DOCUMENT RESUME

ED 107 723		TM 004 531

AUTHOR	Borich, Gary D.; Malitz, David
TITLE	Convergent and Discriminant Validation of Three Classroom Observation Systems: A Proposed Model.
NOTE	16p.
EDRS PRICE	MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS	*Behavior; *Classroom Observation Techniques; Comparative Analysis; Evaluation Methods; Interaction; Interaction Process Analysis; *Models; Teacher Education; *Test Validity; Video Tape Recordings

ABSTRACT

Evaluated is the validity of the behavioral categories held in common among three classroom observation systems. The validity model employed was that reported by Campbell and Fiske (1959) which requires that both convergent and discriminant validity be demonstrated. These procedures were applied to data obtained from the videotapes of 62 teacher trainees to ascertain their usefulness and applicability as a model for the validation of classroom observation systems. The validation procedures employed in this study were found to be an economical and useful method for examining the validity of all classroom observation systems. The advantages and limitations of the method employed are discussed. (Author)

ED107723

CONVERGENT AND DISCRIMINANT VALIDATION OF THREE CLASSROOM OBSERVATION SYSTEMS: A PROPOSED MODEL

Gary D. Borich and David Malitz

The University of Texas at Austin

TM 004 581

Convergent and Discriminant Validation of Three Classroom Observation Systems: A Proposed Model

Gary D. Borich and David Malitz

Numerous instruments have been developed to observe systematically classroom behavior. Such instruments typically consist of a number of categories of teacher-student behavior which an observer tallies or rates periodically as he watches classroom interaction. While the reliability of these systems has been investigated, proper evaluation of their validity has been lacking.

The present study undertook to evaluate the validity of selected categories which several classroom observation instruments held in common. The validity model reported by Campbell and Fiske (1959) was employed which requires that both convergent and divergent validity be demonstrated.

Convergent validity is a confirmation of traits (or variables or categories) by independent measuring methods that requires significant correlation between two methods (or systems) measuring the same trait. Discriminant validity is a requirement that "the correlation between different measures measuring the same trait exceed (a) the correlations obtained between that trait and any other trait not having method in common and (b) the correlations between different traits which happen to employ the same method" (Borich and Bauman, 1972). By determining intercorrelations among categories in a multitrait-multimethod matrix, one can identify categories which pass specified tests of convergent and discriminant validity. The procedures were applied to the following data in order to ascertain their usefulness and applicability as a model for the validation of classroom observation systems.

Convergent and Discriminant Validation of Three Classroom Observation Systems: A Proposed Model

Gary D. Borich and David Malitz

The University of Texas at Austin

Evaluated the validity of the behavioral categories in common among three classroom observation systems. The validity model employed was that reported by Campbell and Fiske (1959) which requires that both convergent and discriminant validity be demonstrated. These procedures were applied to data obtained from the videotapes of 62 teacher trainees to ascertain their usefulness and applicability as a model for the validation of classroom observation systems. The validation procedures employed in this study were found to be an economical and useful method for examining the validity of all classroom observation systems. The advantages and limitations of the method employed are discussed.

Method

Data were obtained from a study of 62 teacher trainees at The University of Texas. All but two of the trainees were female. At the end of the student teaching semester, a video tape was made of 20 minutes of each trainee's classroom interaction. The video tape was observed by two judges who rated the interaction using the Interaction Analysis for the Study of Science Teaching, IAST (Hall, 1972), the Fuller Affective Interaction Record, FAIR (Fuller, 1959) and the Classroom Observation Record, COS (Emmer and Peck, 1973). The IAST, FAIR and COS systems are described in Rosenshine and Furst's chapter in the Second Handbook of Research on Teaching (Travers, 1973) and were chosen on the basis of commonalities in the behavior they purport to measure.

Descriptions of the behavior categories of the three systems were obtained from their coding manuals and categories grouped across systems if, from the category descriptions, it appeared that they measured the same behaviors. From these comparisons, 12 IAST categories were paired with nine FAIR categories; four IAST categories were paired with two COS categories; and, across all three systems, seven IAST categories, five FAIR categories and four COS categories were grouped (there were no COS-FAIR pairings which were not included in the three-system grouping). The exact pairings are identified in Tables 1, 2 and 3.

