THE DEVELOPMENT, VALIDATION AND APPLICATION TO TEACHER TRAINING OF A SYSTEM (PEIAC/LH-75) DESIGNED TO EXPAND THE FLANDERS SYSTEM OF INTERACTION ANALYSIS FOR DESCRIBING TEACHER-PUPIL INTERACTION PROCESS IN PHYSICAL EDUCATION CLASSES

By

Liisa M. Heinilä

University of Jyväskylä, Department of Physical Education

THE DEVELOPMENT, VALIDATION AND APPLICATION TO TEACHER TRAINING OF A SYSTEM (PEIAC /LH-75) DESIGNED TO EXPAND THE FLANDERS SYSTEM OF INTERACTION ANALYSIS FOR DESCRIBING TEACHER - PUPIL INTERACTION PROCESS IN PHYSICAL EDUCATION CLASSES (Supplement)

LIISA HEINILÄ

UNIVERSITY OF JYVÄSKYLÄ

SF-40100 Jyväskylä, Finland

The development of new programmes of practice teaching presupposes the controlling and evaluation of their basic elements. The purpose of this investigation was to develop a system for teacher education and research. The study was based on Flanders theory (1965, 1970) and experimentation (Heinilä 1970, 1971, 1974, 1976, 1979, 1980, 1983).

The measuring device (PEIAC/LH-75) contains three categorical clusters and triple coding was made at six-second intervals. The three clusters observed were 1) teacher's and pupil's speech and movement (12 categories), 2) social access and collective movement activity/passivity (8 categories) and 3) the social form of the class (7 categories), altogether 27 categories.

The construct validity and the objectivity of coding was studied and the applicability of the measuring device was evaluated in live and videotaped observations of in different aspects of P.E.

The scores of six trained observers, each having observed 24 lessons three times in different situations representing different sex, three grade levels and four subjects were analysed. Profiles, matrices of each cluster, indices, r-correlations, factors structures (7), homogeneous structure groups (6) were produced and compared; the most sensitive clusters, categories and discriminating functions identified; the level of objectivity of coding (Scotts' Pi and Kendall's-W) and its relations with variables was determined.

In curriculum evaluation the congruence between intended and actually occuring outcomes was studied. The data cover the courses of microteaching, whose purpose was to develop the verbal indirect teaching behavior of student teachers, arranged by faculty in 1974 and in 1976. The subjects were male and female students (N=48 and N=74), in 275 lessons. The observation instrument (PEIAC/LH-75) was used in a somewhat modified form: two clusters, speech and movement, and altogether 18 categories. The reliability was Scott' Pi .78. The construct validity was analysed by using a multivariate approach. The category frequencies, indices and student evaluations of courses (Questionnaire) were compared using analysis of variance and t-test and chi-square test. The revised 1976 curriculum proved more effective.

The instrument of interaction analysis and its modification used in the courses proved feasible both for research and for teacher training. It facilitated the operationalization of intended behavior, code patterns, and helped to teach discrimination and to create desirable teaching patterns.

Key words: teaching behavior, interaction process analysis, observation instrument, evaluation, reliability, objectivity of coding, construct validity, sensitivity, multiple discriminate analysis, evaluation of curricula , microteaching, congruence between intended and actually occuring outcomes.

Liisa Heinilä
University of Jyväskylä,
Department of Physical Education

THE DEVELOPMENT, VALIDITATION AND APPLICATION TO TEACHER TRAINING OF A SYSTEM (PEIAC/LH-75) DESIGNED TO EXPAND THE FLANDERS SYSTEM OF INTERACTION ANALYSIS FOR DESCRIBING TEACHER-PUPIL INTERACTION PROCESS IN PHYSICAL EDUCATION CLASSES

SUMMARY

The main purpose of this study was to (1) Develop and test a system for describing interaction process in p.e. classes, (2) analyse critically the reliability and validity of the constructed system and (3) apply the system to teacher education in microteaching.

The specific character of physical education requires adaptation of the FIAC system of interaction analysis and taking into account how movement communicates and influences. Consequently, three clusters were included in the PEIAC/LH-75 system: (1) teacher and student talk and teacher's silent activity; (2) students' collective movement activity/ passivity and social access; and (3) the social form of the class. The clusters contain 12, 8, and 7 categories respectively, altogether 27 categories: The time-sample units were taken at six-second intervals. The scores of six trained observers, each having observed 24 lessons three times in different situations representing both male and female teachers, three grade levels and four subject areas were analysed; profiles, matrices of each cluster, indices, r-correlations, factornstructures (7), homogeneous structure groups (6) were produced and were compared; the most sensitive clusters, categories and discriminant functions were identified; the level of objectivity of coding (Scotts'Pi and Kendall's-W) and its relations with variables was determined.

The average level of objectivity of coding (Scotts'Pi) varied according to cluster: I, .61; II, .65; and III, .69. The intercoder agreement was .65, within-coder constancy .69, and between-coder constancy .60 when two observations of the videotape recordings (T2 and T3) were compared. The results indicated that the intercoder agreement was

some what higher within the videotaped material than in the live situation.

The measuring instrument was reliable when estimated by using a nonparametric coefficient of concordance: 23 of 27 categories yielded a value of W significant at the 0.01 level (chi-square test).

In the third phase, the variation of coders was examined. Five discriminant functions were identified, three of which were statistically significant. The first accounted for 58 % of the model's total variance. The inverse character of reliability and validity was highlighted.

The validity of PEIAC/LH-75 was assessed mainly in terms of the question of construct validity, which was demonstrated by convergent and discriminant validitation methods. In this context, a model was developed to define the overall research strategy for the project.

In the primary analyses, it was noted, among other items, that all of the PEIAC/LH-75 categories were used in coding. The results of secondary analysis showed that the discriminant functions clearly distinguished between lesson groups. The first discriminant function proved much more powerful than the other four, with 47 % of the total discrimination of the model. The analysis selected 16 out of the total of 27 categories and set them in sequence according to how much they increased the model's discrimination power. The categories of the second cluster (students' collective activity/passivity and social access) and the categories of the third cluster (social form) proved to possess the highest discrimination power.

It was established that the various subject areas in P.E. and/or the teacher teaching in them are the strongest influence on the shaping of the interaction process as well as on theobjectivity of measurement. Of all categories situation with closed and open ideas' discriminated most clearly.

It was concluded that (1) the instrument possesses a definite degree of construct validity, and that (2) it is sufficiently sensitive to discriminate aspects of direct-nondirect teaching behavior.

In the Application of PEIAC/LH-75 to teacher education the purpose was to evaluate and compare two curricula, whose aim was to develop the verbal indirect teaching behaviour of student teachers. The congruence between intended and actually occurring outcomes was studied. The curricula of courses differed in terms of the following elements: (I) information about (models of) target behaviour (written, audiovisual), (II) timing of instruction of theoretical considerations (before/during the course), (III) size of training groups (5-10), (IV) length of microlessons (5-10 min.), and (V) number of microlessons (2-3).

The data cover the courses of microteaching arranged by the faculty in 1974 and 1976 and the subjects were male and female students who started their studies in 1971 (N = 48) and in 1974 (N = 74), 275 microlessons.

The measurement instrument (PEIAC/LH-75), was used in a somewhat modified form, containing two clusters, speech and movement, with 18 categories. This made it possible to give information about target behaviour, to operationalize model behaviour and to analyze TV-feedback using a systematic observation method. Reliability (.78) was estimated by means of Scott's pi-coefficient. The category frequencies, indices and student evaluations of courses were compared using analyses of variance and t-test (ANOVA), Mann Whitney U-test and chi-square test.

The success of the revised program was reflected in (a) a decrease of teacher talk, (b) an increase of teachers' silent didactic activities, (c) an increase in teachers' response behavior, and (d) a decrease in content emphasis. The increase of indirect behavior was evident in the second session, in which the teachers offered the pupils more opportunities to create ideas and solve problems, observed pupil responses and took advantage of their responses in the progress of the topic treatment.

Thus, the instrument of interaction analysis (PEIAC/LH-75 and its modification) used in the courses proved feasible both for research and for teacher training. It facilated the operationalization, information, evaluation and measurement of intended behavior code patterns, helped to teach discrimination, and to create desirable teaching patterns.

Key words: teaching behavior, interaction process analysis, observation instrument, evaluation, reliability, objectivity of coding, construct validity, sensitivity, multiple discriminant analysis, evaluation of curricula, microteaching, congruence between intended and actually occuring outcomes.

PREFACE

The project reported here consisted of phases, which have been reported separately within a long time period (1970-1983).

The purpose of this report is to do following:

1) create a synthesis of the part reports related to the study project
2) report the basis of decisions made by constructing the observation instrument PEIAC-LH/75
3) present the created measuring instrument
4) report the explorative studies made for determining the assumed capacity of the proposed instrument for gathering and organizing data in physical education teaching event based on a framework developed after surveying relevant research literature, and
5) discuss the results both from development and application perspectives

This report is based on the following original articles and technical reports, which will be referred to in the text:

Heinilä, L. (1970). Opettajan ja oppilaiden välisistä vuoro-vaikutussuhteista liikunnan opetustilanteissa (Report No. 22, pp. 80-94). Helsinki: Liikuntatieteellisen seuran julkaisuja. (Finnish Society for Research in Sport and Physical Education.)

Heinilä, L. (1971). Liikunnan opetustapahtuma sosiaalisena vuorovaikutusprosessina (Teaching of physical education as a process of social interaction). Unpublished master's thesis, University of Jyväskylä, Finland.

Heinilä, L. (1974). Developing a system for describing teacher-pupil interaction in physical education classes. FIEP Bulletin, 44(3), 16-20. (Also published in Education physigie des enfants avant l'epoque de la puberte (1976) (pp. 218-223). Warsaw: Edition Scientifiques de Pologne.)

Heinilä, L. (1976, July). Objectivity of coding in a system (PEIAC/LH-75) developed for describing teacher-pupil interaction in physical education. Paper presented at the International FIEP Congress of Physical Education, Jyväskylä, Finland.

Heinilä, L. (1977). Analysing systems in the evaluation of the teacher-pupil interaction process in physical education classes. FIEP Bulletin, 47(1), 20-34. (Also published in T. Tammivuori (Ed.), Evaluation: International Congress of Physical Education (Report No. 64, pp. 37-58). Helsinki: Finnish Society for Research in Sport and Physical Education.)

Heinilä, L. (1979) Application of interaction analysis to the teacher education in physical education (Research Bulletin No. 15). Jyväskylä, Finland: University of Jyväskylä, Department of Physical Education.

Heinilä, L. (1980). Developing a system (PEIAC/LH-75) for describing teacher-pupil interaction in physical education classes: Objectivity and content validity of coding. In G. Schilling & W. Bauer (Eds.), Audiovisual Means in Sport (pp. 361-370). Basel: Birkhaus Verlag.

Heinilä, L. (1983). Developing a system (PEIAC/LH-75) for describing teacher-pupil interaction in physical education classes: Construct validity and sensitivity. In R. Telama, V. Varstala, J. Tiainen, L. Laakso & T. Haajanen (Eds.), Research in school physical education (Report No. 38, pp. 124-132). Jyväskylä, Finland: University of Jyväskylä, Formation for Promotion of Culture and Health.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

A central task of the university is the planning, realization and evaluation of goal-directed educational programs. This activity should be long-term, comprehensive and integrated with general social planning. It should also be closely linked with decision making concerning all education. The ultimate aim of educational planning should be the quantitative and qualitative development of education (Itälä, 1969). The development of educational programs is a multistage process at several levels and should be based on scientific research.

Attempts have been made since early in this century to apply the methods of scientific research to the problems of school learning, teacher behavior, and teacher education. Within the behavioral sciences there has emerged a sub-discipline of "research on teaching" which Gage (1972) has defined in the following way:

> "Research" is defined as scientific activity aimed at increasing our power to understand, predict, and control events of a given kind. All three of these goals involve relationships between variables. ... "Teaching" in turn may be defined as events, such as teacher behavior, intended to affect the learning of a student. ... Given these definitions of "research" and "teaching," we can define "research on teaching" as the study of relationships between variables, at least one of which refers to a characteristic or behavior of a teacher. If the relationship is one between teacher behaviors or characteristics, on the one hand, and effects on students, on the other, then we have "research on teacher effects," in which the teacher behavior is an independent variable. If the teacher behavior or characteristic serves as a dependent variable in relation to some variable in the program of selecting and training teachers (the teacher education program), then we have "research on teacher education." Both kinds of research taken together make up the field of research on teaching. (pp. 16-17)

This definition does not suggest that other kinds of variables are not also useful, and in fact desirable, in research on teaching. It only specifies that the variables of teacher behavior and characteristics are at the center of concern and must be involved. Figure 1 illustrates the relationships in Gage's definition.

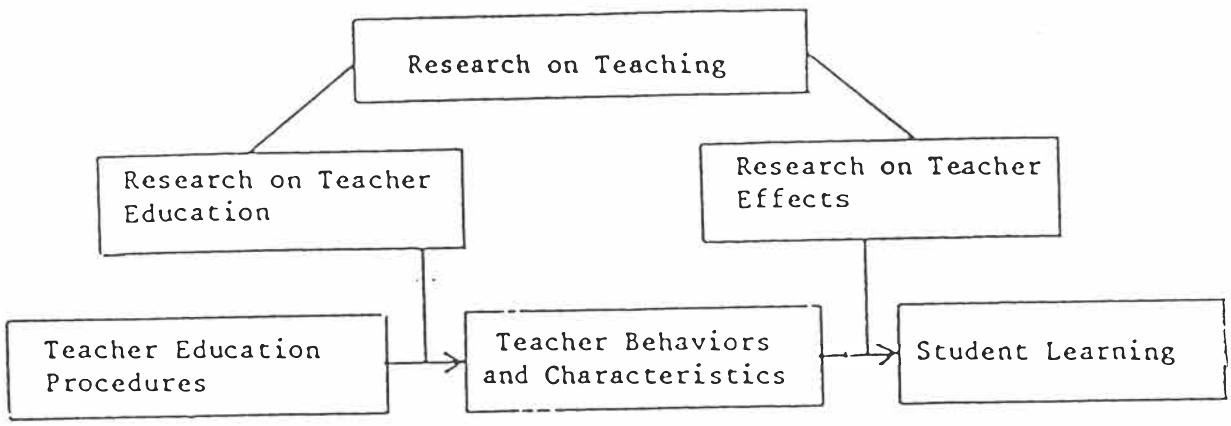


Figure 1. The field of research on teaching. (Gage, 1972, p. 17)

It has been suggested (Binet, 1918) that everything has been said in education while nothing has been proved. It is true that much has been done since the early decades of the century, but it is similarly true that several problems need to be addressed before practice teaching, and indeed, teacher training in general, can be fully developed. Only two of these problems will be taken up here. First, we need to have a feasible and comprehensive conceptualization of the nature of teaching. Second, we need reliable, valid and practicable ways of describing, analyzing and evaluating teaching activities and behaviors. Finally, having addressed these problems, we need to apply what we have found to teacher training programs.

Interaction Analysis Methods

The recent emphasis on interaction and communication between teacher and students and among students, and the subsequent development of methods of interaction analysis have had a profound impact on empirical research on teaching. At an early stage of this new research paradigm, there was a clear interest in studying what contributions interaction analysis might be able to make to teacher education and practice teaching.

Interaction analysis is a label that refers to any technique for studying the chain of classroom events in such a fashion that each event is taken into consideration (Flanders, 1970). The method is based on a conception of teaching as an interpersonal influence whose purpose is to effect pupil learning in line with set objectives. Typical of teacher behavior is human voice and motion, but it may also be frozen in the form of a book or film or a set of programmed instructional materials (Gage, 1972). In the study of teacher behavior, this influence can be observed on the basis of variable values placed on given dimensions such as teacher-centered/pupil-centered, direct/indirect, etc., and event sequences can be described, for instance, by means of a timeline display (cf. Flanders, 1970).

Methods of interaction analysis are based on theoretical considerations and thus contain given conceptual systems. This is true of the systems developed by Bales (1950) and Flanders (1965, 1970). Thus, in using methods of this kind the researcher has not only made methodological decisions but he has also bound himself to a particular theory and set of variables (Heinilä, 1974, 1977). In this way the measuring instrument achieves a central significance. It is therefore not surprising that interaction analysis methods have also proved to be an effective tool in teacher training. They provide a conceptual scheme and simultaneously the means for the operationalization and measurement of variables. Perceptions and communications become more unified and precise, evaluation and comparison attain higher objectivity. The contents of teaching programs refers to the matter being dealt with, such as command words in practice teaching in P.E., or other forms of social interaction, different types of ball games, etc. Form of teaching refers here to the way in which interpersonal communication is organized (Koskenniemi & Hälinen, 1970). It may be group work, problem solving, or programmed teaching, and it may be either direct or indirect. In the past, in the practice teaching of physical education, attention has been directed mainly to the contents of programs, while the development of forms of teaching has occupied a secondary position.

Teacher Education Research

The pedagogical and didactic problems of teacher education are a special subarea of what is now frequently referred to as the "pedagogy of higher education". The Finnish national Commission on Teacher Education (Vuoden 1973 opettajankoulutustoimikunnan mietintö, 1975) has suggested that the most important sectors of research on the pedagogy of higher education concern (a) the problems of the overall aims of higher education, (b) the problems related to the development and investigation of instruction, and (c) the special problems of educational technology and teaching methods. Within this latter area of concern, teacher education, one of the key issues is practice teaching. Researchers and teacher educators are constantly faced with the problems of how the experience should be planned and developed so that the intended competences can be optimally attained.

In January 1974 the Department of Physical Education of the University of Jyväskylä introduced, on an experimental basis, a new type of practice teaching using microteaching. The new course that emerged formed part of the degree requirements and was intended to be given during the last term of the third year as an obligatory course (45 hrs).

The implementation of the course necessitated development work on the methods of interaction analysis. The construction of the interaction model and the related observation instrument, PEIAC/LH-75, (see Chapter 5) that were used with the microteaching exercises was done during the period 1971-1973 (Heinilä, 1974, 1977). The final instrument was the result of empirical pilot studies (Heinilä, 1970, 1971), based on the pioneering work of Flanders (1965, 1970) and drawing on the expertise of the Helsinki DPA project (e.g., Komulainen, 1968, 1970, 1971a, 1971b, 1973, 1974, 1978; Koskenniemi & Komulainen, 1969)

An adapted version of Gage's model of research on teaching illustrates the place the present research occupies in this field (Figure 2).

