DOCUMENT RESUME

ED 196 927 TM 810 017

AUTHOR Carloni, John A.; Kolen, Michael J.
TITLE The Generalizability of Elementary School Student Ratings of Attitudes Toward School Subjects.
PUB DATE Mar 80
NOTE 22p.; Paper presented at the Annual Meeting of the Eastern Educational Research Association (Norfolk, VA, March, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Attitude Measures; Comparative Analysis; Elementary Education; Elementary School Mathematics; Elementary School Science; Item Sampling; Language Attitudes; Rating Scales; Reading Attitudes; Social Studies; *Student Attitudes; *Test Reliability
IDENTIFIERS *Generalizability Theory; Survey of School Attitudes (Hogan); Test Score Variance; *Unit of Analysis Problems

ABSTRACT

Generalizability theory was used to analyze the dependability of elementary school student ratings of attitudes toward school subjects. The rating scales under investigation have been developed to measure the attitudes of students toward four school subjects at both the primary and intermediate levels. Two generalizability coefficients, differing in assumptions of true and error variance, were calculated which estimated the dependability of assessing class attitude. A third coefficient was calculated which estimated the dependability of assessing individual student attitudes. The effects of varying the number of scale items and/or students on the dependability of class means or individual student attitudes were investigated. The results suggest the attitude scales measuring primary student attitudes toward school subjects are not dependable indicators of student attitudes when class is the unit of analysis. There is cogent evidence that both primary and intermediate student attitudes can be dependably evaluated by the attitude scales when the student is the unit of analysis. (Author)

The Generalizability of Elementary School Student Ratings of Attitudes Toward School Subjects*

John A. Carloni and Michael J. Kolen

Hofstra University

Running head: The Generalizability of Elementary School Student Ratings

*Paper presented at the Annual Meeting of the Eastern Educational Research Association in Norfolk, Virginia, March, 1980.

Abstract

Generalizability theory was used to analyze the dependability of elementary school student ratings of attitudes toward school subjects. The rating scales under investigation have been developed to measure the attitudes of students toward four school subjects at both the primary and intermediate levels. Two generalizability coefficients, differing in assumptions of true and error variance, were calculated which estimated the dependability of assessing class attitude. A third coefficient was calculated which estimated the dependability of assessing individual student attitudes. The effects of varying the number of scale items and/or students on the dependability of class means or individual student attitudes were investigated. The results suggest the attitude scales measuring primary student attitudes toward school subjects are not dependable indicators of student attitudes when class is the unit of analysis. There is cogent evidence that both primary and intermediate student attitudes can be dependably evaluated by the attitude scales when the student is the unit of analysis.

The Generalizability of Elementary School Student Ratings of Attitudes Toward School Subjects

Rating scales designed to measure elementary school students' attitudes toward school subjects can provide educators with essential information to assist them in identifying particular students' attitudes, planning instructional programs, determining a class's general attitude toward a curricular area, curriculum evaluation and school attitude research (Hogan, 1975, pp. 4-5). The dependability (reliability) of attitude instruments is an important concern when making decisions of this nature. Factors such as the number of students taking the attitude instrument and the number of items included on the instrument have a pronounced effect on the instrument's dependability for assessing class attitude. The number of items included on the instrument has a pronounced effect on the dependability of assessing individual student attitudes. The manner in which the dependability coefficient is defined is also an important consideration.

Previous research (Gillmore, Kane, and Naccarato, 1978; Kane, Crooks, and Gillmore, 1976; Douglass, Note 1) has demonstrated the usefulness of generalizability theory (Cronbach, Gleser, Nanda, and Rajaratnam, 1972) in analyzing the dependability of student ratings of instruction. These studies

have also addressed a number of concerns in the definition of dependability coefficients

The present study extends the application of generalizability theory to the analysis of the dependability of elementary school student ratings of attitudes toward school subjects. The rating scales under investigation have been developed to measure the attitudes of students toward four school subjects at both the primary (grades 1, 2, and 3) and intermediate (grades 4, 5, 6, 7, and 8) levels. The effects of varying the number of scale items and students on the dependability of individual score and classroom means were assessed for each of the scales at the primary and intermediate levels.

The studies previously cited on student ratings of instruction have considered the class as the unit of analysis. Similarly, in the present investigation the mean rating of a class was considered as the object of measurement in some of the analyses. Other analyses treated the student as the object of measurement since individual student attitude was also considered to be important in the present situation.

