Studies in Educational Evaluation. Vol. 11, pp. 339–357, 1985
Printed in Great Britain.

0191-491X/85 $0.00 + .50

AN EXPLORATION OF THE STRUCTURE OF CLASSROOM BEHAVIOR DURING VALUES DEVELOPMENT LESSONS

A. L. Beem and D. Brugman

Rijksuniversiteit te Leiden

In the past two decades many educators have been concerned with values education. The two main approaches to values education are Values Clarification (Kirschenbaum, 1973, 1977; Raths, Harmin & Simon, 1966) and Moral Development (Kohlberg, 1973, 1975).

The present approach to values education, Values Development, is an educational application of the valuation theory of Hermans (1976), and to some extent integrates both approaches (Beem & Brugman, LICOR report).

In Values Development a value (value area in Hermans' terminology) is defined as anything that pupils find of importance in their life situation. The valuation theory provides a method for investigating the values of a person by this person, by cycling through three processes.

As an important aspect of education is to acquaint pupils with values of other persons, groups and institutions (cf. Rokeach, 1979), these processes were extended to enhance the relevance of the theory for educational applications. Hence, the following valuing processes are distinguished:

1. Making explicit one's own values and feelings.

2. Making explicit values and feelings of other persons and institutions (role-taking).

3.a. Relating values to each other within one's own value system.

3.b. Relating values of different value system to each other, one of which may be one's own.

This research was supported by grant 312 from the Dutch Foundation for Educational Research, S.V.O. We are indebted to Dr. P.L. van der Plas and Dr. D. Bakker for their substantial contributions to the research project, and to Prof. Dr. F.K. Kieviet for his supervision. We thank Dr. I. Smit for correcting the English and M. Cornelissen-Prins for her endurance in typing the manuscript.
Requests for reprints should be sent to A.L. Beem or D. Brugman, LICOR, Rijksuniversiteit Leiden, Stationsplein 12 (kamer 411), 2312 AK Leiden, Holland.

4. Validating values and feelings.

The Values Development program aimed at stimulating the execution of the valuing processes by the pupils. The program was developed in the course of a research project, which was phased according to a descriptive-correlational-experimental loop (Rosenshine & Furst, 1973). Based on results of the correlational phases, all teachers received written training material containing information about behavior expected to stimulate the execution of the processes.

From the results of these earlier phases, described in detail in van der Plas (1981), it is to be expected that indirect instruction (or open teaching) is more effective in this respect. Moreover, Peterson (1979) concluded after reviewing research findings that open teaching, as opposed to direct or traditional approaches, may be more effective for some cognitive outcomes (creativity, problem solving), and for affective outcomes, such as attitudes toward school and teacher. Simon, Howe and Kirschenbaum (1972) have suggested that the use of Values Clarification methods requires teachers to encourage a classroom atmosphere of openness, honesty and respect, to be the model of a good listener, and to show interest in what pupils think. Raths et al. (1966) advised teachers to make their ideas, feelings and opinions known to students.

Lockwood (1978) noticed, that some teachers may be better than others in establishing such an atmosphere. Because the suggestions made are not based on empirical research, he recommended to conduct research on the influence of teacher variables. But the variables mentioned above obviously have a high inference character and are, therefore, not very specific in informing teachers what to do. Therefore, the present research investigates relationships between low inference measures of teacher and pupil verbal behavior during the lessons in order to identify teacher behavior that appears effective to the stimulation of the valuing processes.

The verbal behavior of teachers and pupils was registered with an observation instrument on two occasions: in the first month of the program when the lessons had just been introduced, and several months later in the last month. It is well known (cf. Brophy, 1979) that no single teacher variable will be substantially related to pupils variables. Teacher effectiveness was therefore investigated by analyzing the structure of behavior of teachers and pupils in the first and last month, using principal component analysis. It has also been suggested by Rosenshine and Furst (1973) that behavioral changes of one kind will be accompanied by others. In other words, a change of pattern is to be expected between the first and the last month, if some teachers are more effective than others. Changes in structure were therefore investigated by means of three-mode principal component analysis, which is particularly suited to analyze a change in pattern.

The three-mode analysis gives information about the structure on the first and second occasion. But results of ordinary principal component analyses are also presented because: (a) in a sense they provide an optimal rotation for each occasion, and (b) they give a better approximation to the data with the same number of components. In order to give an impression of the general level of behavioral categories, mean scores and their change will also be presented.

METHOD

CLASSES

Among a group of teachers who had shown interest in the material to be used during the lessons, 33 (15 from elementary and 18 from secondary school) were willing to participate in the research. The interest was a result of various activities, such as a publication about the project, so that the sample was not drawn at random. A random sample was impossible as classes could not be compelled to participate. The classes contained about an equal number of pupils. The age of pupils ranged from 10 to 15 years, with an average of 12. The schools were situated in or around Leiden. In the end 25 classes (12 from elementary and 13 from secondary schools) were left from which sufficient data were available for the analyses to be described. We have no reason to suppose that these classes are not representative for the original group.

PROCEDURE

The lessons were given bi-weekly from the beginning of October until March, but at least ten lessons were given during this period. A lesson lasted about one hour. In general, the pupils together chose one lesson from the workbook every time. Two lessons in October, however, were the same for all classes, but in random order. The same applied to two lessons given in March. The lessons were videotaped and the behavior of teacher and pupils during classroom dialogues was subsequently scored by trained observers, using the observation instrument and scoring apparatus to be described below.

Teachers were provided with written training material. The material contained information about the theoretical background of the lessons and ways to achieve teacher behavior that was expected to have a favorable effect on values development. Teachers were divided randomly in two groups: One group worked through suf-support on these materials; the other group received supervision on four lessons in an indirect supervision style based on systematic observation. Between these groups, there were no statistically significant differences in frequency of behaviors (no significance level was less than .25). Details are discussed elsewhere (Brugman & Beem, 1982).

THE CURRICULUM VALUES DEVELOPMENT

The curriculum consists of a pupil workbook, called "Everything about myself and others", two instructional aids, and exercise units for the teacher. The units were constructed in response to problems teachers experienced during earlier phases of the project. The pupil workbook contains 15 values development lessons. The topics refer to the experiential world of the pupils: school, home, friendship, agression, holidays, money, rules, animals, hobbies, etc. The workbook contains so-called value sheets, to stimulate values development at the individual level.