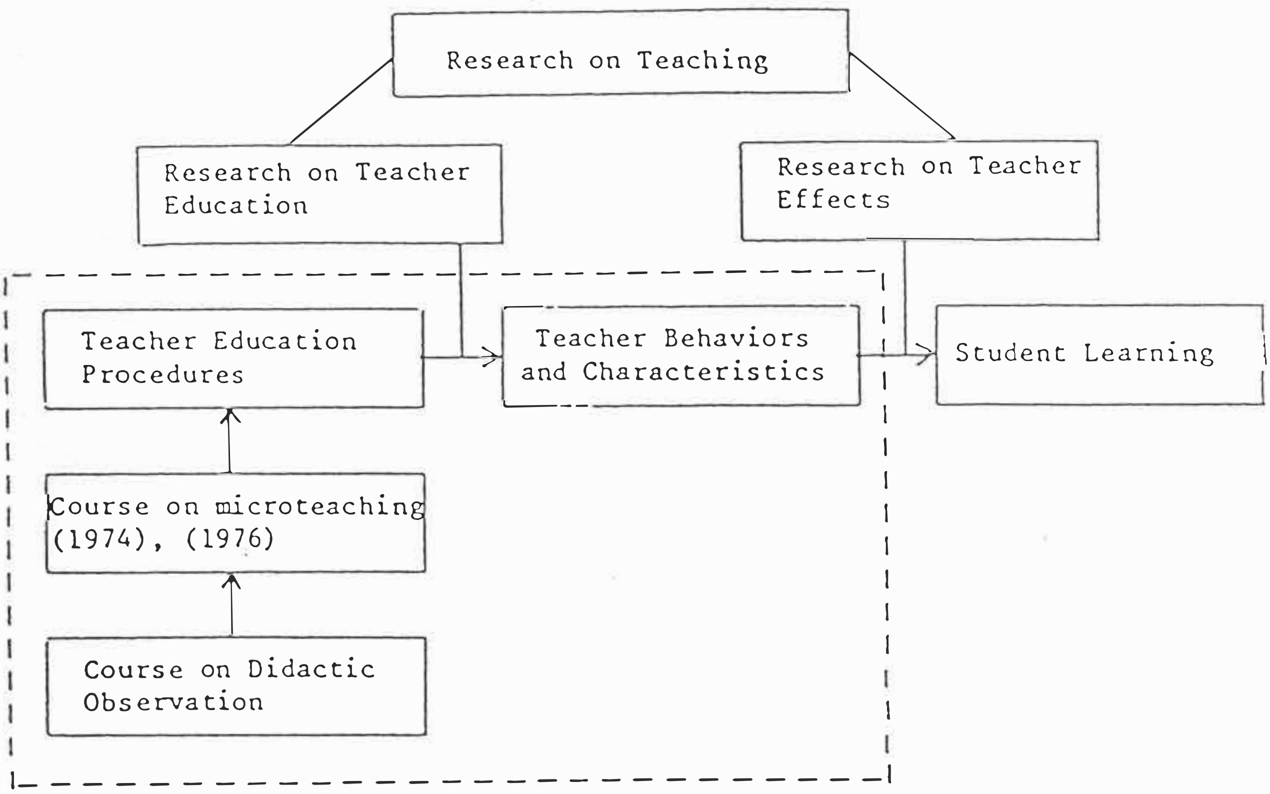


Figure 2. Adapted version of Gage's model of the field of research on teaching.

As can be seen from the model, the task of the project is to identify detailed, observable teacher behaviors that are related to student learning. The task of teacher education is to help student teachers get to know, understand and adopt effective teacher behaviors. So-called performance-based teacher education programs have been based on this outlook and the best known of such programs are microteaching and minicourses such as the one introduced in Jyväskylä. In such courses methods of interaction analysis have been used as a tool to help bring about changes of behavior. It is through the use of these methods, that this study will examine both the problem of describing the nature of teaching and the development of techniques to study these activities and behaviors. At the same time, it will be shown that the methods of interaction analysis provide a new basis for the selection of the forms and contents of teacher training so that the occupational demands of the teaching profession are fulfilled, and theory and practice can be brought closer together.

In summary, then, the purpose of this report is to give a complete account of a research program on the use of interaction analysis in physical education. Drawing on earlier reports, (a) the theoretical framework of the project and its relation to other work on interaction analysis will be described, (b) an account of the construction of the observation instrument will be given, (c) the empirical structure of the instrument will be explored, (d) the measurement properties (reliability, objectivity of coding, construct validity of coding, and construct validity and sensitivity) will be investigated, (e) the application of the instrument in a micro-teaching program and for curriculum evaluation will be described, and finally (f) implications of the study for further research will be discussed.

CHAPTER II
REVIEW OF LITERATURE

Overview

In order to set the present study in its proper context, this chapter will present a review of literature related to research on classroom observation. While Binet's dictum, quoted in the introduction, still is not much of an exaggeration as a summary of the state of education as science, it is true that some researchers in education became interested in analyzing classroom interaction as early as the late 1930's. Since that time, a number of category systems for analyzing primarily verbal interaction in the classroom have been constructed. A survey in the mid-sixties by Amidon and Simon (1965) reported twenty such category systems. Once developed, such category systems have been put to use in a great number of research studies. Early work involving systematic observation in classrooms has been reviewed in the first and second editions of the Handbook of Research on Teaching by Medley and Mitzel (1963) and by Rosenshine and Furst (1973). Medley (1982) has recently written a review of systematic classroom observation in the fifth edition of the Encyclopedia of Educational Research. Thus, any research study which focuses on classroom processes occurs within the context of a well-established research paradigm.

In this chapter we will first discuss the historical development of research on teaching, including the development of interaction analysis. Secondly, the most commonly used observation system in educational research, the Flanders Interaction Analysis Category System, will be described and discussed. This discussion will be followed by a review of research in physical education which has used interaction analysis and observation methods. Finally, these studies will be critically discussed in terms of their success in achieving valid and reliable results.

Historical Development of Research on Teaching

In their article on observation research, Evertson and Green (1986) identified four overlapping phases of the history of this approach to the study of educational processes. Phase One (ca. 1939-1963) was an exploratory phase which attempted to identify teacher-student interactions and other related classroom and instructional behaviors. Phase Two (ca. 1958-1973) was a period of instrument development, and of descriptive, experimental, and training studies. The use of category systems and issues about paradigms for the study of teaching emerged during this phase. During Phase Three (ca. 1973 to present) studies explored teacher behaviors that relate to student achievement, usually on standardized tests. Phase Four (ca. 1972 to present) runs concurrently with Phase Three and is a period of expansion, alternative approaches, theoretical and methodological advances, and convergence across research directions in the use of observational techniques.

This historical review of research on teaching will attempt to explore some of the work done during these phases of study with particular emphasis on the period of expansion, theoretical and methodological advances in the use of observational techniques. Phase four, to which the present study belongs.

Early Research on Teacher Effectiveness

Although research on teaching, as defined by Gage (1972), is relatively new, research on "teacher effectiveness" has been conducted for many years. The early studies were stimulated by the desire to provide an objective basis for the selection, training, employment, and promotion of teachers, but in reality they offered minimal opportunity for a real understanding of teacher effectiveness. In general, as Dunkin and Biddle (1974) emphasized, such studies revealed no more for teachers and educators than the discovery that performance on college examinations and in practice teaching are apparently unrelated to subsequent success in teaching. Many reasons have been offered by reviewers and critics for the failure of these early studies. Dunkin and Biddle have summarized these as (1) the failure to observe teaching activities, (2) theoretical impoverishment, (3) the use of inadequate criteria of effectiveness, and (4) the lack of concern for contextual effects.

With the development of the behavioral sciences in the first half of this century, attempts were made to apply these scientific methods to the problems of teacher behavior, school learning and teacher education. As Dunkin and Biddle point out, perhaps the most significant shortcoming of these early studies is that they consistently avoided looking at the actual process of teaching in the classroom. They further suggest that if teachers vary in their effectiveness, it must be because they vary in the behaviors they exhibit in the classroom. For this reason, the focus of a study on teacher effectiveness must be on the classroom where the teaching actually takes place.

Development of Analytical Research Methods

During the 1960's descriptive analytical research in general education increased considerably and became an independent branch of intellectual inquiry. Its general theoretical orientation became clearer and acquired a more definite direction. Research in this area has been directed towards (1) natural teaching situations; (2) the whole of the teacher-pupil interaction process; and (3) the construction of a uniform theoretical basis and conceptual scheme, within which the newly acquired empirical data can be placed, analyzed and generalized. (See, e.g., Birkin, 1971; Westbury & Bellack, 1971; Dunkin & Biddle, 1974; Heinilä, 1974, 1976.)

This orientation has been greatly influenced by the development of quantitative methods, and observation research has occupied a key position. In this context, observation research refers to the analytical methods based on observation, during which behavior is observed and classified. With this method, a classification system can be based on (1) theory, (2) a theoretical model, (3) existing observational systems, or (4) the results of empirical studies or pilot studies. When the focus of research shifted from teaching efficiency research towards the investigation of the classroom atmosphere and the regularities of the teaching-learning process, observation became the most practicable method.

Development of Observation Recording Instruments

In the field of observation research, the problems of content and method are closely related and they should therefore be examined simultaneously. The use of a measuring instrument implies a theoretical base. Such is the case with, for example, the classic interaction analysis systems by Flanders (1965, 1970) and Bales (1950). When a researcher adopts an instrument of this kind, he has not only made a methodological decision, but he has also committed himself to a particular theory and group of variables. In the study of teacher behavior, the theoretical base might be the observed variable values placed on given dimensions, such as teacher-centered/pupil-centered, direct/indirect, etc.; or the description of event sequences, for instance, by means of time-line display (cf. Flanders, 1970).

Analytical methods based on observation generally include (1) a group of carefully specified categories for the classification of the behavior under observation, (2) a group of standardized procedures which define the observation procedure, (3) instructions for processing, analyzing and presenting the data in a meaningful way which corresponds as closely as possible to the original events (Flanders, 1970; Heinilä, 1970, 1974, 1976). The category system employed will determine the number and quality of the events, which, defined in terms of interaction analysis systems, are exhaustive and mutually exclusive.

During the past two decades a great number of recording instruments have been developed for the study of teaching. (For reviews of some of these see, e.g., Medley and Mitzel, 1963; Simon and Boyer, 1970; Rosenshine, 1971; Rosenshine and Furst, 1973; Biddle, 1967; and Dunkin and Biddle, 1974.) Although these instruments have a common purpose to systematically record teacher-student behavior in the classroom, there are some major differences among them. These differences relate primarily to the dimension or dimensions of the classroom activity to be recorded. Generally, the focus of the instrument reflects the theoretical orientation of the investigator. The particular orientation of the investigator not only guides the general direction of the research work, but is also the key in making decisions concerning the logical steps in the development of the system.

Simon and Boyer (1970) report altogether 92 different recording systems, of which 79 were designed for observing classroom behavior. They suggest foci for categories within recording instruments and classify them as follows:

1. Affective - the emotional content of communication;
2. Cognitive - the intellectual content of communication;

3. Psychomotor - the non-verbal behaviors, posture, body position, facial expressions, and gestures;
4. Activity - what is being done that relates a person to someone or something else (for example, reading or hitting a ball);
5. Content - what is being talked about;
6. Sociological structure - the sociology of the interactive setting, including who is talking to whom and in what roles; and
7. Physical environment - descriptions of the physical space in which the observation is taking place, including the materials and equipment being used.

In a review of almost 500 studies involving the systematic observation of classroom teaching, Dunkin and Biddle (1974) identified six classifications according to content and/or the theoretical "orientation" toward teaching. These classifications are:

1. studies dealing with classroom climate;
2. studies dealing with management and control of pupil behavior in the classroom;
3. studies dealing with the classroom as a social system;
4. studies dealing with the knowledge and intellectual aspects of teaching;
5. studies dealing with logic and linguistics; and
6. studies dealing with the sequential patterns of classroom behavior.

Rosenshine (1971) classified the observation instruments used in fifty-one studies into "category systems" and "rating systems." In a category system, each behavior of the teacher or student was coded whenever it occurred. In a rating, or "sign," system, outside observers or students estimated the behavior of the teacher on a five- or seven-point scale. These observation systems were also classified according to the amount of inference required of the observer or the person reading the research report. The term <u>inference</u> refers, in this context, to the process intervening between the objective behavior seen or heard and the coding of this behavior on an observational instrument. Category systems are classified as "low-inference" measures because the items focus on specific, denotable, relatively objective behaviors, such as "teacher repeats student's idea" or "teacher asks evaluative questions," and also because the behaviors are recorded as frequency counts. The rating systems are referred to as "high-inference" measures because they lack the specificity of low-inference variables. In general, the category systems of observation have been used most frequently. They appear to be more flexible than sign observation and rating systems, provide more data, and have a higher level of objectivity in coding (Rosenshine, 1971; Dunkin & Biddle, 1974).

The task of the category format is to make it easier to organize the work of observers and to express the purpose of the research. On the basis of the degree of category specification and clustering, category formats can be divided into three types containing (1) a number of mutually exclusive categories, which are either unique or constitute a dimension; (2) a number of main categories, all or some of which are subdivided; or (3) a multiple coding system, which consists of a limited number of categories placed into separate clusters. These generally constitute a dimension based on some model of thought (Flanders, 1970; Heinilä, 1971, 1976).

To summarize, the preceding review has indicated that a large number of observational recording instruments have been developed to investigate classroom interaction. These can be divided into "category systems" or "rating systems." The former are regarded as "low-inference" systems because of their high degree of specificity, whereas the latter are regarded as "high-inference" systems, because they operate with more general concepts.

The work of researchers involved in classroom interaction analysis was primarily motivated by a desire to prove that certain preferred interaction patterns are superior for classroom learning. The concepts "integrative/dominative," "democratic/authoritarian," "student-centered/teacher-centered" and "indirect/direct," all spring from a conviction that most teachers could be more effective if they would interact with pupils rather than direct them.

Development of Interaction Analysis

In this section, an attempt will be made to outline the basic assumptions of the traditional interaction analysis paradigm. Given this frame of reference, it should be easier for the author to present a survey of related literature in a succinct form and for the reader to follow the exposition.

Kuhn (1962) introduced the term "paradigm" to denote the fact that some accepted examples of actual scientific practice, including law, theory, application and instrumentation, all together provide models which give rise to coherent traditions of scientific research. Sharing a paradigm means that there is a shared commitment to the same rules and standards for scientific practice. Kuhn suggests that scientists work from models acquired through education and through exposure to a common core of literature. This happens often without an explicit knowledge of why the models have obtained their status. It is even possible that there is no clear-cut underlying body of rules and assumptions for the research traditions.

Kuhn's point is relevant for the interaction analysis paradigm as well. A student of interaction analysis has no single article or theoretical exposition to consult but, instead, needs to get acquainted with a number of paradigmatic articles and research studies. It is partly through such "finger exercise," as Kuhn refers to it, that a researcher learns how to implement an empirical study of classroom processes.

Assumptions of the Traditional Interaction Analysis Paradigm

1. A basic assumption within the interaction analysis paradigm is that the social-emotional climate influences behavior. In a school and class setting, this means that a positive social-emotional climate is beneficial for almost any aspect of education. Various researchers have used somewhat different terminology to express roughly the same basic assumption.

2. It is generally assumed that the social-emotional climate is a group phenomenon and that the teacher's behavior is the most important single factor in creating climate in the classroom.

3. The teacher's verbal behavior is assumed to be a representative sample of his total classroom behavior. As a result of this assumption, it is commonly considered sufficient to observe and record only the verbal behavior of the teacher and students in the classroom.

4. The decision to focus exclusively, or mainly, on the recording of overt verbal interaction is enhanced by the assumption that verbal

behavior can be observed with greater reliability than nonverbal behavior.

5. It is assumed that the study of classroom interaction cannot be done by means of self-reports by the teacher and the students, e.g., through questionnaires or checklists. Interaction must be observed and recorded by an observer who is not simultaneously engaged in that interaction.

6. It has been assumed that observers could be trained to give a faithful record of what actually transpires in the classroom. In addition, it has been assumed that someone trained in the observation method could also decode an observation protocol and, as it were, reconstruct the interaction.

We have already discussed in general terms the development of observation recording instruments, how they have been classified and how the special terms associated with them have been defined. At this point, we will discuss some early studies based on the traditional interaction analysis paradigm. Then, since the principal indebtedness of the present study is to the Flanders system, we will describe the Flanders' Interaction Analysis Categories (FIAC) system. We will then narrow the focus to give an account of the interaction analysis paradigm within physical education. Finally we will discuss studies in physical education that have used adapted versions of the FIAC system.

Early Studies of Teacher Behavior

The formal study of teacher behavior had its origin in the Progressive Education Movement under the influence of Harold Anderson (1939) and the research group consisting of Kurt Lewin, Ronald Lippitt and Ralph White (1939). These early researchers felt a need to make classrooms more student-centered, to abandon the autocracy of education, and to promote the ideals of democracy and group dynamics. The climate of the classroom became very important.

Using the notion of a "social emotional climate," Anderson conducted systematic studies into the effects of teacher behavior upon pupil behavior. The psychological assumptions of these studies are that the child learns less if he is given the answers to his school work, and that he grows less in other respects if the teacher makes all the decisions concerning content and procedure. Anderson quantified behavior phenomena and thus provided the basis upon which Flanders later demonstrated that indirect teacher behavior had a positive correlation with child achievement.

Dominative and integrative behavior of the teacher was observed and identified by Anderson with a category system containing nineteen categories: eleven domination categories and eight integration categories. Anderson also showed that it was possible to compute an index, or ID-ratio, by dividing the number of integrative contacts by the number of dominative contacts, and that teachers could be compared using this index criterion.

Lippitt and White (1943), together with Lewin, conducted a series of laboratory experiments for determining the effects of adult teachers' influence on the organized and voluntary activities of boys clubs. Each club was subjected sequentially to an adult playing the role of an "autocratic leader," a "democratic leader," and a "laissez-faire leader." The results of these studies confirmed or extended the general conclusions of Anderson. As a result of these two basic and independent studies, which produced mutually supportive results, the notion of a social emotional climate was established.

Drawing upon the work of both groups, Withall, through extensive analysis, produced an index of teaching behavior which, though almost identical with the integrative/dominative ID-ratio of Anderson, offered a much more refined category system of classroom climate. Withall (1949) defined the concept "social emotional climate" as the "general emotional factor which appears to be present in interactions occurring between individuals in face to face groups" (p. 348). In practice, this "climate" is considered to influence: "(1) the inner private world of each individual; (2) 'the esprit de corps' of a group; (3) the sense of meaningfulness of group and individual goals and activities; (4) the objectivity with which a problem is attacked; and (5) the kind and extent of interpersonal interaction in a group" (pp. 348-349).

Withall emphasized the importance of the teacher's verbal behavior in determining the classroom climate and identified the preliminary categories of his research instrument by recording regular class sessions and analyzing tape-recorded lessons. From this analysis he devised a system of classifying the teacher's verbalization into the following seven categories:

1. learner-supportive statements;
2. acceptant and clarifying statements;
3. problem-structuring statements or questions;
4. neutral statements;
5. directive or hortative statements with intent;
6. reproving or deprecating remarks;
7. teacher self-supporting remarks. (Withall, 1949, p. 349)

These seven categories ranged from 'learner-supportive' statements (1-3) through 'neutral' statements (4) to teachers' self-supporting statements (5-7).