The first step in completing a generalizability theory study is to estimate the components of variation from the sample data. This is referred to as the generalizability (G) study. The components of variation can then be arranged to form generalizability coefficients which reflect the depend-

ability of measurement in any of a variety contexts. This second stage is referred to as the decision (D) study. Generalizability coefficients are defined as the ratio of the universe score variance to the expected observed score variance. They have the general form of a reliability coefficient in classical test theory if the universe score variance is defined to be equal to the true score variance. Generalizability coefficients differ according to the definition of the sources of variation constituting universe and expected observed score variance, and the number of sampling units included.

In the present study, students were nested within classes and each student was measured with the same instrument. Hence, the components of variation were estimated using a split-plot design with classes, students and items considered to be random effects.

The estimated components of variation considered for this investigation were,

student, $\hat{\sigma}^2(\pi)$, which is an estimate of the true variance of students nested within classes;

class, $\hat{\sigma}^2(\alpha)$, which is an estimate of true class variance;

item, $\hat{\sigma}^2(\beta)$, which is estimate of the true item variance;

Class x item, $\hat{\sigma}^2(\alpha\beta)$, which is an estimate of the class

by Item Interaction variation;

and residual, $\hat{\sigma}_e^2$, which is an estimate of the error variance.

The generalizability coefficients calculated in this study are presented below.

One generalizability coefficient calculated to estimate the dependability of class means was:

$$\hat{\rho}^2(S,I) = \frac{\hat{\sigma}^2(\alpha)}{\left[\hat{\sigma}^2(\alpha)\right] + \left[\hat{\sigma}^2(\pi)/\underline{n} + \hat{\sigma}^2(\alpha\beta)/\underline{k} + \hat{\sigma}_e^2/\underline{n}\,\underline{k}\right]} \quad , \quad (1)$$

where, $\underline{n}$ = number of students in the class, and

$\underline{k}$ = number of items.

This general reliability generalizability coefficient considers students and items as sources of error, and it represents the expected correlation between two means for each class where the means are based on a different set of items and students.

The second generalizability coefficient calculated to estimate the dependability of classroom means was:

$$\hat{\rho}^2(S,I^*) = \frac{\hat{\sigma}^2(\alpha) + \hat{\sigma}^2(\alpha\beta)/\underline{k}}{\left[\hat{\sigma}^2(\alpha) + \hat{\sigma}^2(\alpha\beta)/\underline{k}\right] + \left[\hat{\sigma}^2(\pi)/\underline{n} + \hat{\sigma}_e^2/\underline{n}\,\underline{k}\right]} \quad . \quad (2)$$

This student reliability generalizability coefficient considers students as a source of error, and it represents

the expected correlation between two means for each class where the means are based on a different set of students on the same items. The derivations of equations 1 and 2 are discussed in Kane and Brennan (1977, pp. 271-278).

In order to estimate the dependability of student scores within a class the following within-class generalizability coefficient was calculated:

$$\hat{\rho}^2_{w\text{-}c} = \frac{\hat{\sigma}^2(\pi)}{\hat{\sigma}^2(\pi) + \hat{\sigma}^2_e/\underline{k}} \quad . \tag{3}$$

This coefficient represents the correlation between scores of different sets of items for students within a given class. It would be expected to be approximately equal to the average within-class coefficient alpha value.

Method

Sample

The sample consisted of 160 first, second, third, fourth, fifth, and sixth grade students belonging to a school district located near the New York Metropolitan area. Eight primary and eight intermediate classes, each containing ten students, were included in the analyses. Students were randomly selected from classes containing more than ten students.

Design and Procedure

The design used for this study was a split-plot design (Winer, 1971) and designated as design V-B by Cronbach et. al.

(1972) in which students are nested within classes and crossed with items. Each student responded to each item in all four scales and the variance components were estimated separately for each scale allowing for comparisons among the reliability coefficients of the four scales. Additionally, an analysis was completed for the sum of the four scales. A separate analysis was performed for the primary level form and the intermediate level form and the resulting generalizability coefficients were compared across levels.

Instrumentation

The rating scale used to measure attitudes towards school subjects was the primary (grades 1, 2, and 3) and intermediate (grades 4, 5, 6, 7, and 8) form A of the Survey of School Attitudes by Thomas P. Hogan, 1975 Edition, published by Harcourt Brace Jovanovich, Inc., New York. This survey consists of four scales of 15 items each in the areas of Reading/Language, Mathematics, Science, and Social Studies. All items are three-point Likert-type items.