One of the two instructional aids consists of a handbook with information on the theory and practice of values development. Special attention is paid to five rules of thumb for the teacher: clarify your intention, approach the needs of the students, be personal, propose attractive working forms, and strive for a symmetrical relation. The other instructional aid gives specific suggestions and possibilities about the way each of the fifteen lessons could be carried out.

The exercise units contain examples and general descriptions to clarify important concepts and skills in values education: values and feelings in the classroom, positive feedback, active listening, and open questioning. These descriptions are followed by several exercises: (1) teachers must judge and classify verbal teacher and pupil behavior in written examples of classroom dialogues. This is followed by written feedback on these judgments; (2) teachers are stimulated with specific questions to reflect on behavior of themselves and others both in and outside the classroom (i.e. in daily conversations, e.g., did pupils express many values and feelings?); (3) teachers are asked to complement and fit a sequence of target pupil behaviors (cf. Flanders, 1970), considered as important in values education (e.g., to find an appropriate question that could precede a given pupil's expression of feelings).

The teachers's task is, in the first place, to help pupils made explicit their own values and feelings about the topics, as well as those of others, as explicitly as possible in written statements or in verbal statements during classroom dialogues. Moreover, the teacher will try to stimulate further reflection on these values and feelings: make pupils search for similarities, differences and causalities between different values, to state priorities in value dilemmas, and to reflect upon the moral side of values.

THE OBSERVATION INSTRUMENT

The observation instrument used to register classroom behavior consists of 22 categories, whose content and shorthand descriptions used in the text are described in Appendix A. The categories are mutually exclusive and exhaustive. Teacher categories are generally the same as the FIAC categories (Flanders, 1970).

Like Casteel and Stahl (1975), we assumed that a verbal expression is an indication of the occurrence of a process. Category 14 (expression of personal values) refers to the first valuing process. Category 17 (values of others) refers to the second valuing process. These are affective and social aspects of values development. Category 15 (personal behavior description) and 16 (gives opinion) refer to behavioral and cognitive aspects of these processes. Alternating occurrence of 14 and 17 refers to the third valuing process. The instrument is a final version developed from more detailed instruments. This version does not contain a category for validating values and feelings, because this behavior did not occur during classroom dialogues in former observational studies.

SCORING APPARATUS AND RELIABILITY OF THE OBSERVATION INSTRUMENT

The observers scored classroom behavior in 22 categories by pressing buttons, numbered 1–9, in a keyboard. A category number was automatically linked to time, in seconds, from the start of the lesson. The observers could control and correct their scoring quite easily, because the category number last adjudged to behavior was shown on the screen.

The scores for each category were constructed as follows. For each observer, the percentage of time during classroom dialogue that a class spent in a particular category of the observation instrument was computed. These proportions p_i were transformed by the arcsine transformation: $2\sin^{-2} p_i$. The mean was then computed over observers, both for the first two and for the last two lessons. These averaged scores are referred to,

respectively, as the first and second occasion. To estimate reliability of the scores, a generalizability study was conducted (Cronbach, Bleser, Nanda, & Rajaratnam, 1972): four of the seven observers scored ten of the classes during three of the four videotaped lessons. Observers, classes and lessons were randomized and their effects were evaluated via a completely crossed factorial design.

With small samples it is not possible to estimate variance components for infinite populations accurately. Therefore, the populations were restricted to the four videotaped lessons and the teachers and observers that participated in the experiment. Formulas for estimating variance components for finite populations are given by Searle and Fawcett (1970). Because we are interested in the reliability of mean scores over the same lessons for each class, lessons were treated as fixed in estimating reliability, and because two new observers were randomly allocated to a lesson, observers were treated as random. The reliability estimates are given in Table 1.

Elffers and Tavecchio (1979) suggest that the reliability should be at least .65, a value deduced from probabilities of wrong classifications. With the exception of the categories 1, 17, and 21, the reliabilities exceed this value. In general the reliabilities are quite satisfactory.

TABLE 1: Means for Categories at Two Occasions and Reliabilities

Categories	First occasion	Second occasion	P-value*	Omega squared	Reliabilities
1.Asks open questions	.21	.24	.12	.06	.55
2.Asks closed questions	.30	.34	.07	.09	.84
3.Probes	.35	.38	.25	.01	.91
4.Redirects	.25	.31	.03	.15	.73
5.Gives direction	.87	.73	0	.46	.79
6.Lectures	.80	.70	.02	.18	.82
7.Being personal	.50	.41	.05	.11	.98
8.Summarizes	.27	.35	.03	.15	.67
9.Accepts	.47	.57	0	.34	.71
10.Praises	.12	.14	.34	0	.83
11.Corrective feedback	.23	.22	.23	.02	.78
12.Sharp negative feedback	.26	.30	.08	.08	.91
13.Irrelevant behavior	.14	.10	.06	.10	.73
14.Expresses personal values	.30	.40	0	.21	.92
15.Personal behavior	.65	.61	.39	0	.94
16.Gives opinion	.22	.30	0	.22	.73
17.Refers to values of others	.17	.28	0	.32	.49
18.Gives information	.45	.59	0	.39	.84
19.Asks questions	.15	.16	.54	0	.86
20.Structured working	.50	.40	.06	.10	.69
21.Unordered behavior	.37	.22	0	.32	.58
22.Silence	.38	.45	.05	.11	.71

* Descriptive significance level of t-test for paired observations.

EXPLORATION OF THE STRUCTURE OF CLASSROOM BEHAVIOR

The structure was explored by means of principal component analysis (BMDP4M, Dixon, 1981). The presence of outliers was first investigated (Hawkins, 1980) because, especially in small samples, these may unduly invluence the analysis. The identified outliers are included in the analyses because they turned out only to strengthen existing tendencies in the data.

An interpretational matter should be discussed briefly. For convenience, clusters of variables or directions in the space spanned by the components will be labeled. This does not imply, however, that a latent variable or theoretical construct is postulated. The labeling will be largely done in accordance with Flanders (1970). Thus, the teacher categories 5, 6, 11, 12 and 1, 4, 9, 10 will be labeled directive and non-directive behavior, respectively. The relationship with pupils' categories will be used to qualify the labels (e.g., the categories 14 and 17 are called affective, the categories 15, 16, 18, 19 cognitive). The meaning of the other teacher categories is derived partly from their position in the structure and partly from their content.