Extensive validational procedures followed the development of this category system to determine the objectivity, reliability and validity of the climate index.

The objectivity of Withall's instrument was reported in terms of inter-judge agreement. Data for computing the indices were obtained by coding teachers' statements contained in three typescripts and the percentage of agreement of each of four observers with the investigator was computed. The percentage agreement of each observer with the mean percentage of agreement ranged from 56% to 75%.

Reliability was evaluated by determining the consistency of the instrument. Day-to-day variations in the pattern of statements of three teachers were compared. The chi-square test was used to check the hypothesis that no significant differences occurred from day to day.

To determine the validity of the climate index, four procedures were used: (1) Anderson's Teacher Behavior Categories as the criterion instrument; (2) pupil evaluations; (3) a Teacher Characteristics Rating Scale; and (4) the description of the class situation from three frames of reference.

As a result of these studies, and later those of Ned Flanders (1965, 1970), the school of interaction analysis was created (Amidon & Hough, 1967).

The Flanders Interaction Analysis Category System (FIAC)

Clearly, the research instrument most often used in classroom studies is the Flanders Interaction Analysis Category System (FIAC) and its modifications (Dunkin & Biddle, 1974). This system is based on social psychology and the theory of the leader/subordinate relationship. A knowledge of Flanders' studies and of interaction analysis is important to the understanding of this particular approach to evaluating measuring instruments, since the choice of a system of classification, as well as decisions concerning its modification, involves adherence to a theoretical frame of reference as its basis.

According to Flanders (1970) the main goals guiding the analysis of teaching behavior are (1) to help the teacher develop and control his teaching behavior, and (2) to investigate relationships between classroom interaction and teaching acts so as to explain some of the variability in the chain of events. (Flanders defined an <u>event</u> in terms of time: whatever goes on during a three-second interval is treated as one event and coded as such.) With this in mind, Flanders' theory is an attempt to explain teacher influence and changes in pupil behavior, in which the intervening hypothetical mechanism is the process of goal clarification. Accordingly, teaching is a process of clarifying and implementing objectives, in which the teacher's task is to act flexibly so that there develops a minimum of dependence in pupils (Flanders, 1967b). In developing his theory, Flanders has introduced some basic changes to classroom research by reconceptualizing the continuum of teacher behavior variability, by moderating Anderson's (1939) "Commitment" in which classroom democracy was always advocated and domination avoided, and by including in his new observational instrument additional categories for judging pupil verbal behavior.

Definition of Terms

The following concepts are used in describing tentative hypotheses of teacher influence (Flanders, 1967b).

<u>Direct influence</u> consists of stating the teacher's own opinion or ideas, directing the pupil's action, criticizing his behavior, or justifying the teacher's authority or use of that authority.

Indirect influence consists of soliciting the opinions or ideas of the pupils, applying or enlarging on those opinions or ideas, praising or encouraging the participation of pupils, or clarifying and accepting their feelings.

The word dependence refers to the essential qualities of a superior-subordinate relationship. The opposite of dependence is independence. Independence refers to a condition in which the pupils perceive their activities to be "self-directed" (even though the teacher may have helped create the perception) and they do not expect directions from the teacher. It is assumed that various degrees of dependence or independence exist.

High dependence refers to a condition in which pupils voluntarily seek additional ways of complying with the authority of the teacher.

Medium dependence refers to the average classroom condition in which teacher direction is essential to initiate and guide activities but the pupils do not voluntarily solicit it. When it occurs, they comply.

Low dependence refers to a condition in which pupils react to teacher directions if they occur, but their present activities, usually teacher initiated, can be carried on without continued teacher direction. In the face of difficulties, pupils prefer the teacher's help.

Dimensions of Classroom Learning and Teaching

One aspect of the classroom situation that should make a difference in the pupil's reaction to teacher influence is his perception of the learning goal and the methods of reaching that goal. One can conceive of a situation in which the goal and the methods of reaching the goal are clear to the pupil, and another situation in which these are unclear. Certainly, when a student knows what he is doing, his reactions to teacher influence will not be the same as when he is not sure of what he is doing. The student may also perceive the goal as desirable or undesirable. The attraction of a goal determines motivation, an attribute which Lewin (1935) designated as positive valence or negative valence.

Changing the mode of teacher influence (direct-indirect) along with the process of goal clarification Flanders (1970) calls "flexibility." Flexibility of teacher acts may explain why direct influence may increase or maintain dependence in one situation, and increase or maintain independence in another. This can be illustrated in the following way (Komulainen, 1973):

Mode of teacher influence	Process of goal clarification		
	Unclear	Clear +	Clear -
Direct	+	+	+
Indirect	-	-	-

+ = dependence increases
- = dependence does not increase

In a different context, Soar (1968) has shown that the level of difficulty of the subject matter presupposes that the teacher uses different modes of influence or flexibility. Creative activity demands a freer setting and less control in order to be optimally successful. Thus, the structure of the subject matter is an important factor in determining authority in use.

Later, Flanders (1970) added to his theory the domain of social access, which consists of social contacts and the range of ideas. The presumption of social access for communication means that most of what takes place in the classroom depends on communication. Who talks to whom forms a network of communication which is closely related to physical access, such as the seating arrangements in a classroom. The opportunities to contact other pupils can be at a minumum when the formation is restricted, whereas if mobility permits pupils to select their communication contacts the formation is free. When the ideas discussed are determined primarily by the teacher, the range of ideas is controlled, and when anything can be discussed, the range of ideas is open. "In most instances, free social contacts also permit a wide range of ideas to be discussed" (Flanders, 1970, p. 316).

The measurement of social contacts can be made by asking observers to make a separate assessment of class formation and to record notes whenever this formation changes. Similar evaluation of range of ideas can be made by using pupil questionnaires to determine whether the pupils' perceptions about expressing ideas is controlled or open.

The possible configurations of these four dimensions of classroom teaching and learning, i.e., goal orientation, authority in use, social contacts, and range of ideas, is illustrated by the use of the following figure (Flanders, 1970, p. 317).

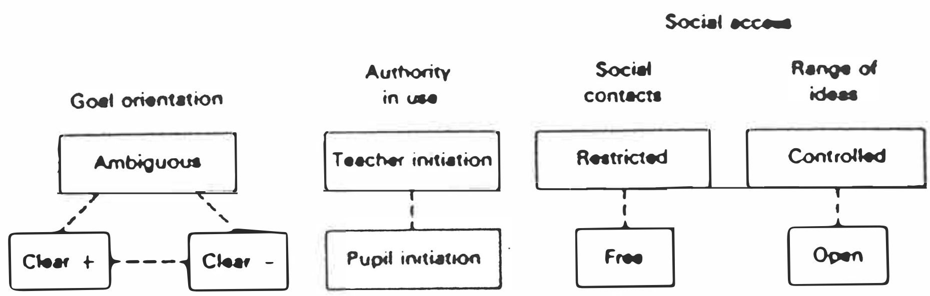


Figure 3. Flanders descriptive model.

Knowing the sequence and variety of the possible configurations in the four domains discussed can help to predict what will happen next. Flanders used the term <u>variety</u> to refer to the total number of different configurations which may occur in the classroom and the term <u>sequence</u> to indicate how many different configuration pairs occurred in a given period.

Flanders sums up his hypotheses concerning the conditional relationships which predict educational outcomes in the following manner:

If...	a certain goal orientation exists (here we begin with the pupils' goal perceptions)
And...	classroom interaction is characterized by a) certain authority in use b) certain social contacts c) and range of ideas social access (here are features of the interaction)
Then we probably expect...	certain educational outcomes, in terms of a) pupil initiation and self-direction b) average pupil attitudes c) average subject matter achievement (Flanders, 1970, p. 320)

The Flanders Observation Instrument

On the basis of his theories, Flanders developed a new observation instrument which was in some ways an improvement to earlier ones and more useful, e.g., as a means of teacher training. Referring to the classifications of Simon and Boyer (1970) mentioned earlier, the Flanders Interaction Analysis Category System focuses upon the first classification, "affective." But, as Flanders points out, it emphasizes both the affective and the cognitive domains in the classroom. In spite of his emphasis on the classroom climate, Flanders was very much aware of the role of the cognitive domain in the classroom. "Every pattern of interaction has a cognitive and an affective component. To understand what goes on in the classroom is to take both into consideration" (Flanders, 1970, p. 270).

Building on Withall's learner-centered/teacher-centered continuum, Flanders identified his teacher talk categories as representing indirect/direct behaviors. Categories 1, 2, 3 and 4 were considered indirect behaviors and categories 5, 6 and 7 represented direct behaviors (Table 1). The continuance of the indirect (integrative)/direct (dominative) dichotomy introduced by Anderson earlier also allowed Flanders to compare teachers in terms of ID-ratios.

The analysis of "initiative" and "response," a characteristic of interaction between two or more individuals, is the major feature of Flanders category system (Table 1). "To initiate," in this context, means to make the first move, to lead, to begin, to introduce an idea or concept the first time, to express one's own will. "To respond" means to take action after an initiation, to counter, to amplify or react to ideas which have already been expressed, to conform or even to comply to the will expressed by others. Flanders (1970) suggests that the teacher is expected, in most situations, to show more initiative than the pupils. His category system was intended to be used to study the balance between initiation and response. He pointed out that a different category system would be needed to investigate other problems of teaching and learning, such as, the effect on class learning of different pupil reactions.

With seven categories of teacher talk and only two of pupil talk in FIAC system, more information is provided about teachers in general, and

TABLE 1. Flanders Interaction Analysis Categories (Flanders, 1970, p. 34)

Flanders' Interaction Analysis Categories* (FIAC)

<table>
<tr><td rowspan="7">Teacher Talk</td><td rowspan="3">Response</td><td>1. Accepts feeling. Accepts and clarifies an attitude or the feeling tone of a pupil in a nonthreatening manner. Feelings may be positive or negative. Predicting and recalling feelings are included.</td></tr>
<tr><td>2. Praises or encourages. Praises or encourages pupil action or behavior. Jokes that release tension, but not at the expense of another individual; nodding head, or saying "Um hm?" or "go on" are included.</td></tr>
<tr><td>3. Accepts or uses ideas of pupils. Clarifying, building, or developing ideas suggested by a pupil. Teacher extensions of pupil ideas are included but as the teacher brings more of his own ideas into play, shift to category five.</td></tr>
<tr><td></td><td>4. Asks questions. Asking a question about content or procedure, based on teacher ideas, with the intent that a pupil will answer.</td></tr>
<tr><td rowspan="3">Initiation</td><td>5. Lecturing. Giving facts or opinions about content or procedures; expressing his own ideas, giving his own explanation, or citing an authority other than a pupil.</td></tr>
<tr><td>6. Giving directions. Directions, commands, or orders to which a pupil is expected to comply.</td></tr>
<tr><td>7. Criticizing or justifying authority. Statements intended to change pupil behavior from nonacceptable to acceptable pattern; bawling someone out; stating why the teacher is doing what he is doing; extreme self-reference.</td></tr>
<tr><td rowspan="2">Pupil Talk</td><td>Response</td><td>8. Pupil-talk—response. Talk by pupils in response to teacher. Teacher initiates the contact or solicits pupil statement or structures the situation. Freedom to express own ideas is limited.</td></tr>
<tr><td>Initiation</td><td>9. Pupil-talk—initiation. Talk by pupils which they initiate. Expressing own ideas; initiating a new topic; freedom to develop opinions and a line of thought, like asking thoughtful questions; going beyond the existing structure.</td></tr>
<tr><td>Silence</td><td></td><td>10. Silence or confusion. Pauses, short periods of silence and periods of confusion in which communication cannot be understood by the observer.</td></tr>
</table>

*There is *no* scale implied by these numbers. Each number is classificatory; it designates a particular kind of communication event. To write these numbers down during observation is to enumerate, not to judge a position on a scale.

(Flanders 1970,34)

therefore how teacher statements influence the balance of initiative and response behavior can be studied only with a particular set of these categories. In general, the quality of the statements is associated with educational outcomes just as much as, if not more than, quantity.

By using Flanders' system, it is possible to identify the quantity and relationship of pupil talk and teacher talk, to classify teacher-pupil behavior, and to record a sequence of verbal events in live classroom situations. The sequence of verbal events can then be displayed in matrix form where frequencies and relationships of various teacher and pupil verbal behavior patterns may be ascertained. With Darwin, Flanders has also considered matrices as first order Markov Chains in order to compare two matrices (Darwin, 1959; Flanders, 1967a). Similar methods of observation and analysis of data have also been applied by Bales (1950) and Pankratz (1967) and in physical education process analylsis by Varstala (1973).

Flanders has summarized his own seven research projects on social emotional climate together with sixteen other projects that have used his 10-category observation system as a base for investigating pupil learning or behavior with an interaction analysis variable. The results obtained by Flanders tend to support the existence of a consistent, causal and often significant relationship between teacher behavior, as quantified by the FIAC system, and the social emotional climate, as measured by attitude scales. Both of these in turn appear to relate to achievement.

> The percent of teacher statements that make use of ideas and opinions previously expressed by pupils is directly related to average class scores on attitude scales of teacher attractiveness, liking the class, etc., as well as to average achievement scores adjusted for initial ability. (Flanders & Simon, 1970, p. 1426)

In order to assess the effects of classroom interaction, Flanders refers to the reports of 18 research projects, the purpose of which has been to investigate at different levels of education the effectiveness of using interaction analysis as a means to facilitate learning. A general objective of such programs has been an awareness of teaching behavior and the development of flexible teaching behavior. The findings of these research projects give rise to the following generalizations:

1. An individual becomes more responsive to pupil ideas ... by learning how to code with categories of interaction analysis and by interpreting displays from specimens of his own teaching and the teaching of another person.
2. Teaching behavior becomes more flexible (or variable) as a result of studying interaction analysis.
3. The attitudes of college students toward teaching and programs of the preparation of teachers become more positive for those who study interaction analysis compared with those who don't. (Flanders, 1970, pp. 354-356)

Interaction Analysis in Physical Education Research

Although descriptive analytic research involving interaction analysis has gained considerable popularity among educators over the last three decades, physical educators for the most part have failed to acknowledge the benefits of such research. In more than a hundred studies reviewed by Dunkin and Biddle (1974) which have dealt with applications of the FIAC system and related instruments, none of them were used in the context of physical education.

After reviewing 700 American descriptive-analytical studies on physical education, Nixon and Locke (1973) concluded that such research was in its infancy in the early seventies and had only begun to come to grips with the problems and prospects of fruitful investigation. It has consisted mainly of fairly unsystematic surveys of various features of teacher-pupil interaction and has generally been colored by attempts to improve the effectiveness of teaching. The focus of these surveys has been sometimes on the teacher, at other times on the pupils, and again on particular behaviors of both parties, such as teacher talk, pupil movement, contents of physical education, etc.

In physical education research, there has been a total lack of a unified theoretical basis, or even a general model of the teacher-pupil interaction process. This has been considered a serious drawback, which slows down the progress of research. As Nixon and Locke state, "it has been difficult to classify, evaluate and co-ordinate investigations"

(Nixon & Locke, 1973, p. 1129). As a result, our knowledge of teacher-pupil interaction in physical education is rather modest. (See, e.g., Mosston, 1966; Anderson, 1971; Locke, 1977; Pieron, 1983).

Observation Instruments in Physical Education Research

In the last decade, there have been attempts to construct measuring instruments for the observation of the teacher-pupil interactive process in physical education classes. Again Flanders' FIAC system and its modifications have been the most frequently used as in similar studies of other classroom situations (Locke, 1977; Cheffers & Mancini, 1978; and Pieron, 1983). The results and experiences gained from these relatively few studies are suggestive of new directions for developing the observation instrument.

In the development of these instruments, perhaps the most crucial question has been to decide to what extent the original Flanders category system should be extended. How many categories, subdivisions, and/or dimensions are needed to get an adequate description of the interaction process in physical education classes and on the other hand how many are feasible? How should the adapted, extended category system be used to gain objective coding results? These questions have been answered in different ways by investigators whose modified observational instruments have been constructed for different purposes. It is useful to review these instruments in terms of the features which were modified, such as content, format (number of dimensions), categories and subdivisions, as well as conceptual posture, units of analysis, and the methods used for determining the reliability and validity of the instruments.

In most cases, the purpose of the investigators in constructing these modified category systems has been to develop and test an instrument for objective observation in order (1) to describe the teaching-learning process in physical education classes (e.g., Cheffers, 1973; Heinilä, 1971, 1974; Nygaard, 1978; Tavecchio, 1977), or (2) to train teachers (e.g., Galloway, 1970; Love & Roderick, 1971; Mancuso, 1973; Underwood, 1977), or (3) to investigate relationships between activities in physical education classes and student growth (e.g., Dougherty, 1971;

Mancuso, 1973; Kemper et al., 1976; Lamarre & Nygaard, 1977). It should be noted that all investigators have considered it necessary, as a prerequisite of validity, to extend the original single-dimensional FIAC system by adding one or more categories or subdivisions for observing the teacher's non-verbal purposeful activities as well (see also, Gasson, 1972; and Splinter, 1980).

Galloway (1962) was the first to attempt to construct an observation instrument for physical education studies. After an extensive analysis for determining the best system for the measurement of nonverbal behavior, he concluded that "no satisfactory procedure for describing nonverbal communication had until that moment been developed" (p. 7). He pursued the topic further and developed an observation instrument based on the FIAC system which was designed to enable an observer to use the categories, time intervals and ground rules of the original Flanders system while recording the nonverbal dimension as well (Galloway, 1970). The new instrument included a procedure for recording nonverbal cues associated with six of the seven teacher behaviors of the Flanders 10-category system. Double coding is used for each behavior recorded, a verbal code from the Flanders system and a nonverbal code from the Galloway system.

Dougherty (1971) used a modification of the FIAC system to discriminate between patterns of teaching. The purpose of this study was to compare the effects of Command, Task, and Individual Program styles of teaching on the development of physical fitness and learning of selected motor skills. The sub-problems were (1) to determine whether a trained observer could, using a modified FIAC system, differentiate between the three styles of teaching used in the study, and (2) to descriptively analyze student attitudes toward the tested styles of teaching.

For the purpose of the study, an eleventh category, "meaningful nonverbal activity," was added to the Flanders system. In addition, the teacher talk categories were subdivided into interaction with the entire group and interaction with individuals. This dimension was not entered into the matrix analysis. A single trained observer was used in this study and no information was provided on the objectivity of the observer nor on the validity of the revised system. However, the scores from the observations were subjected to analysis of variance. The results for the differences among the styles of teaching indicated that the Task and

Individual Program groups had significantly higher ID-ratios than the Command group. It was not, however, possible to differentiate between the Task and Individual Program styles.

Gasson (1972), described the unique setting of physical education as follows:

1. the response of pupils is mainly motor as opposite to verbal,
2. the children are not static but are constantly moving,
3. there are constant changes in spatial relationships between teacher and class,
4. most primary children are eager to move and participate in concrete activities and consequently have a positive attitude toward physical education,
5. the scope of pupils' response is broader than the normal classroom with non-verbal dimension being dominant (p. 3).