In certain cases, a single variable from one system was paired with several variables in another system. For the purposes of constructing the heterotrait-heteromethod matrix, each comparison can be considered unique, even if several comparisons include the same variable. Thus, in the IAST vs. FAIR comparisons, category H consists of "lecture" (IAST) and "lecture" (FAIR), while category I consists of "review"(IAST) and "lecture" (FAIR), both categories having FAIR's "lecture" category in common.

Once the categories to be investigated had been identified, Pearson product-moment correlations were computed. These correlations were used to construct three multitrait-multimethod matrices: IAST vs. FAIR, IAST vs. COS, and IAST vs. FAIR vs. COS. For each matrix, a heterotrait-heteromethod block was formed with those values in which categories coincide but systems differ. A heterotrait-heteromethod block is illustrated in Fig. 1 with the first two categories of behavior listed in Table 1.

For each matrix, a diagonal (called the validity diagonal) is formed through the heterotrait-heteromethod block by the series of cells in which categories coincide but systems differ. Values in the validity diagonal which are significantly different from zero are evidence for convergent validity. Discriminant validity must be assessed in two steps. First, each validity value must be compared with all values in its row and column in the heterotrait-heteromethod block to determine whether the correlation between different methods of measuring the same category exceeds correlations between that category and other categories not having method in common. In a second step, the heterotrait-monomethod triangles are examined to determine whether the correlation between different methods of measuring the same category exceeds correlations between that category and other categories which have method in common. This step is completed by comparing each category's validity diagonal value with values in the heterotrait-monomethod triangles in which that category is involved. This two-step procedure was carried out for each validity diagonal value in each of the three matrices and the results entered in Tables 1, 2 and 3.

Results

For the comparisons between IAST vs. FAIR shown in Table 1, five validity diagonal values failed to show convergent validity by falling short of the .05 level of significance. These five categories (B, G, I, K and L) also failed to show discriminant validity, as they were exceeded numerous times in their hetero-trait-heteromethod block and in their heterotrait-monomethod triangles. Category F was somewhat inconsistent. It did not show strong discriminant validity but did show convergent validity. The remaining cases, however, (categories A, C, D, E, H and J) present strong cases for both types of validity. All of these categories have significant ($p < .05$) validity diagonal values and most are significant at the .001 level. None of the categories was exceeded by more than one of the 22 values in its row and column in the heteromethod block. Four of the categories (A, C, E and H) were not exceeded by any heteromethod value. Categories C, E and H were not exceeded by any monomethod values while the other categories (A, D, J) were not exceeded by more than three of the 22 values.

Overall, the picture for IAST and FAIR shows that categories C, E and H display excellent convergent and discriminant validities with highly significant ($p < .001$) validity diagonal values and perfect records in the heteromethod blocks and monomethod triangles. Categories A and D and, to a lesser extent, J present strong cases for both types of validity with significant validity diagonal values and good records in the heteromethod blocks and monomethod triangles. Category F is an ambiguous case showing some evidence for convergent and discriminant validity but weaker evidence for discriminant validity. The remaining categories (B, G, I, K and L) show no evidence for either type of validity.

Validities appear quite poor in the comparisons of IAST with COS (Table 2). Of the four comparisons, two comparisons (B and C) produced validity diagonal values which were nonsignificant ($p < .05$). The A and D values did, however, reach the .01 significance level. With four comparisons there are only six values in the heteromethod block and in the monomethod triangles with which the validity diagonal value is compared. Thus, if it is exceeded by any of them, this must count heavily against concluding for discriminant validity. Categories B and C are clearly exceeded too many times to have discriminant validity and categories A and D would also appear to be exceeded too often to have discriminant validity. One must conclude, therefore, that in the comparison of the IAST and COS categories, two show convergent validity (B and C) but none display discriminant validity.

In the three-way comparison of IAST, FAIR and COS categories (Table 3), three categories (E, F and H) show excellent evidence for convergent and discriminant validity across all three systems. All three categories have highly significant ($p < .001$) validity diagonal values in all three comparisons. Categories E and F have perfect records in all three heteromethod blocks and monomethod triangles, while category H is exceeded only once in the heteromethod block of the IAST vs. FAIR comparison. Categories A and C show good evidence for validity across the three systems, although discriminant validity is questionable in the FAIR vs. COS comparison, especially for category A. None of the other categories (B, D, G, I) shows evidence for either kind of validity across all three systems.