Results will be presented for the first occasion, the second occasion, and for the change from first to second occasion successively. For all analyses, scores were standardized (mean zero, standard deviation one). Consequently, mutual positions of classes are relative, not absolute. The correlation matrices analysed are given in Table 2.

TABLE 2: Correlations Between Categories on First (below diagonal) and Second (above diagonal) Occasion; Diagonal Contains the Correlations Between the First and Second Occasion for each Category

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	43	34	03	43	-19	-24	-37	-39	26	0	-28	-26	-21	28	22	61	37	18	-10	32	-04	-42
2	02	24	7	17	-05	-03	-10	-29	30	-02	-15	-05	18	-05	-11	-07	-15	0	19	18	-03	09
3	05	02	49	27	-24	-16	-03	16	34	15	-44	-27	04	16	01	-01	-16	29	-24	-21	-41	-18
4	67	22	47	04	-57	-19	-28	20	32	-02	-10	-12	0	27	34	68	18	14	04	-03	-30	-17
5	-34	03	-38	-57	62	05	-33	04	04	-02	31	36	08	-23	-18	-21	-18	-27	01	03	25	-11
6	-07	-02	-21	-28	-10	26	35	-14	24	-32	36	32	06	-67	-44	-31	-53	-42	28	-16	16	14
7	-10	38	-41	-18	04	-02	45	-22	-08	-05	07	-28	-16	-27	-27	-35	-22	-36	-06	-01	05	17
8	-07	29	-15	-05	-08	25	15	0	05	17	15	27	02	-11	-03	10	08	-01	-13	-20	-01	23
9	36	41	61	70	-55	-03	02	18	47	12	07	-09	-01	-29	-05	11	-16	-24	17	0	-17	-35
10	40	13	31	39	-09	-28	-01	10	21	41	-03	-15	-19	-02	01	-06	27	23	02	50	-02	-09
11	-10	-24	-41	-31	37	-21	15	-34	-47	-16	54	62	37	-55	-41	12	-21	-60	06	0	44	09
12	-03	-30	-60	-45	36	-03	16	-02	-54	-01	71	71	34	-56	-40	-07	-34	-40	-01	-32	65	22
13	-10	-17	14	-13	05	-35	-12	-52	-33	-13	13	-11	35	-07	11	01	-31	-30	04	-18	-11	20
14	19	28	46	41	-12	-24	-18	26	31	55	-23	-33	-13	20	64	23	43	48	06	01	-43	-25
15	32	-20	51	63	-54	-30	-43	-02	36	36	-37	-39	02	42	36	34	36	23	19	-07	-59	-45
16	01	17	12	12	-16	-37	-08	09	-03	20	-16	-38	45	30	18	22	32	05	-24	20	-12	-41
17	42	-03	45	57	-40	-09	-07	34	58	33	-37	-39	-33	43	41	09	04	27	-20	35	-16	-19
18	35	-30	61	47	-51	-25	-45	-32	23	11	-28	-52	48	24	61	34	39	17	09	08	-31	01
19	26	-36	-04	05	-17	-01	-30	-04	-12	24	05	30	31	-15	26	15	02	22	28	-09	-10	02
20	-34	15	-32	-35	03	39	13	-15	-34	-54	-20	-20	15	-41	-35	-04	-45	-10	-30	-10	-21	04
21	-07	-37	-42	-40	21	-10	02	-38	-44	-48	57	60	35	-68	-49	-16	-47	-10	28	18	40	19
22	-37	09	-07	01	-09	-29	28	08	-03	-11	14	-03	08	13	10	29	-01	03	-08	-02	-10	28

Note: decimal point omitted.

RESULTS FOR FIRST OCCASION

General level of behavior. The means are presented in Table 1. These scores are averages of arcsine transformations, which are not exactly transformable to averaged proportions, but they can be well approximated. From Winer (1971, p. 872), the averages of .20, .40, .60, .80 are nearly equivalent to the averaged proportions .01, .04, .09, .15. The valuing processes seem to be executed (categories 14-17), but cognitive pupil behavior (categories 15, 18 and 20) dominates. Teacher behavior in general occurs twice as often than pupil behavior.

Conclusions of structural analysis. General conclusions are stated first.

1. Classes can be most clearly divided into two opposite types: those where teacher encourage pupils to speak and take initiative, and those where teacher are commanding, steering, and discouraging pupil behavior. These classes may be labeled non-directive, cognitive as well as affective, versus directive, disordered. In a non-directive and encouraging climate there is, relatively much pupil behavior that is related to values development, cognitive as well as affective. A directive, discouraging climate is related to relatively much unordered pupils behavior.

2. A less clear division can be made into two other opposing types, both of which do not contrast sharply with the first two types: classes with a cognitive, teacher-centered, more structured climate and with a cognitive, pupil-centered, less structured climate.

3. Within classes with a non-directive, encouraging climate, relating of values (i.e., expression of values and referring to values of others) occurs more frequently if the teacher concentrates slightly more on structuring feedback.

4. The following teacher behavior has a positive effect on pupils' behavior related to values development: asking open questions, redirecting, summarizing, accepting, praising. Clearly, negatively related are lecturing, giving directions, and giving corrective and sharp negative feedback. Not effective are asking closed questions and being personal. If teachers display, relatively, much irrelevant behavior, pupils display only cognitive behavior.

Details of analysis. Loadings of the categories on the first five principal components and loadings after a direct oblimin rotation are presented in Table 3. The varimax rotation did not yield such a simple loading pattern. Strong loadings, especially important for the conclusion, are underlined. The five components account for 70% of the variance. The first component is clearly the most important.