For observing this setting, Gasson developed a three-dimensional observation instrument. The instrument used 22 categories to record the verbal behaviors of the teacher and pupils, the location of the teacher, and the nature and amount of child activity. To determine reliability, a "three way checking" was used. That is, the data was obtained in repeated exploratory interobserver reliability tests between himself and two trained observers, using Scott's coefficient. An interobserver reliability of .70 was reported and minimum reliability coefficients were obtained in each of the three dimensions. From the results of this study, Gasson concluded that (1) a reliable instrument had been developed, and (2) there were some indications that some teachers' verbal behavior related significantly to child activity and attitudes.

Mancuso (1973) conducted a study to determine the validity and reliability of an observation instrument which combined the FIAC system with the Love-Roderick (1971) system. To the resulting eleven partly subdivided categories describing the teacher's verbal and nonverbal behavior, she added five categories describing pupil behavior. This single-dimensional system contained 26 categories in all. The data were gathered from simultaneous observations of three observers during a twenty-minute teaching span in a secondary physical education fencing class. A time interval of three seconds was used in coding. The reliability of the instrument was calculated by using Scott's coefficient. Reliability coefficients of .92, .91 and .92 were obtained for the three pairs of observers. The investigator assumed the instrument to be valid because it was based on Flanders' instrument, which was already validated. She concluded, however, that the developed instrument was in need of refinement.

Underwood (1977) developed a single-dimensional interaction analysis system containing nine categories. The first four, Teacher Talk, Demonstration, Class Talk and Class Movement, were subscripted as "response" and "initiate." In addition there was a category of "inactivity." He used two trained observers for live situation recordings. A reliability coefficient of .96 was calculated using Scott's method on data obtained in one lesson recording.

In their studies, Nygaard (1978) and Lammare and Nygaard (1977) used the FIAC system in its basic, unaltered form. They concentrated on analyzing only verbal behavior, applying the system to the observation of audiotaped material. No information concerning reliability was supplied.

The single-dimensional category system (PEIAS) developed by Kemper et al. (1976) contained 17 categories, three of which were identified as Pupil Talk, Actions, and Performances and Demonstrations. In connection with this system, a specially developed computer program was applied for sampling videotaped behavior in real time. Observers coded the displayed behavior by pressing a key on the keyboard of a teletype connected on-line with a LAB 8/e computer. The computer was programmed to record every one-second interval that the key was "on" until the observer pressed another key.

The reliability of the instrument was determined by using Scott's pi. The objectivity of the instrument was operationalized as the degree of interobserver reliability and was assessed with the help of the Kendall coefficient of concordance, W. Three categories yielded a value of W significant at the .05 level, and twelve a value of W significant at the .01 level. Only two categories yielded a non-significant value of W.

The authors note that PEIAS was not standarized or validated. Therefore it was not possible to indicate the absolute position of the teacher on the continuum directive/nondirective, and consequently, it was not possible to say anything definitive about the meaning of inter-teacher differences. They concluded that it was not known which ratio between directive and nondirective teacher behavior is most conducive to learning in physical education. This analysis has been continued using generalizability studies (Taveocchio, 1977; Splinter, 1980).

Cheffers' Validation Study

None of the preceding studies have attemped to test the validity of their modifications of the Flanders instrument. Cheffers (1973) is a notable exception in that he has conducted a comprehensive study which concerns itself with the validation of an instrument designed to expand the FIAC system to describe nonverbal interaction, different varieties of teacher behavior, and pupil responses in physical education. In adapting the FIAC for use in physical education classes, he cited three major limitations on the original system which prevented researchers from identifying the patterns of teacher-pupil interaction during physical education classes:

1. it is concerned only with verbal behavior;
2. it concerns itself with the classroom teacher as the sole body involved in the teaching process; and
3. without ground rule provision, FIAC describes only classes which are conducted in traditional teacher-pupil interaction on a traditional basis without regard for such class structuring as individualized learning and group activity.

The purpose of Cheffers' study was to determine whether his adaptation (CAFIAS) was valid in describing physical activity lessons with greater representativeness (content validity) than the Flanders system. Cheffers' Adaptation of the Flanders Interaction Analysis System (CAFIAS) was a double-category system allowing the coding of behaviors as verbal, nonverbal, or both. In Cheffers' model, the teaching function was not limited to one individual (the teacher), but was identified as either the classroom teacher, another student (coded S), or the environment (coded E). To indicate group or individual teacher interaction, he simply placed either a W (whole), a P (part) or an I (not influencing) beside the relevant code symbol. A five second time interval was used in coding.

For a full analysis, CAFIAS required a 60x60 matrix, which Cheffers reduced to a more workable 20x20 matrix, instead of the Flanders 10x10 matrix. This comprehensive matrix was constructed to describe student behaviors as being predictable, analytical and game playing, or unpredictable and student initiated. CAFIAS was thus meant to be a very flexible research instrument for use in describing educational situations.

Six student volunteers coded the lessons for reliability testing after receiving 15 hours of training to guarantee their proficiency in the use of the new multiple category system. Three of the students used the original FIAC system, and three students used the new CAFIAS along with the investigator. The reliability was estimated by determining the interobserver agreement when lessons were coded using either of these systems. The reliability was then determined by submitting cell rankings to Kendalls' coefficient of concordance, W, and comparing the matrices of the student observers with those of the two main observers. Two comparisons were made, one comparing the main cell (n=10) and the other comparing the total matrices (n was specified 20x20).

All matrices developed for both FIAC and CAFIAS were reported to be concordant to the .05 level of significance and beyond. In two lessons, the badminton lesson and the creative dance lesson, the CAFIAS matrices were significant at the .05 level of significance. All remaining matrices were significant at the .01 level of significance. On the basis of these findings, the instrument was evaluated to be reliable.

Measures of face, content and construct validity were made possible by comparing the scores of trained interpreters answering a questionaire (PAQ). In order to measure the performance of CAFIAS against FIAC, matrices were developed from six carefully selected physical activity classes and were presented to the interpreters. These interpreters were students who were not familiar with either system and interpreted the lessons solely from the information provided by the matrices (known as a "blind" interpretations). This "live" interpretation group served as the control group, allowing comparisons to be drawn between their scores and the scores recorded on PAQ by the two experimental groups. It was found that the control group (outside criterion) scored significantly higher in all interpretations. CAFIAS interpreters were significantly more accurate than FIAC interpreters on the total questionnaire (PAQ), on those questions relative to CAFIAS, and on three of the films of those questions relative to both systems.

Cheffers concluded that observers are able to more accurately interpret physical activity classroom behavior when given a CAFIAS matrix than a FIAC matrix. It also appears that matrices prepared by observers working exclusively on the nonverbal dimensions were not as accurate in representing classroom behaviors as matrices prepared by observers viewing lessons in both verbal and nonverbal dimensions. He also concluded that further tests were needed to determine the sensitivity and feasibility of the instrument for use in physical activity classrooms, such as, e.g., computer programs to make multiple coding systems feasible.

Summary

Some observation instruments have been developed in the last decade for use in physical education studies. The Flanders' system has been applied most frequently and has been modified to a significant extent by varying the coverage, method of data collection and coding procedures, as well as the conceptual posture used. When measuring the affective domain, the results from these instruments are reported in terms of the basic continuum, direct-indirect influence.

Although multidimensional systems have been used most often, the relationships between clusters have not been hypothesized nor generalizations from these relationships made. Correlative techniques were not used to analyze the relationships between the scores of categories of different clusters. The sequence and variety of teaching behavior were analyzed in only a few studies (e.g., Dougherty, 1970; Cheffers, 1973). Critical teaching behavior based on a theoretical model was discussed rarely and only in connection with verbal behavior.

In general, the investigators have considered only observer agreement and have neglected the study of validity. The validation process used by Cheffers with his multidimensional observation instrument (CAFIAS) has been discussed as an example of complicated validational procedures using different types of measurement to determine the degree of face, content and construct validity.

A Critical Discussion of Interaction Analysis Research

In spite of the encouraging results obtained with observation instruments, certain difficulties limiting their use and application, as well as the generalization of results obtained by them, are in general associated with these methods. In addition, each observation method has special problems of its own, and its further development and application depends on the extent to which these problems can be resolved. Several aspects of Flanders-type interaction analyses have been criticized on both theoretical and technical grounds.

The most obvious limitation of the Flanders system is that it measures only a limited portion of the total classroom interaction, verbal. It is based on the assumption that a teacher's verbal behavior is an adequate sample of his total behavior, and that it can be observed with higher reliability than the nonverbal behavior (Amidon & Flanders, 1967b). In discussing methodological problems in classroom research, Dunkin and Biddle (1974) cite Flanders in identifying the crux of the problem.

> One of the best-known series of generalizations stated about teaching is the so-called "law of two thirds" posited by Flanders... . According to this "law", two thirds of the time spent in classrooms is devoted to talk, two thirds of this talking time is occupied by the teacher and two thirds of teacher talk consists of direct influence" (p. 54).

In his investigations of teaching as a stochastic process, Komulainen (1971b) noted other problems associated with the use of this method. For example, the system is suited only to teaching situations where the group of pupils acts as an undifferentiated system under the direction of the teacher. In addition, this method records interaction only in the vertical direction (teacher-pupil), when the system works as an undifferentiated whole (frontal instruction). However, horizontal interaction also occurs in groups of pupils. Komulainen also pointed out that, from the standpoint of models of the instructional process, the forms of teaching are of greater importance than the problems of subject-specificity. "The social form of instructional process decisively affects the number of necessary models" (Komulainen, 1971a, p. 21). (See also Dunkin & Biddle, 1974, p. 416.)

One noteworthy solution for problems of this kind in inter-action analysis is provided by multidimensional parallel codings. Flanders (1967a, 1970) suggested the use of matrices of multidimensional category systems for studying interaction models of critical teaching behaviors. In analyzing other systems, he noted that each one is designed to give emphasis to a particular conceptual framework. In multidimensional systems, elements are grouped into homogenous clusters, and each cluster is given a label. The label is usually, by definition, on a higher level of abstraction than the elements making up each set. Then the relationships between clusters can be hypothesized using the shorthand labels. Finally, from these relationships generalizations can be discussed and predictions made in an effort to apply them in different situations.

Some attempts to resolve the problems inherent in interaction analysis by multidimensional coding and matrix analysis have already been discussed. Cheffers (1973) used a "blind-live" method of validating his instrument, and "outside" and "inside" criteria coded from a videotaped original sequence of events. The comparison was made by using a variance analysis technique. Since this kind of validating procedure is not strictly a laboratory experiment nor simply an experiment in natural surroundings, they are referred to as "quasi experiments" (Cooley and Lohnes, 1976).

The utility of observation instruments is usually determined by indicating the value of the reliability coefficient. Scott's method has often been used for calculating reliability indices. In most cases it signifies intercoder agreement, although within-coder constancy has also been reported in one of the studies (Kemper et al., 1976). The nonparametric coefficient of concordance, Kendalls' W, has also been applied for assessing the reliability of various individual categories or matrices, operationalized as inter-coder agreement.

Perhaps the most critical problem is the conceptual confusion reflected in these instruments. The single-dimensional systems seem to contain overlapping aspects and the categories are not mutually exclusive. This is, however, properly required if Scott's method is to be used for the calculation of the reliability index (Scott, 1955). The multidimensional approach is, from the methodological point of view, more useful than single-dimensional systems. The reliability of the different dimensions must be explored and reported both separately and in combination. The overall reliability method must be supplemented by a method through which the reliability of any individual category can be determined. The level of the reliability index must also be considered.

Reliability coefficients are often based on very small samples of events. The number of observers in the reliability tests reported here has varied from two to six. Using Scott's pi, the values of inter-coder agreement coefficients in Mancuso's (1973) single-dimensional system of 27 categories varied between .91 and .92. In Underwood's (1977) nine category system, a value of .96 was reported. With this method of calculating reliability, these coefficients seem unrealistically high. In the Kemper et al. (1976) 17 category single-dimensional system, the values of within-coder agreement coefficients varied between .67 and .90. With Gasson's multidimensional system, a mean value (pi) of .70 for repeated inter-coder agreement tests was reported, representing the reliability of all three dimensions.

According to Flanders (1967b), a Scott's coefficient of .85 or higher is a reasonable level of performance. Dunkin and Biddle (1974) have also noted that moderately high reliability has been reported in connection with modified single-dimensional FIAC systems. Flanders (1970) has demonstrated that an increase of categories and subdivisions is likely to be related to a decrease in reliability. The same effect has been noted in the studies using multidimensional category systems. The level of .70 accepted by Gasson (1972) seems to be appropriate.

In the studies reviewed above, the instruments have been used only by the developer himself. "Inter-investigation reliability" studies are also needed before making decisions concerning the implementation of these instruments for describing objectively interaction processes, for training teachers, and for testing hypotheses concerning the relationships between context, process and product variables (see, e.g., Rosenshine & Furst, 1973). A more extensive validity and reliability analysis can be demanded of the developer of an observation system intended for widespread use. In such studies it would be appropriate to use different types of reliability coefficients together, because the inadequacy of observer agreements as the sole indices of reliability has been clearly established (Medley & Mitzel, 1963; Komulainen, 1970;

McGaw et al., 1972). It is also necessary for the user and developer of observation systems to provide an adequate sample of data in order to demonstrate that the observations obtained are indeed representative of the universe to which they are claimed to generalize (see Cronbach et al., 1972).

CHAPTER III

REVIEW OF SOME METHODOLOGICAL ISSUES RELATED TO CLASSROOM OBSERVATION

Unit of Analysis

An important decision in developing a measuring instrument is the selection of the unit of analysis. The choice of the unit of analysis for the events of teaching is both a methodological and a theoretical issue. The purpose of the study, the research design, the type of data being sought, and characteristics of the observation instrument need to be considered when selecting a unit of analysis.

Observation instruments differ in their units of analysis according to the teaching events chosen for study. Biddle (1967) has identified the following four possibilities used in different recording instruments:

1. Arbitary unit of time - unit based upon specific predetermined intervals of time
2. Selected naturally - unit depend upon the onset and termination of key events
3. Phenomenal units - indicating a 'natural' break in the sequence of classroom events
4. Analytical units - reflecting the key concepts that are operationally defined by the investigator.

When the aim in selecting a unit of observation is to make it possible to describe the interaction inherent in different dimensions or clusters, and to preserve the sequence of events, the choice of the observation unit is a multistage process related to the rhythm of the events themselves, to the specification of the observation procedures, to the construction of the observation schedule, and to the methods used for analysis.

Selection of Statistical Procedures

There is a variety of studies concerned with the selection of statistical procedures. This selection process is both a methodological and theoretical issue and is related to the validity of the measuring instrument (Flanders, 1970). Most investigators use a class as their statistical unit. In interaction analysis a school class is considered a social system, an indivisible holistic unit, in which the instructional process manifests itself as an interaction process in time, the structural characteristics and sequential processes of which can be described (Bales & Strodtbeck, 1951/1967; Flanders, 1970; Komulainen, 1971a, 1973).

Statistical analysis produces both primary and secondary information. Category distributions and the cell frequencies of sequence matrices represent primary information. From them can be produced `secondary' information, such as indices, factor structures, dimensions, discriminant functions, etc. Flanders (1970) has noted that the utility of the resulting information depends a great deal on the research design, for instance, how time periods are to be combined into a single cumulative display, or how such time periods are related to the purposes of classroom teaching.

In this context some variables describe general characteristics of the teaching-learning situation and the typical progress of events, while others describe differences between teaching situations. Both types of description are needed in the development and evaluation of an observation instrument, when assessing, e.g., the construct validity or the sensitivity of the instrument.

Statistical procedures can be divided into two general types: univariate procedures and multivariate procedures. In univariate procedures a single variable is related to a single outcome, whereas in multivariate studies several variables are combined. The most common procedures are simple correlations and analysis of variance.

Observational studies most commonly use univariate procedures of analysis. The use of multivariate procedures presents serious problems in the interpretation of results and therefore they have been rarely used, as Rosenshine (1971) noted in his review of observational studies. However, these procedures can be used to evaluate the validity and reliability of the measuring instruments (see, e.g., Komulainen, 1973; Koskenniemi & Komulainen, 1969; Medley, 1982).

Problems of Design

Studies of teaching utilize many designs, such as the observation of a single class over many class periods using many variables. In order to organize findings of research on teaching, Dunkin and Biddle (1974) devised a model that grouped variables into four large classes which they labeled presage, context, process and product variables, based on the terminology suggested earlier by Mitzel.

Presage variables concern teacher characteristics such as formative experiences (i.e., age, sex, etc.), teacher training experiences, and teacher properties (i.e., intelligence, motivation, etc.).

Context variables concern the environmental conditions about which the teacher and school officials can do little, and to which the teacher must adjust, e.g., classroom, school and community contexts and pupil characteristics.

Process variables refer to the `actual activities' of classroom teaching, or all observable behaviors of teachers and pupils in the classroom.

Product variables concern the outcomes of teaching. The most frequently investigated product variables are subject matter learning and attitudes toward the subject, both of which involve immediate pupil 'growth' (Dunkin and Biddle, 1974).

Using these terms Dunkin and Biddle classified the designs of research on teaching into four major types: (1) field surveys, (2) presage-process experiments, (3) process-process experiments, and (4) process-product experiments. Most observation instruments designed in the early forties and fifties were aimed at determining relationships between presage and product variables, that is, teacher effectiveness. The validity of measuring instruments is often tested with the use of context-process and presage-process experiments. Experiments concerning process-process relationships are difficult to control (cf. Dunkin & Biddle, 1974; Komulainen, 1978) because teacher behavior is complex and, in part, responsive to pupil behavior. In so called performance-based teacher education programs, methods of interaction analysis have been used as a tool to help identify changes of behavior and to integrate theory and practice.

Process-product experiments have proved to be fruitful in classroom observation. In experiments of this kind, events are manipulated and the effects of different classroom experiences on pupil learning or attitudes are examined. This is the kind of design normally used by Flanders in his experiments (Flanders, 1970). The development of paradigms in this area has led to a division of the teaching process into various component activities which constitute independent variables, and into criteria, such as type of achievement, which are treated as dependent variables, as in the studies of Flanders (1970). This approach has been manifested in the definition of the 'technical skills' of teaching and has led to the development of microteaching and highly controllable arrangements for the modification of teacher behavior (Gage, 1972).

As part of the Finnish investigations into the instructional process (DPA Helsinki Project), Komulainen (1978) studied the developmental changes in the interaction patterns of the DPA classes. For this purpose, he used the content x class x period design in which content and period are repeated measures. This design was limited by the fact that only variables based on unit coding could be used. As a result of this limitation, other factors which might influence development and change are not identified. The methodological examination was confined to the FIAC system. A mixed approach to the analysis, using both hard and soft data, was necessary in drawing conclusions and in interpreting the results and differences between classes for the DPA Helsinki Project (Komulainen & Koskenniemi, 1978).