Discussion

In the various comparisons across the three systems, a number of categories have been shown not to pass tests for convergent and discriminant validity. The failure of certain categories to demonstrate validity could have been caused by failure of the categories to measure the behavior they purport to measure or by improperly equating categories which, in fact, are not equivalent. It is difficult to say from the data which of these factors was operating for any particular category. Hence, it is impossible to say that any category is invalid; the most one can say is that it failed to demonstrate validity. It should be noted that in most cases, categories which failed to demonstrate validity failed to show either convergent or divergent validity. If a large number of variables had shown convergent validity but failed to show divergent validity, one would suspect that strong method variance was outweighing the category (trait) variance. Yet, it was not high values in the heteromethod blocks or in the monomethod triangles which disqualified most categories; it was low, nonsignificant validity values which were easily exceeded by almost any other value. Some strong, significant values were found in the monomethod triangles (e.g., FAIR's "delves" and "initiates" had a correlation of .59, $p < .001$), indicating that a few of each system's categories are not entirely independent of one another. Yet, generally speaking, the monomethod values were low, so that one could conclude that most categories were measuring some unique behavior.

A number of problems were encountered in applying Campbell and Fiske's model to these data. For this study, a subset of categories was selected from each system because some categories in the three systems did not correspond to one another. Corresponding categories had to be picked out and matched up in order to test validity. Yet, while validity is usually thought of in terms of a category's use within its system as a whole, validity was actually tested against the subset. The nature of the test for discriminant validity (comparing one value with a series of other values) makes it more difficult to demonstrate discriminant validities when a large number of categories is being compared. Because each value was compared with a subset of the possible values, it was easier for each value to pass the discriminant validity test than it would have been if all system categories had been compared. This may have given some categories the appearance of discriminant validity which they would not have in the context of their complete system.

Another problem with the Campbell-Fiske method was encountered when one category from one system was paired with several, almost identical, categories in another system. When one pairs categories, one is hypothesizing that the two categories measure the same behavior, i.e., that they will demonstrate convergent validity. But, at the same time, one is hypothesizing that each of the paired categories differs from other categories in its own system and in the other system. In other words, a hypothesis about convergent validity necessarily includes a hypothesis about discriminant validity. It was this second hypothesis which caused trouble, for when the same category appeared in two pairings, it appeared as two "independent" categories in its system. Obviously, when these two "independent" categories were correlated in the monomethod triangle, a value of 1.00 was obtained, precluding any demonstration of discriminant validity for that category. When the "independent" categories were correlated with each of the categories in the other system, duplicate columns or rows appeared in the heteromethod blocks and the monomethod triangles.

To circumvent these difficulties, the correlations of 1.00 in the monomethod triangles were ignored, for in the special case of duplicate categories, a test for the independence of these categories from each other is impossible. In all other respects, however, these duplicate categories were treated like all other categories, for each was a component of a unique pairing with another system's category.

Across the three systems, the results of the study are not encouraging for researchers who choose to measure classroom interaction. One must infer from these results that, of the 88 observational coding systems described by Simon and Boyer (1970), many probably do not meet the standards of convergent and discriminant validity that were proposed in this study. The researcher must be cautious in drawing relationships between research studies which use classroom interaction systems for which the measurement technique itself accounts for greater variation than the behavior being measured or when the same behaviors measured by different systems fail to correlate. Such findings suggest that the descriptive titles of categories and behavioral constructs employed by many observational coding systems may not adequately represent the behavior they purport to measure. The validation procedures employed in this study were found to constitute potentially an economical and useful model for examining the validity of other classroom observation systems.

References

Borich, G. D. & Bauman, P. M. Convergent and discriminant validation of the French and Guilford-Zimmerman spatial orientation and spatial visualization factors. Educational and Psychological Measurement, 1972, 32, 1029-33.

Campbell, D. T. & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Emmer, E. T. & Peck, R. F. Dimensions of classroom behavior. Journal of Educational Psychology, 1973, 64, 223-240.

Fuller, F. F. FAIR System Manual. Austin, Texas: Research and Development Center for Teacher Education, The University of Texas, 1969.

Hall, G. E. A Manual for Users of the IAST: A System of Interaction Analyses. Austin, Texas: Research and Development Center for Teacher Education, The University of Texas, 1972.

Simon, A. & Boyer, E. G. (Eds.) Mirrors for Behavior. Philadelphia: Research for Better Schools, Inc., 1970.

Travers, R. M. W., (Ed.) Second Handbook of Research on Teaching. Chicago: Rand McNally and Co., 1973.