TABLE 3: Category Loadings at Two Occasions

	First occasion											Second occasion										
	Loadings on principal components					Loadings after direct oblimin rotation						Loadings on principal components					Loadings after varimax rotation					
Categories	1	2	3	4	5	1	2	3	4	5	H^{2a}	1	2	3	4	5	1	2	3	4	5	H^{2a}
1. Asks open questions	48	18	45	-29	39	81	-02	32	-07	-17	70	57	50	05	45	-13	12	69	07	32	-44	80
2. Asks closed questions	18	-63	-16	33	32	16	16	-16	76	13	67	07	19	42	34	-14	-11	12	19	02	-54	35
3. Probes	74	18	-22	-03	-12	35	18	-43	-28	23	64	33	-10	47	-30	51	20	-04	79	-02	18	69
4. Redirects	82	07	14	-07	41	89	08	-02	09	12	86	52	54	26	-22	24	09	65	55	-08	06	74
5. Gives derection	-60	-09	14	31	-31	-70	17	27	02	04	57	-41	12	-34	08	-26	-30	-01	-53	-08	03	37
6. Lectures	-21	-41	-13	-67	-23	-17	-05	-37	-06	-77	73	-66	05	47	15	-04	-58	-38	11	-27	-34	68
7. Being personal	-28	-48	06	32	53	11	-13	27	83	07	70	-31	-49	32	26	16	-12	-65	18	15	-21	54
8. Summarizes	16	-64	21	04	-34	-21	67	-12	17	-30	59	-08	11	-28	-57	57	-14	07	22	-07	82	75
9. Accepts	74	-30	03	-16	36	78	15	-22	30	-10	80	03	47	58	20	31	-39	27	61	03	-31	70
10. Praises	52	01	50	29	-16	20	65	28	-14	18	63	21	-11	-27	31	51	02	01	12	65	21	48
11. Corrective feedback	-59	31	41	28	21	-17	-21	75	05	17	73	-67	47	-22	02	03	-76	10	-31	-15	15	72
12. Sharp negative feedback	-66	16	66	04	01	-28	-01	83	-08	-21	91	-71	43	-27	-26	-01	-68	07	-33	-36	33	82
13. Irrelevant behavior	-11	71	-44	28	07	-12	-50	-08	-28	65	79	-26	27	15	-43	-26	-17	11	-04	-61	05	41
14. Expresses personal values	65	-16	13	42	-30	03	73	-13	-04	33	73	78	-20	-13	-24	-34	86	29	-05	-05	0	83
15. Personal behavior	75	32	03	-04	-15	40	27	-20	-43	21	69	69	07	01	-32	-39	68	44	02	-25	-07	73
16. Gives opinion	30	23	-31	56	-10	-11	12	-20	-04	70	56	48	71	-17	-02	07	01	86	10	10	09	77
17. Refers to values of others	73	-14	23	-11	0	52	43	-10	-06	-09	62	59	03	-52	18	06	41	40	-20	51	17	66
18. Gives information	61	61	-32	-10	04	45	-22	-34	-47	37	86	61	-36	-11	-10	05	67	02	12	21	13	52
19. Asks questions	05	57	34	-16	-17	07	0	31	-61	0	50	-07	-05	33	02	-43	07	-09	-05	-33	-42	29
20. Structured working	-42	-18	-72	-25	09	-27	-58	-54	23	-14	78	25	0	-25	70	16	0	14	-12	74	-23	64
21. Unordered	-68	49	08	-16	31	-06	-70	46	-11	-04	84	-61	17	-39	19	0	-75	-09	-47	08	13	58
22. Silence	01	-04	-19	62	11	-15	06	01	33	58	43	-42	-38	-10	-14	14	-13	-52	-10	-04	27	37
% explained variance	28	14	11	10	7							23	11	10	9	8	18	14	10	10	9	

Note. decimal point omitted.

a. communalities.

The first and second conclusion are based on the loadings on the first and second principal component, respectively. The label non-directive versus directive refers to the loadings of the categories 1, 3, 4, 9, 10, 14, 15, 17, 18 versus 5, 11, 12, 21 on the first component. The label teacher-centered, structured, is used for the loadings on the second component of the teacher categories 2 (asks closed questions) and 8 (summarizing), as these categories seem neither directive nor non-directive, but are associated with more structuring teacher talk relative to the non-directive cluster of categories; these categories also have a more cognitive association. Pupil-centered refers to category 19 (asks questions).

Further conclusions are based on the direct oblimin rotated components, as the principal components 3-5 do not differ much in percentage variance explained. The correlations between the rotated components are presented in Table 4. Because the first two rotated components are moderately strongly correlated, categories with the same (opposite) sign on both are in general positively (negatively) correlated. The third conclusion is based on the less negative correlation between the second retated component and the third, than between the second and the first rotated component; the high loadings of the categories 8, 14, and 17 on the second rotated component, and 11 and 12 on the third. Conclusion 4 is a summary of the conclusions 1-3, the loadings of categories on the fourth and fifth rotated component, and the correlations of the rotated components.

TABLE 4: Correlations Between Rotated Components for First Occasion

	1	2	3	4
1				
2	28			
3	-15	-10		
4	-21	07	-05	
5	12	05	-05	-11

Note: decimal point omitted.

RESULTS FOR SECOND OCCASION

General level of behavior. The means are presented in Table 1. Teacher behavior occurs about 1.5 times more than pupils' behavior. Cognitive pupil behavior still dominates.

Conclusions of structural analysis. General conclusions will be stated first.

1. Clases can be most clearly divided into two opposite types, those with non-directive teacher behavior and those with directive, discouraging, and lecturing teacher behavior. In the first type of classes, relatively, much cognitive and affective pupil behavior related to values development occurs. In the other type, there is, relatively, much unordered behavior.

2. Within the non-directive group, relatively, more non-directive and moderate discouraging behavior is related to more cognitive behavior with respect to values development, and affective behavior is primarily represented by referring to values of others. If there is less non-directive behavior and less discouragement, the accent in the mixture

of cognitive and affective pupil behavior is on expressions of personal values.

3. Being personal is positively related to informing, and negatively to behavior related to values development. It is also negatively related to non-directive behavior.

4. Other behaviors do not have very strong relations with the types mentioned. "Summarizing" is negatively related to asking open and closed questions, but it is not related to the valuing processes. "Accepting" is positively related to various kinds of asking questions, but is not related to discouraging behavior and only slightly negatively to the valuing processes. "Praising" is positively related to referring to values. Given favorable conditions, more referring to values is therefore related to more praising by teachers.