Reliability Concept in Observation Studies

Each time an instrument is developed, it should be tested for reliability and validity. Reliability and validity are not regarded as a property of the instrument but as that of measurement. The observer and classification system together form the measuring instrument. The distinction between reliability and validity is a problem in observational studies. In general, reliability is the agreement between two efforts to measure the same trait through maximally similar methods. Validity is represented in the agreement between two attempts to measure the same trait through maximally different methods. However, even though a correlation between dissimilar subtests is probably a reliability measure, it is even closer to validity (Campbell & Fiske, 1959).

As stated earlier, reliability is not a property of an instrument but of measurement. It reflects the ability of the instrument to resist the effects of chance and to provide the same measurement results in varying circumstances. The instrument itself is neither reliable nor unreliable. It is that only when the instrument has been used to collect data and the data have been manipulated in some way to produce scores.

In observation studies, the concept of reliability has an entirely different content and significance from what it has, for example, in psychometric testing. An observation instrument is a set of procedures by means of which a trained observer can record and categorize behaviors and features in a quantifiable form. It consists of a number of items, to which the observer responds in some way dependent on the behavior (or feature) he has observed (Rowley, 1976). Categorizing in observation research typically means the placement of each time-unit into certain classes according to a pre-designed plan. Thus, when examining the reliability of a coding problem associated with the development of observation systems, the phase of categorization has to be considered. Because the observer and classification system together form the measuring instrument, the observer becomes an additional source of error of measurement. The measurement results may be more or less reliable depending on the manner in which the instrument is used, on the subjects or features observed, on the number, skill and training of the observers, and on the observation circumstances.

As Komulainen (1973) has noted, "the value of results depends crucially on the accurate use of the metalanguage of the classification system in the coding process" (p. 11). Therefore, in examining the reliability of a coding problem associated with the development of an observation system, attention must be paid both to the quality of informatin utilized and, above all, to the way in which it is used in the coding process.

The questions to be answered, then, are which data yield the reliability index, and, secondly, how can it be computed. Once this much is accomplished, the adequate level of reliability may be determined.

The concept of reliability is understood in various ways, and various methods have been used to determine reliability in observation studies (e.g. Dunkin & Biddle, 1974; Emmer, 1972; Rosenshine, 1971; Rosenshine & Furst, 1973). These differences in turn are due to varying research objectives and methodological solutions (Medley, 1971).

Within the area of classroom observation instruments, the most commonly used form of reliability measure is observer agreement. The agreement coefficient is usually based on whether two (or more) observers were similar in their tally of total events of each type using such terms as between-observer agreement, inter-rater agreement and inter-coder agreement (Rosenshine & Furst, 1973). Komulainen (1970) uses the term inter-coder agreement to emphasize the objective and mechanical nature of observation in contradistinction to the subjective element inherent in judgments. Bellack et al. (1966) and Flanders (1967b), among others, specify reliability only in terms of observer agreement.

A second commonly used form of reliability measure is stability, or coder consistency. This term has many different meanings, but the central idea is that the coder must be capable of repeating his coding later in the same way. Roughly speaking, it refers to the constancy with which the same observer codes identical audiovisual tapes or transcripts at two different times (Rosenshine, 1971).

In addition, the consistency of the trait to be measured is receiving increased attention. As early as 1953, Borgatta and Bales (1953) pointed out that if common elements exist in the condition under which the behavior occurs (i.e., the task, subject, size of groups, etc.), a certain degree of consistency in the interaction pattern may be expected. They also pointed out that in observation studies the term "consistency of observed phenomena" becomes a more correct identification than "reliability of test." Therefore, indices of observer agreement should not be cited as evidence of reliability.

The problem with a series of reliability indices is that each of them measures the effect of only one or two sources of error. The range of sources of error with the multifacet concept and technique of observational procedures is large. Therefore, a major problem is to decide which sources of error in measurement are relevant. In general, the magnitude of errors is regarded as primarily dependent on the type of decisions to be made from scores, as well as on how they were collected. In constructing the theory of generalizability of scores and profiles, Cronbach et al. (1972) state that "there is a universe of observations, any of which would have yielded a usable basis for the decisions." In connection with this theory, the question of reliability, too, resolves into a question of the accuracy of generalizations, or of generalizability. The term "universe" is applied to conditions under which the subjects (or aspects) might be observed, and the term "facet" to conditions of a certain kind. The observations and measures may be classified according to the facet, the observer, the setting in which the observation is made, etc. The facets, alone or in combinations, define the universe. The universe to which an observation is generalized depends on the practical or theoretical concern of the decision maker (Cronbach et al., 1972).

Heinilä (1974) used the term frame factors instead of the term facet in connection with the model constructed for describing the general elements of the research into the interactional process in physical education and of the research strategy. The term "frame factors" emphasizes the characteristic role that different conditions play in regulating the formation of the interaction process. The term "frame factor" will be used here as well for the same reason. The frame factors regulating the formation of the coding process, used alone or in combination, define the universe of generalizability of results.

In the observation studies of Medley and Mitzel (1963), each observer is regarded as a source of variability in addition to the between-person variability. In this study, reliability signified the extent to which the differences between different classes are greater than differences among codings of the same class. Medley and Mitzel used an analysis of variance for estimating the variation attributable to each facet. In this connection the variability of the object of observation was shown to be the most important source of error variance. The inadequacy of inter-observer agreement as the sole estimation of reliability was also indicated.

However, Rosenshine (1971) noted that this meaning of reliability has been regarded as "intriguing" and difficult to interpret, because it asks not only whether the coders are coding in the same way, but also whether the teachers (or classes) are different in the variables of interest. McGaw et al (1972) refined this method by elaborating on the means for measuring differentiation in a situation where teacher behavior is expected to vary. This variance component approach is based on Cronbach's generalizability theory (Cronbach et al., 1972) which enables the researcher to discover multiple sources of error. This method has been applied, e.g., by Tavecchio et al. (1977) to determine the reliability of an instrument constructed to measure the interaction process in physical education classes.

Komulainen (1970), too, in connection with a study to determine the objectivity of coding of a modified Flanders Interaction Analysis System, presented a method in which both reliability components, observer agreement and observer consistency, are taken into account. Videotaped situations were used in this study, with the two codings occurring on occasions placed at three month intervals. The definitions involved in this method are based on the assumption of the presumably high constancy of the trait to be measured. The reliability problem was not regarded as related to the permanence of various features, as in Medley and Mitzel's (1963) study, but to the dependability of the measurement of the features (Komulainen 1970) as in McGaw et al. (1972). Komulainen (1970) determined both the within-occasion reliability (agreement) and between-occasion reliability (stability) indices, and considered the variation of the coefficients computed attributable to different "facets" (school subjects, coder pairs and coding occasions). This assessment was based on the evaluation of the quality of the measurement scale. In this connection Komulainen considered the range of the variation of Scott's coefficient to have properties similar to those of the coefficient of correlation (Cohen, 1960; Komulainen, 1970).

Komulainen (1970) defines <u>inter-coder agreement</u> as the similarity between the codings performed by two independent observers at the point of time T; <u>within-coder constancy</u> as a reliability indicator resulting from recategorizing from a videotape and comparing various codings done by the same person; and <u>between-coder constancy</u> as the agreement between codings of the same situation performed at different points of time. The following simplified schematic representation of a two-observer case indicates how the various agreement indices are formed:

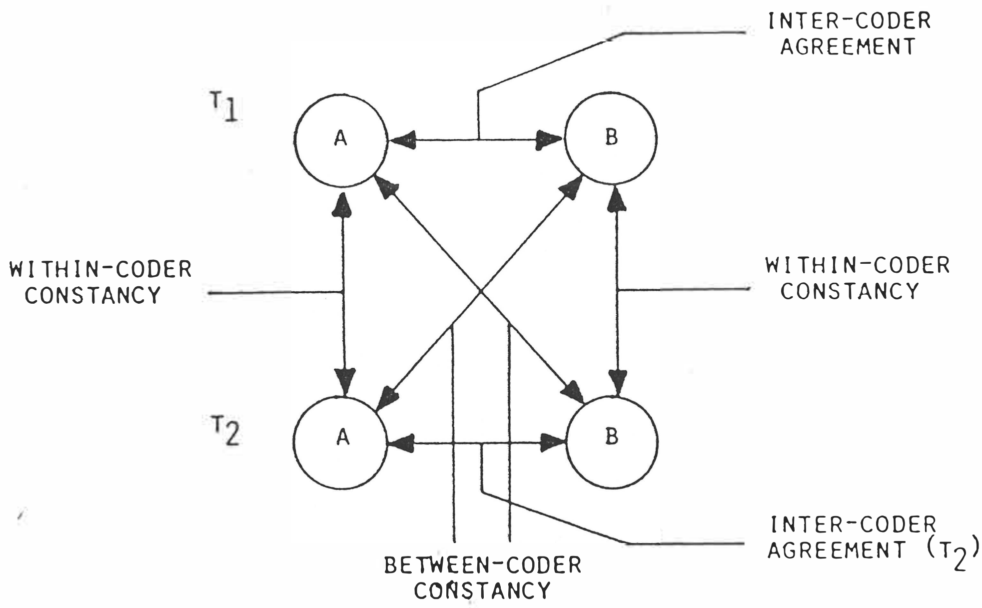


Figure 4. How various agreement indices are formed.

The method presented by Komulainen also enables the researcher to examine multiple sources of error and their characteristics, especially those caused by the coder. As Cronbach et al. (1972) and Komulainen (1973) point out, the lack of reliability does not mean that the majority of classifications occur by chance. The coder's interpretation of the situation and use of the metalanguage of the classification system have been noted to be quite unique. Thus, this "source" is an additional factor causing disagreement. Komulainen (1973) has illustrated it with the following model, showing the factors contributing to reliability, the relations between these and their nature:

/______________ / ______________ / ______________

reliable coder's unique chance

systematic random

According to Komulainen this type of error is a somewhat more important source of error within an observation schedule, however, since it is usually unavoidable. Therefore, the number of coders to be used, as well as their selection and training, need to be studied in assessing the usefulness of a classification system.

Rosenshine and Furst (1973) also address the same problem when comparing observation studies, in which different investigators have used the same observation instruments. They labelled this issue of reliability "inter-investigation agreement." The potential influence of observers is also closely related to the problems in determining the representativeness of coding results. If we accept that there are likely to be systematic differences between observers, then it follows that "error" variation will be greater with a team of observers than if a single observer had been used. However, by using a team the universe of interest is broadened.

In addition, if many items are used, as in a multidimensional classification system, the "error" variation will be greater than if a single dimensional system is used, because the influence of observers will be simultaneously multiplied. Thus the increase in relibaility is almost certain to be accompanied by a decrease in valdiity. Therefore (in this context), the classic theory of measurement erors (where reliability is regarded as a necessary but insufficient precondition of validity) is less descriptive (Cronbach et al, 1972; Komulainen, 1973; Smith & Meux, 1970).

The review of these issues of reliability helps us to confront the problem of multiple criterion measures. Batteries need to be produced which permit multivariate designs. In developing an observational system intended for widespread use, it is important to establish a good <u>within-occasion reliability</u> (agreement) as a necessary but not sufficient condition for stability. It is also important for its own sake when the instrument is intended to be used for feedback in connection with a performance-based teacher education program, where teacher performance is compared to a certain criterion skill used as target behavior. <u>Between-occasion reliability</u> (stability) and associated problems of representativeness are perplexing and need to be studied in this investigation, in assessing the degree of objectivity of coding. <u>Constancy</u> is also important when the observation system is intended to be used as a research tool and the object of the study is to determine if the observed variables are related to some outcome variables (see, e.g., Emmer, 1972; Rosenshine & Furst, 1973).

Unreliability may also be due to very small differences among the objects of observation on the dimension observed. It has, however, been regarded as inappropriate to delete some variables from an observation instrument even if they do not differentiate across classrooms (see, e.g., Bookhout, 1967; Rosenshine & Furst, 1973). It is important to take this point of view into account in developing an observational system, because it needs to be demonstrated that the observations obtained are indeed representative of the universe into which they are claimed to generalize. And as noted earlier, the universe of observations is characterized with respect to one, two or more facets (frame factors).

Estimation of Reliability Indices

The reliability coefficient indicates a correlation between two different uses of the same measurement. The numerical value of it can be calculated by different methods depending upon the research objectives and the nature of the material.

The reliability indices may be estimated on the unitizing level or on the distribution level. In observation studies, we are concerned with measurement events carried out by one or more persons (1, 2...n) on the same or different coding occasions (T_1, T_2, ..., T_n). For example, if two coders carry out a coding of <u>n</u> events independently of each other within an all inclusive and mutually exclusive group of <u>C</u> categories, the result is a square matrix, C x C, portrayed in Figure 5:

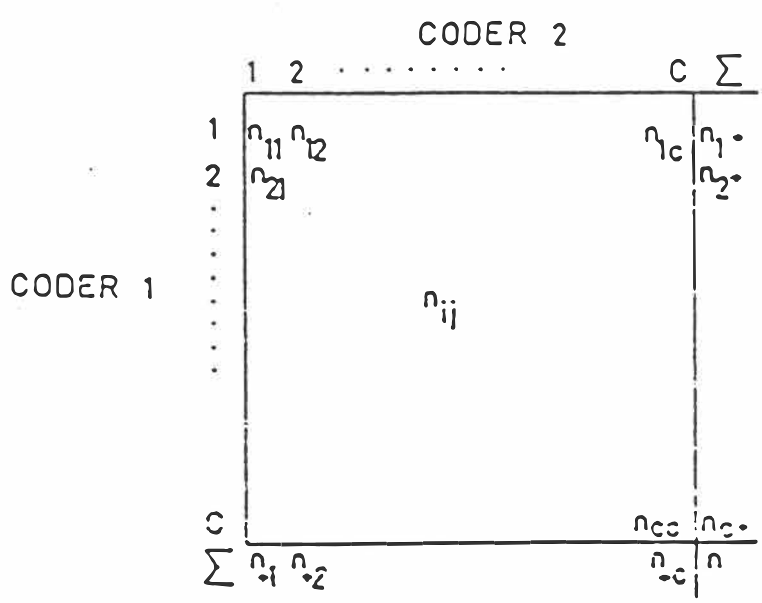


Figure 5. Coding occasion of two coders, with symbols used. (Komulainen, 1974, p. 2)

The reliability coefficients on the distribution level are based on marginal distributions ($n_1 + n_2 + \ldots n_c$), those on the unitizing level on diagonal frequencies ($n_{11} \ldots n_{cc}$) (Komulainen, 1974).

If we wish to study interactional sequences and are using matrix cell frequencies for units of analysis, reliability should be evaluated on the unitizing level. Where the nature and structure of the process are to be studied, marginal distributions may be used as the basis for reliability evaluation (Rosenshine & Furst, 1973). The indices may be applied to single categories or averaged across all categories. Thus they are used to describe the overall reliability of the observation system. In the present study both systems were applied.

For estimating reliability, several indices of agreement and stability have been used, including percentage of agreement, intraclass correlation (usually the product-moment, but occasionally the rank-order coefficient) between two sets of scales, the indices based on perceived agreement give a misleading picture of reliability. For example, where few categories are involved, as in dichotomous coding, the role of chance agreement is great: disagreement in one means agreement in the other, the "errors" are compensating each other.

Therefore, by examining the objectivity of the coding of a multi-dimensional observation instrument with different numbers of categories in each cluster is no reason to align the reliability problem of a category system with the normal measurement of quantitative scales, where reliability is defined as the ratio of true to observed variance (Komulainen, 1973; Valkonen, 1971).

For his Content Analysis, Scott (1955) developed an improved method for estimating reliability in the case of nominal scale coding. Scott's coefficient is a method for estimating observer reliability using any system which assigns events to mutually exclusive categories. It is applied to several categories and takes chance agreement into account by subtracting from each category the proportion of frequencies which would be expected to be in agreement by chance alone. Scott's pi takes into account the fact that the agreement to be expected on the basis of chance does not equal the theoretical expectation but varies according to the relative frequency of occurence of each category (P) in the sample to be analysed. The mean value of the coders' category distribution of the entire sample, and from this the role of chance is computed. Scott's coefficieent provides information not on individual categories, but on the mutual consistency of two coders' entire codings.

Scott's pi is virtually the only reliability index used with the Flanders Interaction Analysis Category System (FIAC). Flanders (1965) argued for this method when comparing it with the adaptation of the Chi-square proposed by Bales, and noted that Scott's method (1) is unaffected by low frequencies, (2) can be adapted to per cent figures, (3) can be estimated more rapidly in the field, and (4) is more sensitive, at higher levels of reliability. Scott's coefficient pi used by Flanders (1965) is determined by the two formulae below:

(1) $$\pi = \frac{P_o - P_e}{1 - P_e}$$

where: P_o = observed percentage agreement

P_e = percentage agreement to be expected on the basis of chance, as obtained from (2)

(2) $$P_e = \sum_{i=1}^{k} P_i^2$$

where: P_i = the proportion of tallies falling to each category

k = the number of categories

(Scott, 1955, p. 321-325)

In formula one, "p" can be roughly interpreted as the amount by which two observers exceeded chance agreement divided by the amount by which perfect agreement exceeds chance (Flanders, 1967b).

Originally Scott's coefficient was designed for computation on the unitizing level (Scott, 1955). However, it is also considered applicable to reliability coefficient computation on the distribution level. Among others, Komulainen (1973) suggests, on the basis of studies on differences of individual categories between agreement coefficients on the unitizing and distribution levels, that the danger of mutually compensating errors due to the use of the frequency totals is not serious.

It can be concluded, after reviewing the possibilities for estimating reliability indices, that the criterion to be used has relevance to the measurement scale, to the role of chance, to the level of calculation of indices, to the choice of the methods to be used for calculating the coefficient, as well as to the objectivity of coding. In addition, the problems of observer training need to be taken into account in this context.

Effective training of coders requires immediate feedback regarding how they have learned to make category discriminations. For that purpose, Flanders (1967b) developed a method which makes it possible to estimate reliability quickly in the field by using a pocket slide rule. He modified Scott's method by converting tallies into percent figures and by developing a graphical method for estimating "P" from the size of the two largest categories. (See Flanders 1967b, 161-166)
This method is also appropriate for the examination of the reliability of the multidimensional observation instrument.

However, coders must be given at least some training before they are able to use observation instruments. Flanders (1967b) graphically describes the problem of observer training as twofold, "first, converting men into machines, and, second, keeping them in that condition while they are observing" (pp. 158).

It was found that individuals differ in their ability to become reliable observers. In general, the persons who have become successful observers have had counseling experience, a broad background in social psychology, or experience as observers in some other system of interaction analysis. Also successful teaching experience, particularly on the elementary level, was found to be a strong predictor of a reliable observer (Flanders, 1967b).