Results

The estimated variance components for the primary and intermediate scales are presented in Table 1. Following the suggestion of Kane and Brennan (1977, p. 290), negative estimated variance components have been replaced by zeros.

Insert Table 1 about here

Except for the primary mathematics scale, the remaining primary scales had an estimated class variance, $\hat{\sigma}^2(\alpha)$, equal to zero. The value of the estimated class variance was also equal to zero for the sum of all four primary scales, and the estimated variance for the class by item interaction, $\hat{\sigma}^2(\alpha\beta)$, was equal to zero for the primary social studies scale. The estimated student variance, $\hat{\sigma}^2(\pi)$, was greater than the estimated item variance, $\hat{\sigma}^2(\beta)$, for each of the primary and intermediate scales.

Generalizability coefficients estimating the dependability of class means are presented in Table 2. Coefficients are tabulated for samples of five and ten students and five and fifteen items. This was done to indicate the effects of varying the number of scale items and students on the dependability of class means.

Insert Table 2 about here

Four three of the four primary scales and the sum of scales, the general reliability was zero. The exception was the primary mathematics scale since it had a nonzero estimated class variance.

Except for the primary mathematics and social studies scales, the student reliability tended to decrease with increasing numb of items for the remaining primary scales.

This resulted from an estimated class variance equal to zero in the numerator of the student reliability generalizability coefficient.

For both the intermediate and primary scales, the general reliability coefficient was always smaller than the corresponding student reliability coefficient because it contained only the estimated class variance in the estimated universe score variance. In particular, for small numbers of items, the student reliability coefficient was considerably larger than the corresponding general reliability coefficient for the intermediate reading/language and science scales and the primary math scale.

Increasing the number of students tended to have a greater effect on the magnitudes of the nonzero generalizability coefficients than increasing the number of items, for both the primary and intermediate scales. Furthermore, large increases in items had a relatively small effect on the magnitude of generalizability coefficients. For example, the student reliability coefficient for the intermediate reading scale which was .61 when $n = 10$ and $k = 15$, increased to only .64 when $k = 60$ and n was held constant.

Estimated within-class generalizability coefficients are presented in Table 3 for samples of five, fifteen, and sixty items.

Insert Table 4 about here

Except for science, the primary scale within-class generalizability coefficients were greater than the intermediate scale coefficients for all numbers of items. The estimated within-class generalizability coefficients calculated in this study for samples of fifteen items are presented in Table 4 with the corresponding Coefficient Alpha estimates, r_{kk}, calculated by Hogan (1975 [a], p.8; 1975 [b], p.8). The Coefficient Alpha estimates were calculated from the responses of over 2,000 primary and 4,000 intermediate students. Of special note is the generally close agreement between the corresponding estimated within-class and Alpha Coefficients. The coefficients for the primary scales agreed much more closely then did the intermediate scale coefficients; however, this may have been partially due to the absence of grade 7 and 8 students from the present study.

Discussion

Of the three generalizability coefficients presented in this study the student reliability generalizability coefficient is perhaps the least useful because generalization is usually desired beyond the fixed set of items included on the instrument. The general reliability generalizability coefficient is useful when the class is the unit of analyses

and generalization is desired beyond the finite number of students and items included in the study. The within-class generalizability coefficient is a useful estimate of the dependability of scores when the student is the unit of analysis.

The results of the present study indicate that the estimated item variance is consistently less than the estimated student variance. This result is in agreement with studies that have been done on student ratings of instruction (Gillmore et. al., 1978, p. 1). Thus, increasing the number of items generally has a lesser effect on the dependability of class means than increasing the number of students.

There was a tendency for the primary scales to have estimated class variance at or near zero which was reflected in the low general and student reliability generalizability coefficients. This result suggests that in the primary grades the class or teacher has relatively little effect on student attitudes toward school subjects. Thus, the attitude scales measuring primary student attitudes toward school subjects are not dependable indicators of student attitudes when the class is the unit of analysis. However, there is cogent evidence that both primary and intermediate student attitudes can be dependably evaluated by attitude scales when the student is the unit of analysis.

The present investigation might have been improved by using an attitude scale containing more than three response options per item. This would probably result in relatively larger estimated variance components. Results might also have been more consistent had larger classes been used, and had grades 7 and 8 been included in the analyses.

The SSA (Hogan, 1975) uses three faces (a smiling, neutral, and frowning face) for response options for each item. It would be interesting to complete a study on the effect of this kind of response option compared to other types of response options on the dependability of elementary student responses.