Details of analysis. The conclusions are based on the loadings presented in Table 3. The five components account for 61% of the variance. It should be noted that for interpretation some categories with low commonalities should not be given too much attention. The first component is the most important. The first conclusion is based on the strong loadings on this component of the categories 1, 4, 14-18 versus 5, 6, 11, 12, 21, 22. It should be noted that the positive encouraging categories 9 and 10 (accepts and praises) are not clearly related to either of these opposite types.

The other components hardly differ in percentage of variance explained. The other conclusions are, therefore, based on the varimax rotated components (direct oblimin did not converge). In contrast to the first occasion, the varimax rotation does yield a simple loading pattern. Conclusions 2 and 3 are based on the loading pattern on the first two rotated components. The first rotated component separates "discouraging" directive behavior, with unordered pupil behavior, from values development related behavior, both cognitive and affective. The second rotated component separates "informing" and "being personal" from non-directive behavior and behavior related to cognitive values development. Conclusion 4 is based on the loadings on the other rotated components.

ANALYSIS OF CHANGE IN STRUCTURE BY MEANS OF THREE-MODE PRINCIPAL COMPONENT ANALYSIS

Three-mode principal component analysis is a method to perform principal component analysis on the three modes of a data block simultaneously. The technique was introduced by Tucker (1963, 1964). A recent and up-to-date treatment is given by Kroonenberg (1983). For details the reader is referred to these authors and to Kroonenberg and De Leeuw (1980). Here only a short description will be given. Appendix B provides some technical details on the model, and the way it is used here.

In this example the modes consist of classes (mode 1), categories (mode 2), and occasions (mode 3). In the three-mode model classes and categories have loadings on separate class and category components, respectively. The relationships between the class and category components are represented in a so-called core matrix. The entries in the core matrix can be interpreted as the scores (loadings) of each component of one mode (classes) on each component of a second mode (categories). The larger a particular entry, the stronger the relationship. In this application the TUCKALS2 model is used, which amounts to the estimation of one core matrix for each occasion separately. Differences between corresponding entries in the two core

matrices are evidence of a change in structure. They indicate that the scores of class components on variable components change from the first to the second occasion. When these changes are substantial relative to the elements in the core matrix, the change in structure is substantial relative to the whole structure. However, changes of elements in the core matrix do not reveal which categories are involved. This will be investigated by computing, for each class component separately, the distances between the component and the categories at the two occasions, and the change in distances. This amounts to deriving a pattern of behavioral change for classes that would load exclusively on the component concerned. Although some classes have a substantial loading on only one component, no class has a non-zero loading on only one component. Because this pattern is inferred for a class component, its importance can be judged from the percentage variance accounted for. Hence, a change pattern is derived that is an idealization of actual patterns. The notion of an "ideal" class is justified in our opinion, as it is implicit in teacher training programs. Such a class may also be actually observed, in contrast with a latent variable derived from observed variables.

General level of change. The means of the first and second occasion, the descriptive significance level of a t-statistic for paired observations and omega squared (Hays, 1972, p. 417) for each category are presented in Table 1, and the correlations between category scores on the two occasions are presented at the diagonal of Table 2. Although there is an increase in giving sharp negative feedback, the differences between the first and second occasion can in general be interpreted as favorable (i.e., more values development related behavior occurs).

From the correlations on the diagonal in Table 2 it can be concluded that on the second occasion classes differ much in their relative scores for the categories 4, 8, 17, 20 and to a lesser extent 14, 18. Their scores do not differ much for the categories 1, 3, 7, 9, 10, and even more so for 5, 11, 12. After orthogonal rotation to congruence, congruence coefficients between the corresponding principal components of the first and second occasion were .90, .46, .29, .20, .49 for the components 1-5 respectively. The fifth component of the first occasion and the fourth component of the second occasion had a relatively substantial congruence coefficient of -.22. In general, the structure appears to have changed substantially. It should also be noted, that although a low congruence coefficient is evidence for a change in structure, perfect congruence does not imply the reverse. For example, permutation of the rows of a data matrix would result in perfect congruence coefficients for the loading pattern.

Conclusions of the analysis.

1. Classes with a relatively favorable, non-directive, encouraging climate for values development at the start of the experiment, have a relatively favorable climate at the end. The same holds for classes of the opposite type. Clearly (un)favorable climates do not change easily.

2. Classes with a cognitive, pupil-centered climate at the start change into a relatively favorable direction (i.e., values development related behavior is relatively more frequenct at the end of the experiment). Similar changes occur to a lesser degree in teacher-centered, structured classes.

3. The change mentioned in 2 is related to more "asking open questions," "redirecting," "summarizing," "accepting," "praising" and less "lecturing," "sharp and corrective negative feedback" and "unordered

pupil behavior."

Details of analysis. Before the conclusions are discussed, some general remarks will be made. The first five category and class components account for 23, 9, 7, 6, 5 and 20, 9, 8, 6, 5% of the variance, respectively. The model accounts for 48, 54, and 43% of the total variance, the variance of the first and the variance of the second occasion respectively. The common structure therefore resembles more the structure on the first occasion. Especially the first two common category components are highly similar to the first two components of the ordinary analysis on the first occasion.