The training procedures used and the length of the training period required need to be considered. In general, the training procedures are related to the observation system used. The more complex the instrument, the more training is required before coders are able to use it reliably. For example, when using the Flanders FIAC system, the categories are first memorized. Then the training begins using a variety of tape recordings of classroom interaction which provide unusual examples of direct or indirect influence patterns. There is an exact category distribution for each tape used. Six to ten hours of preliminary training with tapes is necessary before coders are able to move to the second phase of training, observing in "live" classrooms. During this phase of training the presence of experienced trainers is needed.

Consistent observation by a team requires group training, discussion of common ground rules, each observer's understanding of his own unique biases, and regular meetings after training to discuss unusual categorization problems (Flanders, 1967b).

Flanders described an experiment in which the original Flanders Interaction Analysis System (FIAC) was subscripted to 22 categories. The training period for the new system consisted of 18 hours. Eighteen of nineteen reliability checks produced a Scott's coefficient between .70 and .86 with the median .79. One of the lowest coefficient (.56) occurred during a "difficult" observation and was followed by creating some ground rules which eliminated the difficulty. When all the observations were collapsed to the original 10 categories, all reliabilities were about .05 to .10 higher (Flanders, 1970, p. 141).

On the Concept of Validity

Both reliability and validity require that agreement between measures be demonstrated. A common denominator which most types of validity concepts share in contradistinction to reliability is that this agreement represents the convergence of independent approaches. In connection with observational studies, independence is, of course, a matter of degree. The concept of independence is usually indicated by such phrases as "outside criterion", "external variable", "criterion performance", etc. (Campbell & Fiske, 1959).

To assess validity for an instrument one normally compares scores generated by it against some criterion measure that is known to reflect the phenomenon in which we are interested. To establish validity for an instrument when no criterion is available, Dunkin and Biddle (1974) propose "that we have a theory suggesting a relationship between the phenomenon and something else. If our investigation produces the predicted relationship it is then assumed that the measurement we have made was also valid" (p. 79).

An observation instrument can be examined in terms of its face, content, or construct validity. <u>Face validity</u> refers to the need to show that the instrument is somewhat "obviously" on target with its goal when compared with non-relevant instrumentation. The level of face validity depends on the quality of the category system and of the category definitions, and on whether or not the latter form a facet. The category set forms a facet if the categories provided are mutually exclusive and provide an unambigious classification for each event that is to be coded to one or more facets (Dunkin & Biddle, 1974; Foa, 1965; and Guttman, 1954). For example, in the case of a physical education class, we might use the categories "pupils are collectively moving" and "not passive" to code examples of movement behavior. These two categories form a facet. It is also possible that the instrument may include two or more facets for which the events of teaching should be coded. Most instruments developed for research on teaching using live observation are single-faceted, such as the FIAC system. However, in studies which can take advantage of video-recordings for more complete data, multifaceted category instruments are possible. If the observational instrument includes many facets, the possibilities of recording need to be considered.

Content validity is concerned in observational studies with the relevance of categories to the content area addressed. It measures the degree to which the instrument accurately measures what it seeks to measure in relation to content. Content validity is commonly confirmed through outside criteria, such as a literature search, and through cognitive debate and interaction among specialists in the relevant field.

Construct validity can be defined as the ability of the instrument to distinguish between groups known to behave differently on the construct under study. Construct validity is not related solely to particular investigative procedures, but also to the orientation of the investigator. Once a test constructor hypothesizes that two individual groups will perform differently on his test, and designs an experiment to test this hypothesis, he is exploring its construct validity. When the researcher has no defined criterion measure of the quality with which he is concerned, and must use indirect measures, he will ordinarily test his instrument for construct validity (Safrit, 1973). Here the trait of the quality underlying the test is of central importance, rather than either the test behavior or scores on the criteria (Cronbach & Meehl, 1955).

Campbell and Fiske (1959) discuss convergent and discriminant validation and clarify the criteria to be found in cumulative evaluation considered jointly in the context of the multitrait-multimethod matrix. They show that to demonstrate construct validity, one needs to show that a test not only correlates highly with those variables with which it should (convergent validation), but also that it does not correlate with variables from which it should differ (discriminant validation). The multitrait-multimethod matrix is a systematic experimental design for this type of validation. To examine discriminant validity, and to estimate the relative contribution of trait and method variance, more than one trait as well as more than one method must be employed in the validational process. A careful examination of the multitrait-multimethod matrix (discriminant matrix) will indicate which concepts need sharper definition, and which concepts are poorly measured because of excessive or confusing method variance. Validity judgements based on such a matrix should be taken into account during the development of the instrument, along with the postulated relationships among them, the level of technical refinement of the methods, the relative independence of the methods and any pertinent characteristics of the samples.

The increased use of technical equipment in observational studies makes the testing and evaluation of measuring instruments more efficient. Audiovisual recordings have an immediate appeal for research purposes, because they provide a wealth of details of the two media in which most classroom interaction takes place. However, measurements cannot be valid if the results are subject to error connected with the measurement situation. The effect of using an internal television system for classroom observation has been studied by Komulainen (1968). It was found that the disturbing influence of the television system declined in about three weeks to a level from which it did not decrease any more (Komulainen, 1968, 1971). Honigman (1970) and Cheffers (1973) used audiovisual recordings to validate their multidimensional observation instruments. Both tested the construct validity of their instruments by using the "blind-live" method, assuming that the encoded and decoded data arrays were sufficient to rival "live" or "on the spot" observation. Both found that their data descriptions were more accurate than those taken from live observations, although they did not achieve the same sensitivity as the live observers attained. A number of possible systematic biases were isolated in these studies which may be connected with outside effects such as the technical equipment.

Flanders (1970) deals with the problem of validity in terms of models, and states that although no classroom interaction can ever be completely recreated or repeated, the issue of validity in coding does not rest on the impossibility of recreating what took place. Instead it depends on whether what was encoded did in fact exist and whether the elements of the original situation are recreated in their proper perspective during the decoding process. Validity, therefore, requires accurate interpretation during both decoding and encoding.

CHAPTER IV
RESEARCH PROBLEMS

The main purpose of the present study was to develop and test a system for describing instructional procedures in physical education classes. It is especially concerned with providing good descriptions of teacher-student interactions and does not attempt, for instance, to test hypotheses or to evaluate the effects of such interactions. Since there were no well-established or well-tested procedures for describing instructional procedures in physical education classes when the present project was started, the primary concern was to construct a feasible system.

Therefore this study has a clear methodological orientation. Drawing on theories of the teaching-learning process and on available research, the first research task was to develop a theoretically justifiable system for describing and analysing what happens in the physical education classroom. It is not enough, however, to construct a new instrument or system. The constructed system must be tested to ascertain how "good" it is. Thus, the second major research task of this study was to test the ability of the procedure to yield a faithful description of what transpired in the instructional process.

The new system was developed as a result of three main assumptions: (1) that P.E. classes differ from other classes, especially due to the greater role of the non-verbal behavior; (2) that P.E. classes vary to some extent in terms of their interaction patterns according to the type of class; and (3) that the interaction patterns in P.E. classes vary according to grade level; and (4) that the interaction patterns in P.E. classes vary according to subject area in P.E.

Based on these needs and assumptions, the present study sought to answer the following questions:

1. How can we develop an instrument that is suitable for the description of the instructional process in physical education through observation?
   1.1. What is the state-of-the-art theoretical view of the instructional process?
   1.2. What kinds of instruments have been used in the observation of teaching (a) in general, and b) in

physical education?

1.3. What does research say about the suitability of such instruments?

1.4. What should be the structure of an instrument that is designed to be used for the observation of the instructional process in physical education?

On the basis of such considerations, a system for observing and describing Inteactin process, in P.E. classes was developed.

2. How can we validate the developed instrument?

2.1. How reliable is the system in observing and describing interaction in physical education classes

(a) in live vs. video-recorded situations

(b) at different grade levels

(c) dealing with different types of classes (subject areas)

(d) in relation to the observation of other classes in P.E. using other systems, and

(e) in relation to observations of other classes in other school subjects (particularly the Flanders FIAC system)?

2.2. How valid is the developed system?

(f) What are the proportions of talk vs. movement using the developed instrument as opposed to the proportion of talk in FIAC-type studies? Are there expected differences here?

(h) Does the instrument distinguish reliably P.E. classes held at three different grade levels?

(i) Does the instrument distinguish reliably classes dealing with four subject matter areas?

(j) How does the empirical structure of the obtained data correspond to the theoretical construct structure?

(k) How invariable is the empirical structure across three grade levels?

(e) How invariable is the empirical structure across four subject matter areas?

Definitions

Before embarking on a discussion of the design and methodology of the study, we will define some of the key terms used in the study.

Teaching process

Instruction is seen as a mainly interactive process within school life, aiming at the development of the pupil's personality in accordance with educational objectives. Instruction consists of various situations which are distinguishable from each other by the way activities are arranged. Instruction is a purposive process where teaching is carried out according to internalized goals. The form of instruction refers to the way in which interpersonal communication is organized. It may be group work, problem solving, or programed-teaching, and it may be either direct or indirect.

Interaction is the basic unit of instruction. It presupposes communication between persons, and may be either indirect or direct by nature. In interaction two levels can be distinguished on which communication takes place, the content level and the process level. The interaction process is an advance which proceeds in real time. This interaction process includes the phases of orientation, labor and evaluation.

In communication the following components can be distinguished: message, channels (visual, auditive, psychomotor), sender and receiver.

The content level of communication refers to the subject under discussion and the material that is dealt with during teaching.

The process level of communication is the dual effect of individual behavior on one's self and on the other members of the group.

Observation instrument

An observation instrument is a set of procedures by means of which a trained observer can record and categorize behaviors and features in a quantifiable form. Two observation instruments discussed in this study are:

FIAC: The Flanders Interaction Analysis Category System.

PEIAC/LH-75: Physical Education Interaction Analysis Category System developed by Liisa Heinilä (1974). This system is based on

Flanders' theory (1970) and is a modification and expansion of his FIAC-System (see Heinilä, 1976).

The term nonverbal behavior refers to observable human behaviors which are not expressed verbally. Verbal behavior refers to audible, spoken behavior. Motor activities are those goal-directed movement activities normally considered to be part of the subject matter of physical education such as games, gymnastics, dance, and fundamental movements.

Direct influence refers the teacher's verbal and nonverbal actions which direct the pupil's actions or restrict the pupil's freedom of participation and/or initiation of activity, or criticize his behavior, or justify the teacher's authority or use of that authority. Indirect influence refers to those verbal statements or nonverbal actions of the teacher which encourage a student's participation and/or initiation of activity.

Categorizing means the placement of each time unit into certain classes in each cluster according to a predesigned plan.

Coding means conversion of the content of the instructional process into a form amenable to quantitative treatment.

The term occasion refers to the situation where trained observers are coding with a rule agreed in advance.

The term frame factors refers to the conditions under which the observations and codings are made.

Objectivity of coding signifies the degree of independence between the final results and the coder himself (Komulainen, 1970, 1974).

Inter-coder agreement is the similarity between the codings performed by two independent observers at a point of time (T_1, T_2, or T_3).

Within-coder constancy is the similarity between the codings done from the videotaped material at the point of time 1 (T_2) and the re-coding of the same material at the point of time 2 (T_3) by the same observer.

Between-coder constancy is the agreement between codings of the same material performed by different coders at different points of time (T_2-T_3).

Coding content constancy signifies the independence between the final coding results and the consistency of the coding target in inter-coder agreement, within-coder constancy and between-coder constancy.

Validity

Content validity refers to the degree to which the instrument accurately measures what it seeks to measure in relation to content.

Construct validity signifies the ability of the instrument to distinquish between groups "known" to behave differently on the construct under study.

Sensitivity is the ability of an instrument to make the discriminations required for the research problem (Cheffers 1973).

CHAPTER V
RESEARCH DESIGN AND METHODOLOGY

Chapter Overview

The procedures for constructing and testing the observation instrument are presented and discussed in this chapter. The focus of the first part of the chapter is on the general background and theoretical framework of the research project, and describes the decisions made in constructing the observation instrument. The focus of the last part of this chapter is on the procedures and strategies used for determining the reliability and validity of the observation instrument, and on data collection and analysis.

Construction of the Observation Instrument

The preliminary construction of the research model, and the observation instrument based on it, was done during the period of 1971-1973 (Heinilä, 1974). The observation system developed was based on Flanders' theory (1965, 1970) and on the empirical studies of Heinilä (1970, 1971, 1974, 1976).

The research strategy used for developing the observation instrument and analysis system is illustrated in Figure 6 (Heinilä, 1976).

In general, the decisions made in developing and analyzing the system proceeded along the following lines:

1. specification of the entry situation and selection of a valid theoretical and conceptual framework;
2. the construction of mutually exclusive and exhaustive observable behavior categories derived from the conceptual framework;
3. the selection of a unit of observation and the development of adequate coding procedures for accurate system use;
4. the selection of a unit of analysis derived from the conceptual framework;
5. the determination of acceptable levels of inter-coder reliability (agreement) and intra-coder reliability (constancy levels).

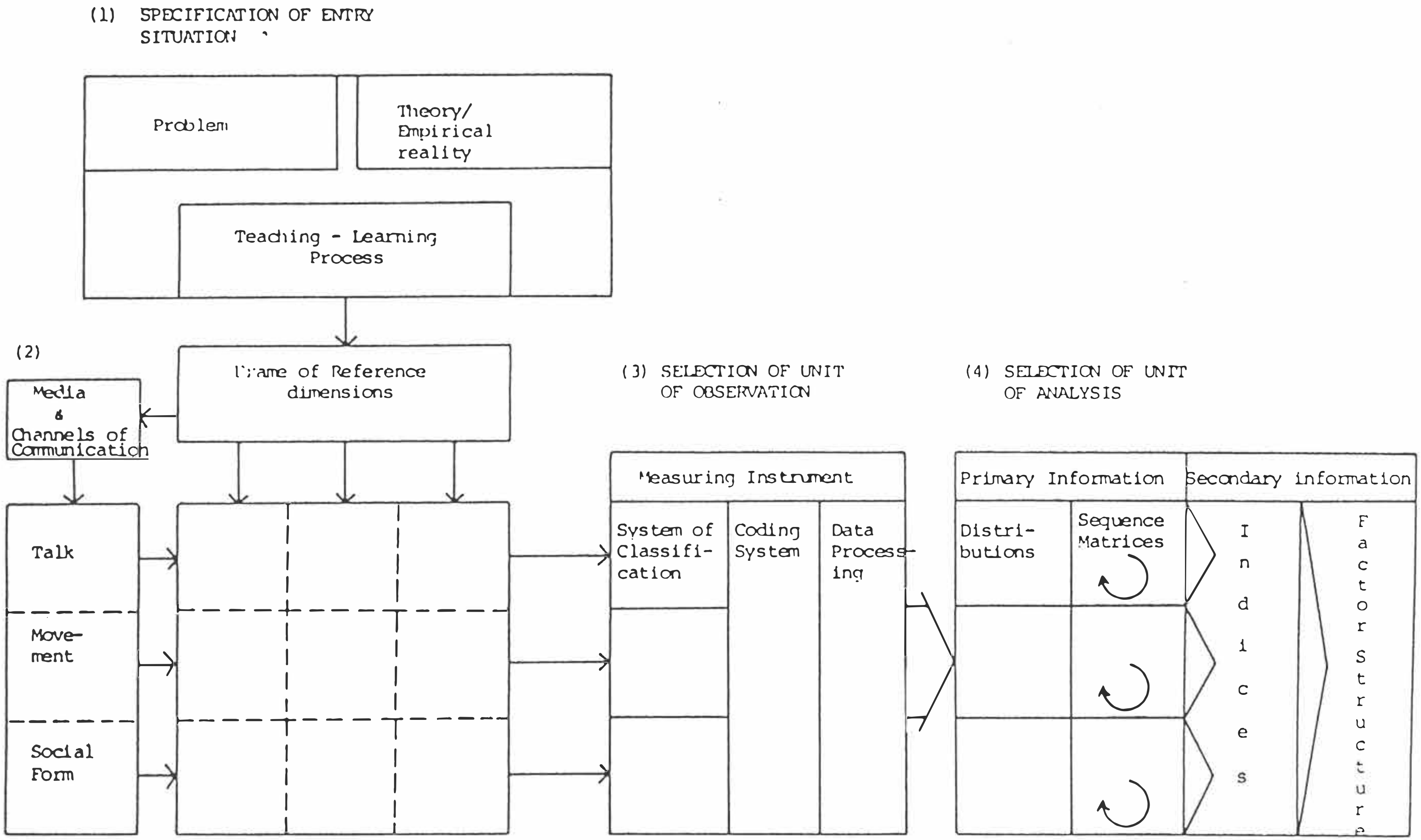


Figure 6. Stages and components in developing a system of analysis (PEIAC/LH-75).

A central problem was the construction of a method for the analysis of the teacher-pupil interaction in physical education in which the different factors of the interaction process and the aspects of communication could be adequately described, and so that the relevant variables of the adopted theory would be sufficiently well represented. The main task of this investigation was thus to have an adequate conception of physical education teaching, and to create an improved system for the scientific measurement, analysis, and evaluation of the physical education teaching process.

The selection of perspective was an important first step because the primary task of descriptive research is to produce an accurate record of significant real-world events. An unlimited number of objects for description and their dimensions may be identified. It is necessary to clarify which events and aspects might be significant to the development of physical education teaching, and to limit the investigation to these aspects.

Problems of content and method in the field of observation research are closely related, and should therefore be examined simultaneously. Often the measuring instrument will also include the theory, as in the classic Bales Interaction Analysis method, and the Flanders Interaction Analysis method, which is perhaps the system most used in process research in the educational sciences. In choosing methods of this kind the researcher has not only made methodological decisions, but has also bound himself to a particular theory and group of variables. In this way the measuring instrument achieves a central significance.

Because of this close relationship between content and method, the basic functions and construct features (characteristics) of physical education teaching events are of particular importance, and must be included in the model developed for the study. Physical education teaching is an interpersonal interaction that is related to the social process of the teaching event and aims at the furthering of the pupils' personality development along the lines laid down by the educational objectives. This social interaction is located in a particular culture and way of life and has certain limitations. By taking these facts as a point of departure, the factors that become base-elements are identified as (1) the teacher and pupils, (2) the teacher-pupil interaction process, and (3) the factors regulating its constructional formation, such as, objectives, materials, and various environmental factors (Heinilä, 1971; Parsons, 1968). With these base elements as a starting point, then, the following model of the interactive process of instruction was developed:

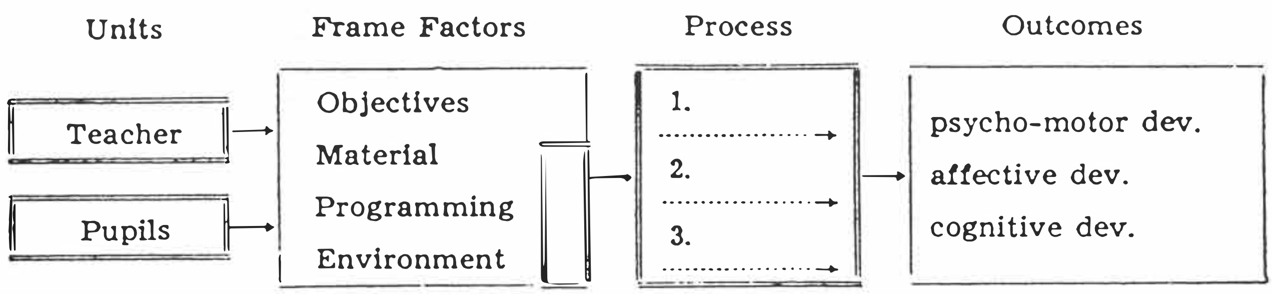


Figure 7. A descriptive model of the teacher-pupil interactive process in physical education (Heinilä, 1976, p. 221)

It is assumed that between the elements of the model, the units, frame factors, processes and outcomes, there is a particular interrelational form which manifests itself as the selection of alternative means as the activity is directed towards the goal.

