very frequent to very infrequent.

TABLE 1

Predictability of Rankings of Alternative Choices on 15 Analogy Items
for the TSU Students

| | Analogy Rankings (Criterion) | | | | | Abs. Diff. | rho |
	1	2	3	4	5		
1	3	5	2	1	4	10	-.200
2*	1	4	3	2	5	4	.675*
3	2	1	4	3	5	4	.800
4	2	1	4	3	5	4	.800
Items 5	2	3	1	4	5	4	.700
6	1	2	5	3	4	4	.700
7	1.5	1.5	4	3	5	3	.875
8	1	2	3	4	5	0	1.000
9	1	4	2	5	3	6	.500
10	3	1	2	4	5	4	.700
11	1	2	3	4	5	0	1.000
12	1	4	3	2	5	4	.600
13	1	2	3	4	5	0	1.000
14	2	3	1	4	5	4	.700
15	2	3	5	4	1	8	-.100
Sum	24.5	38.5	45	50	67		

Note: Criterion rankings (row 1) are based on 48 Texas Southern University students'
analogy solutions. Predictor rankings (body of table) are based on 56 TSU stu-
dents' WRRS ratings. $\chi_r^2 = 26.28$ (p<.001, df = 4). Mean rho = .65.

*Includes a 3.5 tie on the analogy rankings: one is therefore considered a 3,
the other a 4, in a conservative direction for calculation of χ_r^2. The tie
is included in the calculation of rho.

TABLE 2

Mean Rank-Order Correlations of Analogy Solutions and
Word Relatedness Ratings of Lycoming College Students
and Texas Southern University Students

Variables	2	3	4
1. Analogy Solutions (LC) N = 50	.53	.49	$.25^{*}$
2. WRRS (LC) N = 48		.69	.72
3. Analogy Solutions (TSU) N = 48			.65
4. WRRS (TSU) N = 56			

Note: χ^2_r values association with each mean rho provided above were
all highly significant (p<.01 or better) except * (p>.10).

TABLE 3

Mean Rank-Order Correlations of Word Relatedness Ratings and Two
Sets of Norms (Palermo-Jenkins and Nunnally-Koplin) for Lycoming
College Students and Texas Southern University Students

Variables	2	3	4
1. Palermo-Jenkins Norms	--	.75	.92
2. Nunnally-Koplin Norms		.78	.60
3. WRRS (LC) N = 48			.63
4. WRRS (TSU) N = 56			

Note: χ^2_r values associated with each mean rho provided above
significant (p<.05 or better).