TABLE 5: Distances Between Classes and Categories

	$(C_1H')'$ for first occasion					$(C_2H')'$ for second occasion				
Categories:	Class components					Class components				
	1	2	3	4	5	1	2	3	4	5
1. Asks open questions	-2.04	-0.87	2.45	-0.75	-1.75	-2.47	1.57	1.21	1.05	-3.00
2. Asks closed questions	-0.59	-2.36	-1.06	1.50	-1.29	-1.34	-1.54	0.63	1.69	-0.89
3. Probes	-3.48	-0.69	1.10	0.80	0.62	-2.44	-0.41	0.28	0.98	1.66
4. Redirects	-3.35	-1.01	1.62	-0.63	-0.21	-2.57	0.85	-0.13	0.27	-0.25
5. Gives direction	3.01	0.74	-0.58	-0.44	-0.67	2.01	0.27	0.32	-0.41	-1.57
6. Lectures	1.59	-0.87	-0.68	0.97	0.03	1.60	-2.68	1.54	1.26	1.32
7. Being personal	1.19	-2.36	-2.37	0.62	-0.35	1.04	-2.55	-0.29	0.38	0.71
8. Summarizes	0.08	-2.06	-1.22	-1.72	-0.10	0.71	-0.61	-1.57	-1.35	0.14
9. Accepts	-2.38	-3.04	0.76	0.23	-1.07	-1.96	-1.58	1.16	1.73	-0.24
10. Praises	-1.67	-1.95	0.66	-2.82	-0.38	-0.63	0.82	-1.41	-1.49	-0.76
11. Corrective feedback	3.18	1.14	1.30	-1.57	-0.25	3.07	-0.19	1.74	-0.24	-0.68
12. Sharp negative feedback	3.47	0.87	1.57	-2.18	-0.36	3.53	-0.36	1.93	-0.37	-0.78
13. Irrelevant behavior	0.00	3.34	1.62	1.56	1.40	0.16	0.48	1.54	0.85	1.77
14. Expresses personal values	-3.34	-1.09	0.46	-0.98	0.06	-2.56	1.41	-1.74	-0.85	-0.23
15. Personal behavior	-3.67	0.43	1.33	-0.85	0.81	-2.59	2.08	-1.52	-0.99	0.53
16. Gives opinion	-1.23	1.31	0.80	0.89	-0.49	-2.09	2.11	-0.06	0.40	-1.69
17. Refers to values of others	-2.54	-1.74	0.53	-2.19	-0.65	-1.95	1.70	-1.85	-1.34	-1.52
18. Gives information	-3.43	1.66	1.41	0.35	1.42	-2.44	1.84	-0.93	-0.43	1.51
19. Asks questions	-0.58	1.11	2.38	-1.25	0.89	0.63	0.21	1.13	-0.14	1.29
20. Structured working	1.33	1.25	-2.53	2.55	-0.47	-0.66	0.97	-1.16	0.17	-1.47
21. Unordered behavior	3.59	2.13	1.00	-0.12	-0.51	2.57	0.23	2.12	0.46	-1.34
22. Silence	0.66	0.17	-0.92	-0.72	1.47	1.76	-0.98	-1.02	-1.26	2.52

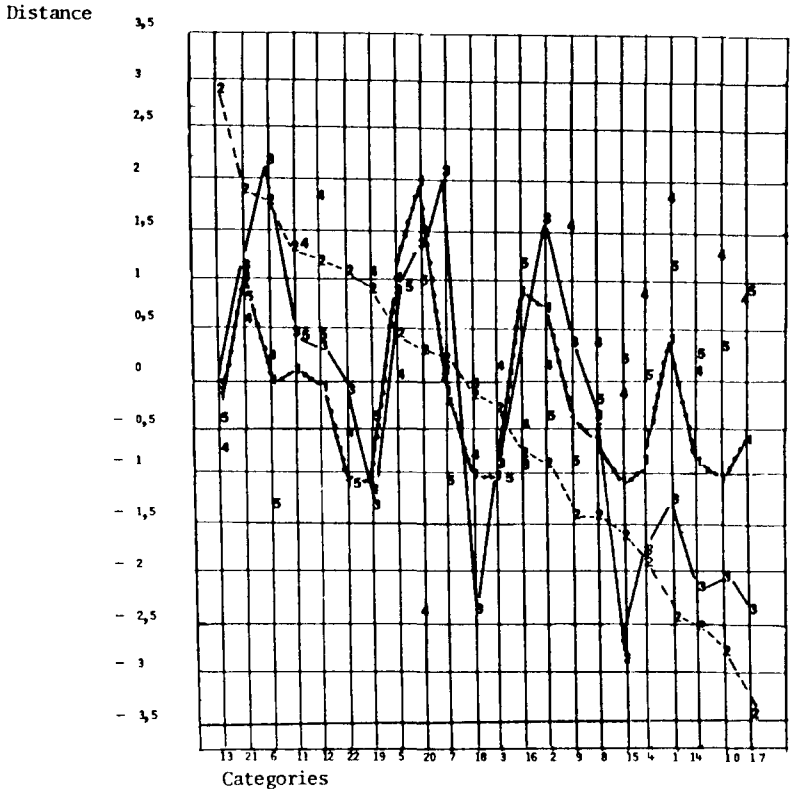
Class components can be characterized according to their distances to categories on the first occasion, if these distances are large relative to other class components. Classes with a positive (negative) loading on a class component are close to positive (negative) values (and far from negative (positive) values) in the columns of Table 5, which contains the distances. From these, the first, second, and third class components can be characterized on the first occasion as, respectively: non-directive, cognitive as well as affective classes, versus directive, disordered classes; teacher-centered, cognitive (categories 2, 7, 8, 9, 10) versus pupils-centered, cognitive (categories 13, 16, 21). Although a similar distinction was made in the analysis of the first occasion, in the first group some more encouraging occurs here, and in the second some more disordered behavior; non-directive cognitive (categories 1, 19), versus teacher-centered, structured (categories 7, 20).

TABLE 6: Core Matrices for First and Second Occasion, and Their Difference

Class Components	1	2	3	4	5
	Category Components First occasion (C_1)				
1	11.30	-0.07	1.08	0.53	-1.43
2	2.83	-6.39	-2.94	-0.31	-1.98
3	- 2.22	-4.96	1.01	1.42	3.59
4	0.55	1.38	-5.24	2.55	-2.41
5	- 0.13	-1.44	-2.37	-2.83	0.52
	Second occasion (C_2)				
1	9.29	-0.44	1.16	-1.88	1.15
2	- 3.52	-4.52	1.53	-0.09	-2.70
3	3.43	-1.54	-0.92	3.64	2.99
4	0.36	0.95	-2.17	3.69	1.08
5	0.11	0.40	-4.39	-3.97	2.59
	Differences C_1-C_2				
1	2.01	0.37	-0.08	2.41	-2.58
2	6.35	-1.87	-4.47	-0.22	0.72
3	- 5.65	-3.42	1.93	-2.22	0.60
4	0.19	0.43	-3.07	-1.14	-3.49
5	- 0.24	-1.84	2.02	1.14	-2.07

Note: important difference underlined.