Assumptions of the Study

Physical education is an action situation in which the form of teaching assumes a central position. In addition, the subject matter contains a lot of affective substance and elements of creativity. A major goal of physical education is the development of pupils' independence and self-direction, i.e., a way of life characterized by physical activity and a permanent interest in physical activity (Heinilä, 1971, 1974, 1976; Komiteanmietintö, 1970a, 1970b).

Movement and physical exercise are typical characteristics of the interaction process in physical education. Movement communicates and movement influences. It is the goal and at the same time a means of attaining the goal. The physiological functions of exercise are realized only through movement activity. Goal-oriented teaching of physical activity is characterized by physical activity. Consequently its occurrence is an essential indicator of the teacher's mode of influence and flexibility. Therefore, the pupils' collective activity and passivity constitute an important dimension in the PEIAC/LH-75 system (see Figures 8 and 9), and at the same time represent the domain of the pupils' activity and social access.

In an active physical education situation, the social form of the participating group and the situation as a whole provide learning experiences. The social form is largely dependent on the teacher's mode of influence, which can be either a stable or transitory feature of the teaching-learning interaction process. Pupils may have different behavioral functions and roles as members of the social group. In this context, behavior refers to activities expressed by members of the group by means of verbal concepts or in symbolic terms, such as movements. Functions are forms of behavior which are purposefully directed towards forming a group or helping it to carry out tasks (Heinilä, 1971, 1974). The teacher can influence the social form of the group by the distribution of labor and responsibility within the group. Labor refers here to the behavior forms and functions that occur in the teaching situation and are similar for all members of the group or specific for individuals or groups. The execution of certain sets of functions by members of the group is referred to as roles.

The observation instrument PEIAC/LH-75 was created to enable researchers to gather valid and reliable empirical data on selected process variables of physical education classes. Such data gathering would provide a comprehensive index of teaching behavior in physical education classes upon which future teaching strategies could be based. Further, it would guide the selection and implementation of teacher training programmes if significant correlations were obtained between the scores of the student rating scale and the behaviors recorded with the observational instrument. It was assumed that with the greater number of clusters, variables and associated techniques for describing and classifying teacher-pupil behavior, the expanded instrument would be more useful and more descriptive in the physical education setting than the original (FIAC) (Heinilä, 1974, 1976).

A cursory examination of the results of the pilot study of this project (Heinilä, 1971) revealed the following: (a) there was a great variety of different configurations connected with the social form, division of labor and responsibility within each lesson and between lessons observed, (b) the data from 15 lessons was characterized by a diversity of content and different forms of teacher-pupil and pupil-pupil interaction, and (c) the face-to-face situation was not common. Consequently, the need for a multidimensional observation instrument was clearly indicated (Heinilä, 1970, 1971).

The Frame of Reference

A frame of reference delimits the area of research, and defines central variables and dimensions and is determined by the research problem and a theory relevant to the exploration of the problem. It also guides the selection of the units of observation and analysis.

The balance between teacher initation and response behavior was the focus of the observation, to be objectively measured and described in this context. This frame of reference is presented in Figures 8 and 9. It describes the theoretical and conceptual framework adapted for the instructional process in physical education.

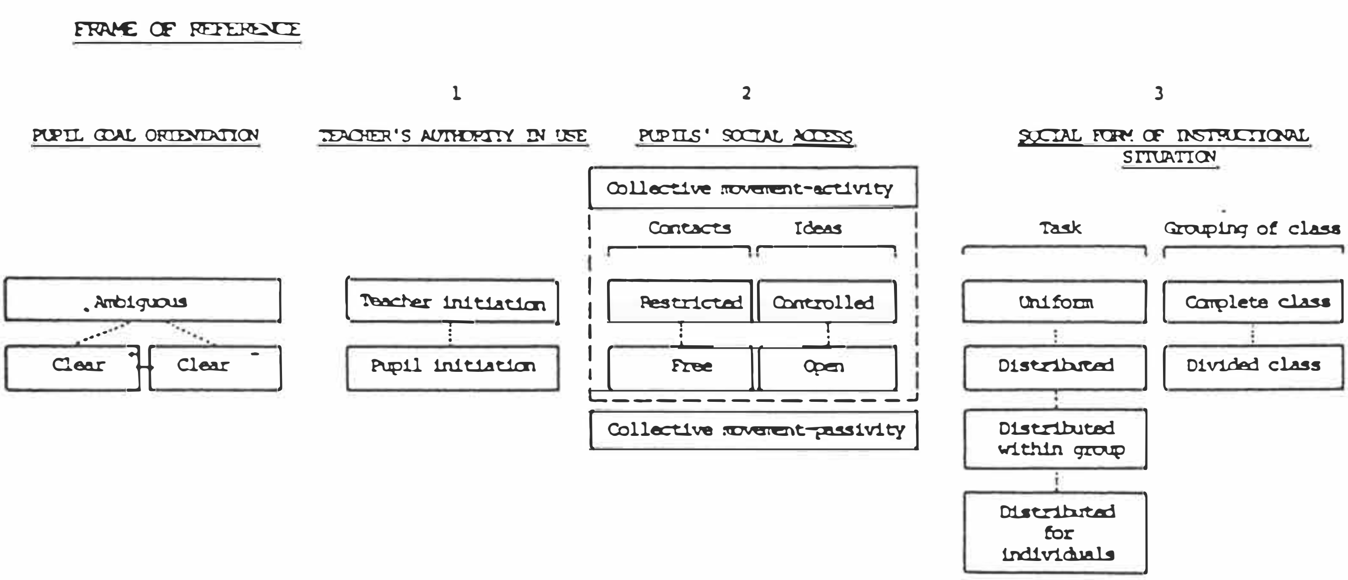


Figure 8. Frame of reference: Dimensions for describing the interaction process in physical education classes.

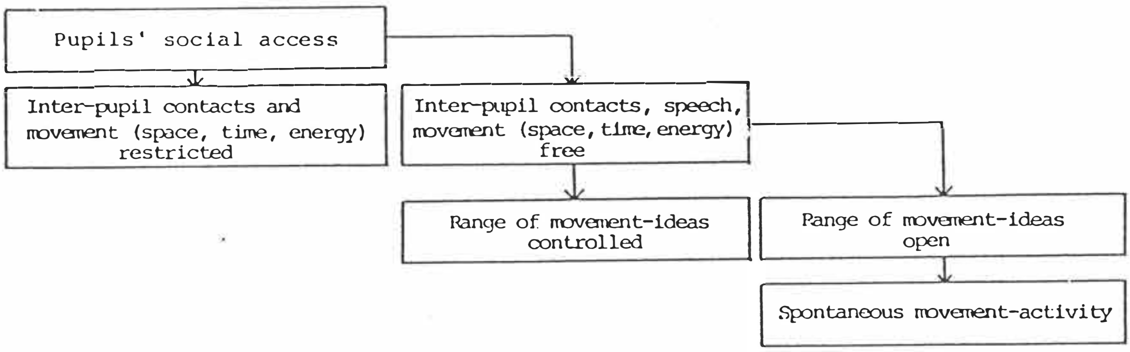


Figure 9. Sequence in degree of freedom of pupils' social access.

Given the research task of developing an observation instrument based on Flanders' theory, the first step was to adapt FIAC to better analyse and describe the interaction process in physical education classes. Flanders' theoretical model of verbal interaction was expanded by adding two aspects which characterize interaction in P.E. classes: (1) the social access in movement activity, and (2) the social form. Accordingly, the three dimensions used to describe teacher-pupil interaction in physical education were (1) the degree of the teacher's authority, (2) the pupils' collective movement activity/passivity and social access, and (3) the social form of the instructional situation. The channels of communication and the media were taken into account in selecting the unit of observation. Thus Flanders' statement concerning his theory of changes in pupils was modified for the PEIAC/LH-75 project to read:

If ...	a certain goal orientation exists (here we begin with the pupils' goal perceptions)
And ...	classroom interaction is characterized by a) certain authority in use b) certain social contacts social access c) range of ideas in pupils' movement activity* (here are features of the interaction) d) and certain social form (here division of labor and responsibility*)
Then we probably expect ...	certain educational outcomes, in terms of a) pupil initiation and self-direction b) average pupil attitudes c) average subject matter achievement.

(* indicates PEIAC/LH-75 modification)

Thus, this adapted theoretical model is an attempt to explain teacher influence and changes in pupil behavior in which an intervening hypothetical mechanism is the process of goal clarification (Figure 10). Each dimension contains a certain aspect of teacher authority in use, but the channels of communication and forms are variable.

Mode of teacher influence:	Dimension 1-3	Pupil goal orientation: ambiguous	clear-	clear+
Direct	verbal/nonverbal movement, social access social form	+	+	+
Indirect	verbal/nonverbal movement, social access social form	-	-	-

+ = dependence increases
- = dependence does not increase

Figure 10. Theoretical model for describing hypothetical mechanism.

The criterion of pupil change toward independence was believed to be an appropriate measure. The strength of this approach resides in the hope that pupil performance of required and self initiated work may be more positively identified and more precisely measured than consequent pupil change.

Selection of the Unit of Observation

The selection of the unit of observation is a process which reflects both questions of principle and technique. The PEIAC/LH-75 system is based on the observation that individual classroom events are meaningful in as much as they constitute part of a sequence, and particularly as they form a sequence of interaction between teacher and pupils. Process is always in a given state. When the aim is to describe the interaction inherent in the talk, movement and the social form of the situation and to preserve the sequence of events, the choice of the observation unit is a multistage process. This is true of both the specification of the methods of observation and coding, and of the construction of the observation schedule.

In the PEIAC/LH-75 system, a unit of time occurring at given intervals was used and tallies were entered in the coding protocol at regular intervals. When category observation is based on regular time intervals, the unit of time also becomes the unit of observation. For this study, an interval of six seconds was used with triple coding. That is, each event was recorded in three different clusters. The nature, extensiveness and specificity of the unit were determined partly by the content and structure of the observation schedule and partly by the time interval.

Variables describing the sequence of events are particularly important in the study of teaching behavior since they may be related to learning outcomes. The sequence of events can be described by means of cell frequencies or indices, or by the models of behavior sequences developed from them.

The selection of the units of analysis for the description of the variables of the teaching-learning process of physical education demands careful consideration and, above all, a continuous development of research methods and their creative application.

Development of Categories

The primary aim of PEIAC/LH-75 was to produce a flexible research instrument for use in describing teachers' authority in use in different physical education situations and periods. The categories of the instrument and their respective dimensions/ headings are described in the following order:

Cluster I Teacher's talk, pupil's talk, teacher's silent activity, other

Cluster II. Pupil's collective movement activity/passivity and the social access

Cluster III. Social form (division of labor and responsibility)

The first cluster was adapted and extended from the Flanders ten-category system presented in Table 1 by making six modifications:

1. Combining FIAC-categories 1 and 2 to form the first PEIAC/LH-75 category, which thus contains acceptance, praise and encouragement by the teacher. The second category of PEIAC/LH-75 is for corrective feedback.
2. Adding to the content of the third FIAC-category (the use of the ideas), "movement patterns suggested by pupils."
3. Adding to the content of the fourth FIAC-category (asks questions), "initiates, terminates movement activity."
4. Adding to the fifth FIAC-category, "demonstration of movement pattern".
5. Adding to the sixth FIAC-category "gives direction, comments during activity (pupil expected to comply)."
6. The addition of two categories for meaningful nonverbal teacher activity:
   "Category 10. Teacher follows pupils' activity, silent guidance," and
   "Category 11. Teacher silent participation in movement activity (such as dancing, playing games)."

Thus, the final categories of the PEIAC/LH-75 system are as shown in Table 2.

The classifications in Cluster I were determined not only by the teacher's but also by the pupils' verbal expressions, as a result of which a certain social form was described.

In Cluster II, collective activity (categories II/1-II/4) refers to movement-activity which has a learning function. The classification was made through observation of the activity in the entire class and the degree of the pupils' freedom in movement, social contacts and range of ideas. It was used when one half of the pupils were moving. The category spontaneous activity (II/4) was used when pupils were allowed to

Table 2. PEIAC/LH-75 Categories.

Cluster I: Teacher talk, movement, pupil talk, other

TEACHER TALK

01. Accepts and clarifies an attitude or the feeling/tone of a pupil in a non-threatening manner. Feelings may be positive or negative. Predicting and recalling feelings are included. Praises or encourages pupil action or behavior. Jokes that release tension, but not at the expense of another individual; nodding head, or saying "Um hm" or "go on" are included.

02. Gives corrective feedback, directs, clarifies, answers pupil's questions.

03. Makes use of the ideas and movement patterns suggested by a pupil or group of pupils: clarifies, expands, builds, questions and movement initiations on the ideas expressed by a pupil. Summarizes pupil's ideas or movement patterns, asks a pupil to demonstrate. Compares the ideas or movement patterns expressed by one pupil to those of another or to those given, repeats pupil's ideas, asks a pupil to demonstrate.

04. Asks questions, initiates, terminates activity: Asks questions requiring narrow answers, initiates short-term activity, terminates activity. Broad, open questions which clearly permit choice in ways of answering and moving.

05. Content emphasis:
Presents information, opinions, demonstrates movement patterns, make a pupil demonstrate or citing an authority other than a pupil.
Organizes pupils, material, division of labor and responsibility.

06. Gives directions, commands during activity (pupils expected to comply)

07. Criticizes pupil behavior, rejects movement pattern, justifies authority. Statements intended to change pupil behavior from nonacceptable to acceptable pattern, bawling someone out; stating what the teacher is doing; extreme self-reference.

Table 2. (cont.)

PUPIL TALK	08.	Pupil answers question made by the teacher.
	09.	Pupil initiates speech, asks for instructions, expresses own ideas or movement patterns.
SILENT TEACHER ACTIVITY AND OTHER	10.	Teacher follows pupil's activity, silent quidance
	11.	Teacher's silent participation in movement activity.
	12.	Confused situation, uproar, periods of confusion in which communication cannot be understood by the observer.

Cluster II: Pupils' collective movement activity/passivity and social access

ACTIVITY	1.	Inter-pupil contacts and movement, space, time, energy restricted; range of ideas controlled
	2.	Inter-pupil contacts and/or movement free; range of ideas control
	3.	Inter-pupil contacts free; range of ideas open
	4.	Pupils' spontaneous activity
PASSIVITY	5.	Pupils follow instruction, demonstration
	6.	Pupils organize themselves, assist in organization
	7.	Pupils wait for turn
OTHER	8.	Confused situation

Cluster III: Social form (division of labor and responsibility)

SITUATION	1.	Complete class, uniform tasks
	2.	Divided class, uniform tasks
	3.	Divided class, differentiated tasks
	4.	Divided class, differentiated tasks distributed amongst groups & within group
	5.	Individual work, uniform tasks
	6.	Individual work, differentiated tasks
	7.	Other situation, confused situation

TABLE 2. Physical Education Interaction Analysis Category System (PEIAC/LH-75)

		I CLUSTER - TEACHER TALK - PUPIL TALK - SILENT TEACHER ACTIVITY category			II CLUSTER - SOCIAL ACCESS (PUPILS' COLLECTIVE MOVE-MENT ACTIVITY/PASSIVITY) category	III CLUSTER - SOCIAL FORM (DIVISION OF LABOUR AND RESPONSIBILITY) category
TEACHER TALK	RESPONSE	01. Accepts, praises, encourages 02. Gives corrective feedback, directs, urges 03. Uses pupils' ideas, accepts, clarifies, develops ideas, movement, tasks suggested by pupils	PUPILS' COLLECTIVE MOVEMENT-	ACTIVITY	1. Inter-pupil contacts and movement (space, time, energy) restricted; range of ideas controlled 2. Inter-pupil contacts and/or movement free; range of ideas controlled	1. Complete class, uniform task 2. Divided class, uniform task
TEACHER TALK	INITIATION	04. Asks, initiates and terminates activity 05. Presents information, uses demonstration, describes, organizes pupils/material 06. Gives directions, commands during activity (pupil expected to comply) 07. Criticizes pupil behaviour, rejects movement pattern	PUPILS' COLLECTIVE MOVEMENT-	ACTIVITY	3. Inter-pupil contacts and/or movement free; range of ideas open 4. Pupils' spontaneous activity	3. Divided class, differentiated task 4. Divided class, differentiated tasks distributed amongst groups & within group
PUPIL TALK	INIT./RESP.	08. Answers question/clarifies, demonstrates 09. Initiates speech (asking for instructions expressing own ideas, movements)	PUPILS' COLLECTIVE MOVEMENT-	PASSIVITY	5. Pupils follow instruction, demonstration	
TEACHER SILENT ACTIVITY		10. Follows pupils' activity, silent guidance 11. Silent participation in movement activity	PUPILS' COLLECTIVE MOVEMENT-	PASSIVITY	6. Pupils organize themselves, assist in organization 7. Pupils wait for turn	5. Individual work, uniform task 6. Individual work, differentiated task
OTHER		12. Confused situation, uproar			8. Confused situation, uproar	7. Other situation,confusion
		The decision on classification is made on the basis of the didactic function of the activity.				

L. Heinilä 1975

move in a certain situation under the teacher's supervision and given facilities, the teacher assisting and guiding if needed. The problems were set by the pupils. On the other hand, movement response (II/1-II/3) means the movement activity which was initiated by the teacher's direct or indirect actions based on his own and/or collective decisions. The term collective movement-passivity (II/5-II/7) indicates that pupils were not moving but were involved in other activity which had a learning function.

The Cluster III observation looks at the social form of instructional situation as a whole, which appears in the division of labor and responsibility. To classify the division of labor and responsibility, those behaviors, functions and roles which the group members displayed during the instructional situation were observed. Functions are behaviors directed purposefully toward building the group and toward helping it accomplish its task. These functions may be permanent of occasional, more or less conscious. The characteristic playing of certain sets of functions by group members is referred to as roles. If tasks are distributed within the group, it is the role functions which are often in question. PEIAC/LH-75, Table 2.

Decisions concerning classifications were made in all clusters on the basis of the didactic function of the activity.