		IAST accepts A	IAST questions B	FAIR values A	FAIR delves B
IAST	A	(.16)*			
	B	.23	(.70)		
FAIR	A	.43	-.10	(.58)	
	B	-.12	-.01	-.14	(.84)

*Interjudge reliabilities.

Figure 1. Simplified Illustration of the Validation Model.

The validity diagonal = .43, -.01; the heterotrait-heteromethod block = .43, -.01, -.10, -.12. The monomethod triangles = .23 and -.14, respectively.

Table 1. Validities of Variables
from the IAST and FAIR Classroom Observation Systems
N = 62

Variable Names IAST/FAIR		Validity Diagonal Value	Convergent Validity: Highest Value in Heteromethod	Convergent Validity: No. Higher	Discriminant Validity: Highest Value in Monomethod	Discriminant Validity: No. Higher
accepts feelings/values	A	.429	.272	0	.539	2
questions student's stmt./delves	B	-.011	.701	16	.595	19
confirms student's stmt./OK	C	.812	.259	0	.306	0
open question/initiates	D	.825	.701	0	.595	0
criticizes/criticizes	E	.904	.299	0	.549	0
looks at notes/tangential	F	.253	.272	1	.539	3
non-functional behavior/woolgathering	G	.006	.234	19	-.268	18
lecture/lecture	H	.713	-.250	0	-.308	0
review/lecture	I	-.135	.713	6	.336	4
read aloud/lecture	J	.413	.713	1	.549	1
substantive closed stmt./questions	K	.038	-.317	18	-.268	18
substantive open stmt./questions	L	.088	.225	8	-.268	5

Table 2. Validities of Variables from the IAST and COS Classroom Observation Systems
N = 62

Variable Name (IAST/COS)	Value	Convergent Validity: Highest Value in Heteromethod	Convergent Validity: No. Higher	Discriminant Validity: Highest Value in Monomethod	Discriminant Validity: No. Higher
closed question/convergent eval.	.3375	-.4979	2	-.2991	0
closed student stmt./converg. eval.	.1720	-.4979	3	-.2991	2
open question/higher cognitive	.2431	.4415	3	.5241	3
open student stmt./higher cognitive	.4415	.4979	1	.5241	1

Table 3. Validities of Variables from IAST, FAIR, and COS Classroom Observation Systems

Variable Name (IAST/FAIR/COS)	IAST vs. FAIR					IAST vs. COS					FAIR vs. COS				
		Converg. Valid.		Disc. Valid.			Converg. Valid.		Disc. Valid.			Converg. Valid.		Disc. Valid.	
	Value	Highest Value in Hetero.	No. Higher	Highest Value in Mono.	No. Higher	Value	Highest Value in Hetero.	No. Higher	Highest Value in Mono.	No. Higher	Value	Highest Value in Hetero.	No. Higher	Highest Value in Mono.	No. Higher
accepts feelings/values/ positive affective A	.4292	.2003	0	.2497	0	.3819	-.2927	0	-.3610	0	.2585	-.3789	1	-.3610	1
question stmt./delves/ teach init. prob. B	-.0113	-.2675	8	.5945	10	-.0018	.6524	9	.3360	9	.5169	.6921	1	.5945	1
open question/delves/ teach init. prob. C	.7011	.8255	1	.5945	0	.6524	-.2981	0	-.3075	0	.5169	.6921	1	.5945	1
question stmt./initiates/teach init. prob. D	-.0048	.8255	10	.5945	10	-.0018	.6524	9	.2536	9	.6921	-.5025	0	.5945	0
open question/initiates/teach init. prob. E	.8255	.7011	0	.5945	0	.6524	-.2981	0	-.3075	0	.6921	-.5025	0	.5945	0
looks at notes/lectures/ teach present F	.7132	.4127	0	-.3075	0	.5869	.4249	0	-.3610	0	.8980	-.3789	0	-.3610	0
reviews/lectures/ teach present G	-.1349	.7132	2	.3360	3	-.0658	.5869	8	-.3610	9	.8980	-.3789	0	-.3610	0
reads aloud/lectures/ teach present H	.4127	.7132	1	-.1956	0	.4249	.5869	1	-.3160	0	.8980	-.3789	0	-.3610	0
non-functional behavior/ woolgathering/level of attention I	-.0060	.2340	9	.2536	9	-.2185	-.2981	2	.2536	1	-.0082	-.5025	6	.1419	7