The core matrices of the first and second occasion, and their differences, are presented in Table 6. On both occasions the first class component loads heavily on the first variable component exclusively. Conclusion 1 is based on this observation. It is evident from Table 6 that the changes for the second and third class components are the most important. In Figure 1, the difference between corresponding columns of Table 5 are plotted. If a difference is positive, classes with a positive (negative) loading on the class component corresponding to a column in Table 5 have lower (higher) standardized scores for the category concerned on the second than on the first occasion. To facilitate the comparison with the differences of the first and second columns, the differences for the third and fourth column were multiplied by -1. In Figure 1, the categories are ordered according to the greatest changes. Conclusions 2 and 3 are based on Figure 1.



Note: Rows 3 and 4 of ((C_1-C_2) H')' multiplied by -1.
The numbers in the figure refer to class components 1-5.

FIGURE 1: Plot of Changes of Distances Between Classes and Categories

DISCUSSION

A cluster of interrelated teacher behaviors, consisting of asking open questions, redirecting, summarizing, accepting, and praising, has been identified as being positively related to the valuing processes. Another cluster, consisting of giving directions, negative feedback, and unordered pupil behavior, was negatively related to the valuing processes. The two clusters form a dimension that may be labeled non-directive versus directive. However, other teacher behaviors do not coincide with this dimension, but they still account for sizable portions of behavioral variance. Of particular interest among these behaviors are lecturing and being personal. All three analyses suggest that lecturing is negatively related to values development, and especially the analysis of change in structure suggests the same for being personal.

The results can be interpreted as generally conforming to the recommendations of Raths et al. (1966) and Simon et al. (1972) about how teachers should behave. Summarizing, which requires active listening, seems particularly important. However, contrary to the recommendations of Raths et al. it is not recommended that "being personal" (i.e., the teacher's expression of own values and feelings, opinions, and situation) is used as initiating teacher behavior. It may be more appropriate as a response to

pupil questioning. Teachers should at the same time structure pupil activities in an unobtrusive way. This is probably the most difficult task for the teacher.

Because the data are correlational, causal interpretations are not really justified. The analysis of change suggests, however, that the non-directive pattern may as a whole be causally interpreted, although confounding factors cannot be entirely ruled out. This hypothesis should be tested in future research, to establish more firmly the relations between these behaviors and values development processes, and to disentangle cause and effect within this pattern. Future research should also clarify why classes initially differ with respect to non-directiveness and values development, and why some classes change more than others toward this pattern. These differences may be partly explainable by teacher and pupils' behavior during regular lessons. For instance, the first class component hardly contributed to the structural change which, for the more directive pattern, implies stability of relatively large amounts of negative feedback and unordered pupil behavior. Rosenshine (1979) has suggested that this pattern probably indicates difficulty in controlling students, which would be unfavorable to any program. Observations of classroom dialogues during regular lessons would reveal whether or not this pattern is restricted to values development lessons. More generally, such observations would show to what extent teachers allow pupil initiative during regular lessons. Evidently, teachers should not be uncomfortable with pupil initiative.

Observations during regular lessons is also important for another reason. The behavioral pattern most effective for values education is obviously not compatible with direct instruction, which has been shown to be effective for teaching basic skills, at least in elementary grades (Brophy, 1979; Gage, 1984; Rosenshine, 1979). However, Peterson (1979) claimed, after a review of research findings, that care should be exercised in generalizing about the effectiveness of direct instruction as opposed to open teaching, which is surely related to non-directiveness. Open teaching is apparently more favorable to creativity and problem solving. (Parenthetically, this would explain the ineffectiveness of "lecturing" and "being personal," as the execution of the valuing processes cannot consist of applying some learned rule or algorithm.) There is also some evidence that a highly structured and teacher-controlled classroom climate is negatively related to continuing motivation to learn (Pascarella, Walberg, Junker, & Haertel, 1981). In our opinion, there are good reasons for considering values education an important educational objective too (cf. Fenton, 1976), aside from achievement. Hence, it seems important to investigate whether teachers can be trained to be effective for both teaching basic skills and for stimulating values development.

APPENDIX A

THE VALUES DEVELOPMENT INTERACTION CATEGORY SYSTEM

Teacher talk categories

1. Asks open questions: asks relatively open-ended questions which call for unpredictable, person-specific response, about a new topic.

2. Asks closed questions: asks questions requiring predictable response, short reply or yes-or-no answer, about a new topic. Response can often be judged as right or wring.

3. Probes: asks new, often more specific question about the same topic, mostly of the same pupil.

4. Redirects: asks same question about same topic, of another pupil.

5. Gives direction: all directions, commands, procedures or orders which a pupil is expected to comply with.

6. Lectures: factual information to clarify a topic, procedure.

7. Being personal: personal values and feelings, situations, opinions (cf. categories 14, 15, 16).

8. Summarizes: tries to state the crux of the whole matter based on one or more pupil statements. If the teacher brings his own opinions or judgments into play, then shift to the feedback categories 10, 11, 12 or category 7.

9. Accepts: accepting response (hm, hm; yes) repeating pupil statements.

10. Praises or encourages: positive, enthusiastic response. Direct motivating and supporting response.

11. Gives corrective feedback: correcting response about behavior of a pupil with the intention to change pupils' behavior from non-acceptable to acceptable by reminding the pupil of accepted rules or creating a new rule or agreement.

12. Gives sharp negative feedback: correcting response about pupil or pupils behavior with the intention to change pupils behavior from non-acceptable to acceptable. Note the intonation.

13. Irrelevant behavior: other teacher talk/all teacher talk that does not fit in one of the aforementioned categories.

Pupil talk categories

14. Expresses personal values and feelings: utterance in which a pupil explicitly expresses a feeling, an appraisal.

15. Personal behavior description and situation description: talking about own behaviors, situations the pupil was confronted with, without referring to feelings, appraisals, opinions.

16. Gives opinion: own opinion, judgment without making feelings verbally explicit.

17. Refers to values and feelings, behaviors, situation descriptions, opinions of others: repeats, summarizes, reflects values and feelings of others.

18. Gives information: see category 6.

19. Asks questions: any question posed to teacher or other pupils.

20. Structured working, other pupil talk: all task-oriented behavior that does not fit in one of the aforementioned categories.