Reference Note

1. Douglass, J. B. Generalizability of behavior-specific vs. general items for student ratings of instruction. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1979.

References

Cronbach, L. J., Gleser, G. C., Nanda, H., and Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley, 1972.

Gillmore, G. M., Kane, M. T., and Naccarato, R. W. The generalizability of student ratings of instruction: Estimation of the teacher and course components. Journal of Educational Measurement, 1978, 15, 1-13.

Hogan, T. P. Survey of school attitudes: Manual for administering and interpreting (Intermediate level form A and B). New York: Harcourt Brace Javanovich, Inc., 1977. (a)

Hogan, T. P. Survey of school attitudes: Manual for administering and interpreting (Primary level forms A and B). New York: Harcourt Brace Javanovich, Inc., 1975. (b)

Kane, M. T., and Brennan, R. L. The generalizability of class means. Review of Educational Research, 1977, 47, 267-292.

Kane, M. T., Gillmore, G. M., and Crooks, T. J. Student evaluations of teaching: The generalizability of class means. Journal of Educational Measurement, 1976, 13, 171-184.

Winer, B. J. Statistical principals in experimental design (2nd ed.). New York: McGraw Hill, 1971.

Table 1

Estimated Variance Components for Primary and Intermediate Scales

Estimated variance component	Reading/ language	Mathematics	Science	Social studies	Sum of scales
Primary					
Class	.0000	.0218	.0000	.0000	.0000
Students within class	.1034	.1505	.0977	.1437	.0956
Items	.0283	.0303	.0112	.0272	.0301
Class x items	.0156	.0215	.0297	.0000	.0183
Residual	.4364	.4291	.4489	.4435	.4677
Intermediate					
Class	.0172	.0042	.0101	.0178	.0102
Students within class	.0886	.1119	.1173	.0989	.0604
Items	.0610	.0252	.0211	.0346	.0370
Class x items	.0228	.0055	.0124	.0056	.0138
Residual	.4339	.4265	.3636	.3906	.4474

Table 2

Estimated Generalizability Coefficients for Reliability of Class Means

Scale	Items per scale: 5		15	
	Number of students: 5	10	5	10
Primary scales				
General reliability				
Reading/language	.00	.00	.00	.00
Mathematics	.30	.44	.37	.53
Science	.00	.00	.00	.00
Social studies	.00	.00	.00	.00
Sum of scales	.00	.00	.00	.00
Student reliability				
Reading/language	.08	.14	.04	.07
Mathematics	.36	.53	.39	.57
Science	.14	.24	.07	.13
Social studies	.00	.00	.00	.00
Sum of scales	.04	.07	.02	.03

Table 2 (continued)

Estimated Generalizability Coefficients for

Reliability of Class Means

Scale	Items per scale: 5		15	
	Number of students: 5	10	5	10
Intermediate scales				
General reliability				
Reading/language	.30	.44	.41	.57
Mathematics	.09	.17	.13	.23
Science	.20	.32	.26	.40
Social studies	.33	.49	.41	.58
Sum of scales	.37	.53	.42	.59
Student reliability				
Reading/language	.38	.56	.44	.61
Mathematics	.12	.21	.14	.25
Science	.25	.40	.28	.44
Social studies	.35	.52	.42	.59
Sum of scales	.40	.57	.43	.61

Table 3

Estimated Within-Class Generalizability Coefficients for Primary and Intermediate Sclaes

	Items per scale		
Scales	5	15	60
Primary			
Reading/language	.54	.78	.93
Mathematics	.64	.84	.95
Science	.52	.77	.93
Social studies	.62	.83	.95
Sum of scales	.80	.92	.98
Intermediate			
Reading/language	.51	.75	.92
Mathematics	.57	.80	.94
Science	.62	.83	.95
Social studies	.56	.79	.94
Sum of scales	.73	.89	.97

Table 4

Comparison of the Within-Class Coefficient in the Present Study With the Results of Hogan (1975 [a], p.8; 1975 [b], p.8)

	Primary scales		Intermediate scales	
Subject	$\hat{\rho}^2_{w\text{-}c}$	r_{α} calculated by Hogan	$\hat{\rho}^2_{w\text{-}c}$	r_{α} calculated by Hogan
Reading/language	.78	.81	.75	.84
Mathematics	.84	.85	.80	.92
Science	.77	.79	.83	.84
Social studies	.83	.84	.79	.84