21. Unordered, interruptive behavior: considerable noise of some or all pupils, which disrupts planned activities.

22. Silence during classroom dialogues.

Note: In the test, the categories are referred to by the underlined parts of the descriptions.

APPENDIX B

Let z_{ijk} be the score of class i (i = 1, .., n) on variable j (j = 1, .., m) at occasion k (k = 1, 2), then the three-mode model as used in this paper can be written as:

$$z_{ijk} = \sum_{p=1}^{n} \sum_{q=1}^{m} g_{ip} h_{jq} c_{pqk} \quad \text{or} \quad Z_k = GC_kH'$$

where g_{ip} and h_{jq} are the loadings of the ith class and jth category on the pth class and qth category component, respectively; G and H are matrices of order classes by class components and categories by category components, containing loadings of classes and categories on class and category components. The pxq core matrix C_k contains the elements c_{pqk}: the loading of the pth class component on the qth category component on occasion k. Of course, Z_k (of order classes by categories) is often only approximated as GC_kH' by a small number of class and category components. These approximated scores are called distances in the paper, since they are reproduced from the inner-products of class and category vectors in the spaces of class and category components (cf. Kroonenberg, 1983, p.164). Since $Z_1 - Z_2 = G (C_1 - C_2) H'$, a necessary and sufficient condition for a structural change is that $C_1 - C_2 \neq 0$. The matrix $(C_1 - C_2) H'$ (or alternatively $C_1H' - C_2H'$) contains for each class component the information to which categories the changes apply. The jth element (corresponding to the jth category) in the pth row of this matrix is proportional to the change that would be observed for a class which loads exclusively on the pth class component. These elements are plotted in Figure 1.

REFERENCES

Beem A.L., & Brugman. D. (LICOR, Report). The effects of values development lessons on classroom behavior and pupils outcomes as measured by questionnaires, and the relations between classroom behavior and those outcomes. Internal Report. Leiden: LICOR.

Brophy, J.E. (1979). Teacher behavior and its effects. Journal of Educational Psychology, 71, 733-750.

Brugman, D., & Beem, A.L. (1982). Een experimenteel onderzoek naar de effecten van oefeneenhaden voor leerkrachten bij lessen waardenontwikkeling. In M. Boekaerts & C.E.M. van Lieshout (Eds.) Sociale en motivationele aspekten van het leren. Lisse: Swets & Zeitlinger.

Casteel, J.D., & Stahl, R.J. (1975). Value clarification in the classroom: A primer. Santa Monica, CA: Goodyear.

Dixon, W.J. (Ed.). (1981). BMDP statistical software 1981. Berkely, CA: University of California Press.

Elffers, H., & Tavecchio, L.W.C. (1979). Variance components in test generalizability research: which, when, why? Vereniging voor Onderwijs-research: V.O.R.-publikatie 9.

Fenton, E. (1976). Moral education: the research findings. Social Education, 40, 61-86.

Flanders, N.A. (1970). Analyzing teaching behavior. Reading, MA: Addison-Wesley, 1970.

Gage, N.L. (1986, October). What do we know about teaching effectiveness? Phi Delta Kappan, 87-93.

Hays, W.L. (1973). Statistics for the social sciences (2nd ed.). London: Holt, Rinehart and Winston.

Hawkins, D.M. (1980). Identification of outliers. London: Chapman and Hall.

Hermans, H.J.M. (1976). Value areas and their development. Amsterdam: Swets & Zeitlinger.

Kirschenbaum, H. (1973). Beyond values clarification. In H. Kirschenbaum & S. Simon (Eds.), Readings in values clarification. Minneapolis: Winston Press.

Kirschenbaum, H. (1977). Advanced value clarification. La Jolla, CA: University Associates.

Kohlberg, L. (1973). The child as a moral philosopher. In H. Kirschenbaum & S. Simon (Eds.), Readings in values clarification. Minneapolis: Winston Press.

Kohlberg, L. (1975). The cognitive-developmental approach to moral education. Phi Delta Kappan, 56, 670-677.

Kroonenberg, P.M. (1983). Three-mode principal component analysis: theory and applications. Leiden: DSWO Press.

Kroonenberg, P.M., & De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. Psychometrika, 45, 69-89.

Lockwood, A.L. (1978). The effects of values clarification and moral development curricula on school-age subjects: a critical review of recent research. Review of Educational Research, 48, 325-364.

Pascarella, E.T., Walberg, H.J., Junker, L.K., & Haertel, G.D. (1981). Continuing motivation in science for early and late adolescents. American Educational Research Journal, 18, 439-452.

Peterson, P.L. Direct instruction reconsidered. In P.L. Peterson & H.J. Walberg (Eds.), Research on teaching. Berkely, CA: McCutchan Publishing Corporation.

Raths, L., Harmin, M., & Simon, S. (1966). Values and teaching. Columbus, OH: Charles E. Merrill.

Rokeach, M. (1979). Value education in educational settings. In M. Rokeach (Ed.), Understanding human values. NY: The Free Press.

Rosenshine, B. (1979). Content, time and direct instruction. In P.L. Peterson & H.J. Walberg (Eds.), Research on teaching. Berkely, CA: McCutchan Publishing Corporation.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R.M.W. Travers (Ed.), Second handbook of research on teaching. Chicago: Rand McNally.

Simon, S., Howe, L., & Kirschenbaum, H. (1972). Values clarification: A handbook of practical strategies. New York: Hart.

Tucker, L.R. (1964). The extension of factor analysis to three-dimensional matrices. In H. Gulliksen & N. Frederiksen (Eds.), Contributions to mathematical psychology. NY: Holt, Rinehard and Winston.

Van der Plas, P. (1981). Waardenontwikkeling in het onderwijs. 's-Gravenhage: Stichting voor Onderzoek van het Onderwijs S.V.O.

Winer, B.J. (1971). Statistical principles in experimental design. (2nd ed.). Tokyo: McGraw-Hill, Kogakusha.

THE AUTHORS

A. LEO BEEM is a Senior Researcher at the Center for Educational Research of the University of Leiden. His field of interest is research methodology, motivation and learning.

DANIEL BRUGMAN is a Senior Researcher at the Center for Educational Research of the University of Leiden. His field of interest is classroom interaction and motivation.