Procedures in Observation and Coding

PEIAC/LH-75 is multidimensional and, therefore, some modification to the observation procedures used in FIAC-system was necessary. Instead of Flanders' three second time interval, a six second time interval was used and the triple coding produced three clusters. The dominant characteristics of the time interval were coded. Naturally, the clusters of the instrument can also be used separately, and with the first cluster, the three second time interval can be used, if preferred.

The procedures of observation in the PEIAC/LH-75 system were as follows: The observer placed himself where he could hear and see both the teacher and the pupils, or the video recording on the TV monitor. He observed the first five minutes from the beginning of the lesson without marking the card. The observation period was started and terminated by marking "1287" in the first and last row of the appropriate column. Then every six seconds, either on hearing the signal or by following the hands of a large clock placed on top of the TV receiver, the observer decided which of the three classes of observation in the classification system the events of the previous six seconds best belonged to. The observer then wrote down the numbers selected while following the events of the next period. Thus he continued for twenty minutes making four digit markings in the appropriate row of the answer card in the six second columns, ten markings per minute. The chronology of the events was retained. A louder signal marked the end of a five minute period, whereupon the observer continued marking in the first column of the row reserved for the next five minutes.

Where certain events in the observation period were unclear, an indication was made in the rows (2 vertical lines) at the beginning or end of that period and a more precise explanation was given on the right-hand edge of the card or on the back. Other features which were necessary for the later interpretation of results were indicated; for example, whether or not the class was divided, the size of the group observed, etc.

The classification time sheet (see Appendix A) was the same as an ADP coding sheet in which the lesson material variables were coded in columns 1-8, the sequence number of the card in columns 9-10, and the observations on the teaching process within the time units in columns 11-78. Before the observation period began, the observer recorded basic information in the first ten columns of the time sheet.

It was essential that the sequence of events be carefully preserved as it was transferred from the observers' coding sheets onto computer punch cards for the statistical processing of the material.

Matrix Analysis

As stated earlier, the purpose of interaction analysis is to preserve selected aspects of interaction through observation, encoding, tabulation and then decoding. Validity in coding depends on whether what was encoded did in fact exist and whether these elements of the original situation are recreated in their proper perspective during the decoding process.

In order to preserve the elements of the original situation for accurate decoding, Flanders used a method of analysis, called the matrix analysis, which records the sequence of events in a classroom in such a way that certain facts become readily apparent. The sequence of number codes were entered into a row/column table, or matrix, in which each column and each row corresponds to one of the observation categories. In the Flanders system, a 10 x 10 matrix was used (Amidon & Flanders, 1967b).

The sequence of events is represented by pairs of code symbols. For example, the sequence 5,5,4,10,10,10,4,5, will read from left to right:

1st pair   3rd pair   5th pair   7th pair

5 5 4 10 10 10 4 5

2nd pair   4th pair   6th pair

The first number of any pair designates the row and the second number designates the column. Note that, except for the first and last symbol, each code symbol is used twice in forming the pairs. When you use this method of pairing, there will be one less tally in the matrix than there were numbers entered in the original record (N-1), and n-1 pairs. This is a convenient way to check the tabulations in the matrix for accuracy.

In order to check for errors in recording, the first step in tabulation is to add the same number (usually the code symbol for silence or confusion) to the beginning and the end of the sequence. When a sequence of code numbers, which begins and ends with the same number, is entered into a matrix without error, the sum of each corresponding row and column will be equal. When this occurs, the matrix is said to be balanced.

In PEIAC/LH-75, the sequence of numbers of the three category clusters was entered separately by cluster, so that the first cluster forms a 12-row by 12-column matrix, the second cluster a 8 by 8 matrix, and the third cluster a 7 by 7 matrix. Separate matrices were made for each episode, with each matrix representing a single type of activity, such as class verbal/nonverbal behavior, movement activity/passivity, or social form.

Interpretation of PEIAC-LH-75 Matrices

There are different arithmetic procedures that can be used to make comparisons between two or more matrices. They all use proportions, so that direct comparisons of numbers can be made, regardless of how long a particular observation lasted. For the PEIAC/LH-75, two general methods were used. First, all column totals were converted to a percent of the matrix total and then were calculated as ratios for which there were normative expectations. This is called a frequency matrix. Second, composite matrices involving thousands of tallies were converted to a common base of 1000. This is called a millage matrix.

Two assumptions concerning the indices were applied in this context. First, when two numbers in a matrix were added or divided, as in the calculation of a percent, the assumption was that tallying within the category system proceeded at a constant rate and each tally was presumed to be an equivalent unit. Second, as soon as an assertion, based on the matrix, was made about the classroom interaction or the social form, it was assumed that the total number of tallies and their configuration adequately represented those aspects of the original interaction which were encoded, within the limitations of the PEIAC/LH-75 category system.

There were certain steps of matrix interpretation used in the PEIAC/LH-75 system, which together formed a situational setting. Adapted from the five steps of matrix interpretation used in the FIAC (Flanders, 1970, p. 98), the first cluster consisted of five steps, the second and third clusters of four steps each.

CLUSTER I

1. Check the matrix total in order to estimate the elapsed coding time (which was usually the same for the three clusters).
2. Check the percent of teacher talk, pupil talk, silence and confusion, and teacher's silent activity, and use this information in combination with...
3. ...the balance of teacher response and initiation in constrast with pupil verbal and nonverbal initiation.

4. Check the initial reaction of the teacher to the termination of pupil talk, or the initiation or termination of movement activity.
5. Check the proportions of tallies to be found in "content cross" and "steady state cells" in order to estimate the rapidity of exchange, tendency toward sustained talk, toward work, and toward sustained nonverbal content emphasis.

CLUSTER II

1. Check the matrix total in order to estimate the elapsed coding time.
2. Check the percent of pupil collective movement activity and passivity, and confusion, and use this information in combination with...
3. ...the balance of teacher response and initiation (social access) with pupil collective movement activity.
4. Check the proportion of the tallies to be found in the "steady state cells" in order to estimate the rapidity of exchange, tendency toward sustained movement activity, and tendency toward sustained movement passivity.

CLUSTER III

1. Check the matrix total in order to estimate the elapsed coding time.
2. Check the percent of the sex differences in social forms and configurations, and use this information in combination with...
3. ...the balance (social form) of teacher response and initiation by division of labor and responsibility.
4. Check the proportion of the tallies to be found in the "steady state cells" in order to estimate the rapidity of exchange, and tendency of social form.

As a final step, consider emerging matrices in combination, together with certain presage and context variables (as classified according to teacher, grade level, and subject area of physical education.)

The Major PEIAC/LH-75 Parameters and Their Calculation

The major PEIAC/LH-75 parameters and the formulas for their calculation are listed in Table 3. These parameters were intended to stimulate thinking about the interaction process in P.E. classes as a sequence of coded symbols and as patterns within a matrix. The indices were based on unit coding, and the statistical procedure used was category frequency matrices, with the data presented in percentages and ratios. They were computed from matrices of the three clusters of PEIAC/LH-75: indices 1-8 and 10 from the Cluster I matrix, indices 11-14 from the Cluster II matrix, and indices 15-18 from the Cluster III matrix. Index 9 was calculated by using marginal frequencies of the categories from the matrices of Clusters I and II. They can be used in interpreting and comparing PEIAC/LH-75 matrices.

It is important in comparing two or more matrices to examine the matrix totals and consider whether the sample is appropriate for the stated purposes. Matrix interpretation must then begin with certain primary features of interaction and continue with the more complex features. These primary and complex features are discussed below for each of the three clusters.

1. The proportion of teacher talk (TT), and...
2. ...the proportion of pupil talk (PT) in percent. Monopolizing talking time is one way to dominate a situation and express one's will. Since power, maturity, authority, and initiative usually lie with the teacher, it is not surprising to discover in P.E. classes that the teacher talked more than half of the elapsed coding time in most visits (Heinilä, 1971.)
3. The proportion of teacher's sustained activity ratio (TSAR) can be determined by calculating the percent of all tallies that lie within the 12 "steady state" cells. This ratio reflects the tendency of teacher and pupil talk, and teacher silent activity to remain in the same category for periods longer than six seconds. The higher

Table 3. PEIAC/LH-75 Indices and Their Calculation

No	Symbol	Name of Index	Cluster	Formulas for calculation of ratios
1	TT	Percent teacher talk	I	$\frac{01+02+03+04+05+06+07}{N_I(\text{=row totals cluster I})} \cdot 100$
2	PT	Percent pupil talk	I	$\frac{08+09}{N_I} \cdot 100$
3	TSAR	Teacher sustained activity ratio	I	$\frac{\text{Matrix I diagonal cells}}{N_I} \cdot 100$
4	TSGPR	Teacher silent guidance and participation ratio	I	$\frac{10+11}{01+02+03+04+05+06+07+10+11} \cdot 100$
5	TRR	Teacher response ratio	I	$\frac{01+02+03+11}{01+02+03+11+06+07} \cdot 100$
6	TQAR	Teacher question and activity initiation-termination ratio	I	$\frac{04}{04+05} \cdot 100$
7	CCR	Content emphasis ratio	I	$\frac{04+05}{N_I} \cdot 100$
8	PVIR	Pupil verbal initiation ratio	I	$\frac{09}{09+08} \cdot 100$
9	PIR	Pupil initiation ratio (verbal and nonverbal)	I,II	$\frac{09}{08+09} \cdot 100 + \frac{3+4}{1+2+3+4} \cdot 100$
10	TPR	Teacher praise ratio	I	$\frac{01}{01+07} \cdot 100$
11	PCA	Percent pupil collective activity	II	$\frac{1+2+3+4}{N_{II}(\text{= row totals cluster II})} \cdot 100$
12	PSAR	Pupil sustained activity ratio	II	$\frac{\text{Matrix II diagonal cells}}{N_{II}} \cdot 100$
13	PSAR	Pupil social access ratio	II	$\frac{3+4}{1+2+3+4} \cdot 100$
14	PIOR	Pupil collective activity following instruction, organizing ratio	II	$\frac{5+6}{N_{II}} \cdot 100$
15	SGWR	Pupil social group work ratio	III	$\frac{3+4}{1+2+3+4+5+6} \cdot 100$
16	PIWR	Pupil individual work ratio	III	$\frac{5+6}{1+2+3+4+5+6} \cdot 100$
17	SFVR	Social form variability ratio	III	Number of categories used (max 6)
18	SSFR	Sustained social form ratio	III	$\frac{\text{Matrix III diagonal cells}}{N_{III}} \cdot 100$

this ratio, the less rapid is the interchange between the teacher and the pupils on the average, and the pupils may, in fact, be quite silent.

4. The teacher's silent guidance and participation ratio (TSGPR) is defined as an index which corresponds to the teacher's tendency to use silent guidance and participation in pupil activity as, e.g., in pupils' games or dance. The higher this ratio, the more dominant movement communication is in the interaction process.
5. The teacher's response ratio (TRR) is defined as an index which corresponds to the teacher's tendency to react to the verbal and nonverbal ideas and feelings of the pupils. The formula is designed so that the index will be a percent figure, never higher than 100 and never less than zero. This ratio indicates, for example, that the teacher responded to pupil talk or movement activity more often in matrix X than in matrix Y. This index is adapted from the ID-ratio of the Flanders system.
6. The teacher question and activity initiation-termination ratio (TQAR) is defined as an index representing the tendency of the teacher to use questions, and to initiate and terminate movement activity when guiding the more content oriented part of the situation. The TQAR is the percent of all category I/04 and I/05 statements which are classified in category I/04.
7. The content emphasis ratio (CCR) is rather poorly named, since many statements in categories I/03, I/06, I/08, and I/09, as well as the teacher's silent activity categories, I/10 and I/11, are also concerned directly with content. However, the content emphasis does isolate those teacher statements which are least likely to be involved with certain process problems which every teacher must solve, such as presenting information or initiating and terminating movement activity.

8. The pupil verbal initiation ratio (PVIR) indicates what proportion of pupil talk was judged by the observer to be an act of initiation.
9. The pupil initiation ratio (PIR) indicates what proportion of pupil talk and movement activity was judged by the observer to be an act of initiation.
10. The teacher praise ratio (TPR) is defined as the tendency of the teacher to praise or integrate pupils feelings into the class discussion, or movement activity, at the moment the pupils stop talking or moving, or while they are still moving.
11. The pupil collective activity ratio (PCA) indicates what portion of pupil time was judged by the observer to be movement activity, which is a general feature of the interaction process in P.E. classes. When this index is average or above, it reflects the teacher's tendency to use movement activity.
12. The proportion of pupils' sustained activity (PSAR) can be determined by calculating the percent of all tallies that lie within the 8 "steady state cells" of the matrix Cluster II. It corresponds to the tendency of pupil collective class time to rest in the same category for periods longer than 6 seconds. The higher the ratio, the less rapid is the interchange between the different forms of movement activity/passivity.
13. The pupil social access ratio (PSAR) indicates what proportion of pupil collective movement activity was judged by the observer to be a movement activity of pupil initiation. It is defined as an index which corresponds to the teacher's tendency to use and to react to the ideas and feelings of the pupils in movement activity.
14. The pupil collective following instruction, organizing ratio (PIOR) indicates what proportion of pupil time was judged by the observer to be this kind of movement passivity in preparation for movement activity.

15. The pupil social group work ratio (SGWR) indicates what proportion of pupil time was judged by the observer to be group work based on pupil responsibility. When this ratio is average or above, it reflects the teacher's tendency to divide responsibility among groups of pupils.
16. The pupil individual work ratio (PIWR) is an even more sensitive index which reflects the tendency of the teacher to delegate labor and responsibility to individual pupils, when the ratio is average or above.
17. The social form variability ratio (SFVR) reflects the tendency of the teacher to use different social forms and division of the labor and responsibility in the P.E. class interaction process when the ratio is average or above.
18. The sustained social form ratio (SSFR) can be determined by calculating the percent of all tallies that lie within the 7 "steady state cells" of social form. It reflects the tendency of the teacher to divide the social form. The higher the ratio, the less often labor and responsibility divided.

Training of Observers

When the PEIAC/LH-75 system is used as a research tool, it is employed by trained observers in order to collect reliable data regarding teaching behaviors as a part of a research project. Systematic and thorough training procedures are needed in order to ensure this reliability.

The observers were three men and three women holding bachelor degrees in Physical Education. Their university studies had included, in their second or third year, a 32-hour basic observer course with theory and exercises, in addition to which they received 20 hours' further training for this particular task. During the initial stages of training, the observers coded from tapescripts, audiotapes, and videotapes. The last part of the training program included discussions and illustrations of the perspective. During this period the measuring instrument was given finishing touches and preliminary experiments were made on its applicability. Ground rules for coding were developed to supplement some of the operational definitions for the clusters and categories. At the end of the training period, the intercoder agreement was estimated by using Scott's pi. It was shown to have reached an adequate level (MD .89). Because reliability was controlled during the training period, controls were not applied during the study itself.

After a basic fifteen hour observation course, the categories are memorized, and training begins with tape recordings of interaction in the gymnasium. There should be a variety of training tapes that provide examples of different indirect or direct influence patterns, different aspects of pupils' social access in movement activity, and different social forms. Working with tapes in teams of two or more is recommended. Trainees can then start and stop the playback to discuss each classification. Ten to fifteen hours of preliminary training with tapes is often necessary before proceeding to live situations.

Reliable observation requires consideration of the total situation being observed in order to understand the individual and collective acts and social form being classified. Trainees need to be give ground rules in order to be consistent when choices occur. The general ground rules established by Flanders were adapted to the PEIAC/LH-75 system and applied for categorizing classroom interaction (Amidon & Flanders, 1967a).

Rule 1: When not certain in which of two or more categories a statement belongs, choose the category in Cluster I (speech) and Cluster II (movement and social access) that is numerically farthest from categories I/05, II/2 and II/5.

Rule 2: If the primary tone of the teacher's behavior has been consistently direct or consistently indirect, do not shift into the opposite classification unless a clear indication of shift is given by the teacher (in Cluster I). The same principle will be applied in Cluster II in observing forms of social access and in Cluster III in observing social forms.

Rule 3: The observer must not be overly concerned with his own biases or with the teacher's intent.

Rule 4: If more than one category occurs during the six-second interval, choose in Cluster I the category describing interaction between the teacher and pupils. If no change occurs within six seconds, repeat that category number.

Rule 5: If a confused situation is longer than six seconds,it is recorded as 12 in Cluster I, 8 in Cluster II and 7 in Cluster III.

Flanders (1967b) noted in considering the problems of observer training and reliability, that ground rules two and four seem to be an invitation to biased observation. Yet there is a theory of the "unbiased, biased observer," which recognizes that the observer is biased in the sense that his categorization must be consistent with his general assessment of the teacher's intent for a given sequence of action, but he is unbiased in that he remains open to all evidence that the general intent of the teacher may be changing. During preliminary training, the problem of distinguishing these shifts in categories usually arises. The solution is never fixed or final, but "the observer must learn to be skeptical of verbal habits which are often unreliable cues compared with the total time the teacher talks, the nature of the learning activities, and other more general evidence" (Flanders, 1967b, p. 159). Multiple coding with category clusters is the most flexible system but standardizing the observation procedures and establishing observer reliability may prove difficult.

In general, the observation training and cluster development occurred simultaneously in this study. Observation practice revealed weaknesses in category definitions, with particular categories presenting difficult coding problems. As a result, changes in the coding system were made during the training procedure. Observers need enough training so that the mechanics of recording in three clusters does not interfere with encoding and the more common events are coded consistently. The tempo of recording must be fast enough to accomplish the purpose of the investigation. In this investigation, the training period consisted of 20 hours to guarantee the proficiency of the six observers in the use of the new three-dimensional Physical Education Interaction Analysing Category System, PEIAC/LH-75.

Research Design

Observation always has a definite purpose. Before observation begins there must be a carefully prepared account of the problems the research is meant to explain. This specification will determine the selection of behavior traits, data collection, statistical analysis, and the interpretation of results. The resulting classification system can be based on 1) a theory, 2) a theoretical model, 3) an existing observational system, or 4) the results of empirical research or pilot studies.

The measurements must be directed at what we wish to measure in order to fulfill the requirement of validity. Measurement cannot be valid if the results are subject to different types of sources of error mainly associated with the measurement situation.

The measurement must also be reliable. The greater the effect of random factors on the obtained results, the less reliable the obtained data. The reliability of observational measurement is largely dependent on how objectively the person who does the classification can function. In systematic observation, the important question is how carefully the manual has specified which action should be placed in a certain category, and on the other hand, how well the person who does the classification has understood the manual. In order to verify the coders' classifications, a judge should determine, first, whether or not the classifications correspond to the manual, and, second, to what extent the classifications done by two or more persons coincide. The proposed system needs to be subjected to validation and reliability measures before it can be accepted as a feasible research instrument and as a tool to be used in teacher education.

General elements for testing the validity and reliability of the observation instrument and the research strategy used are illustrated in Figure 11.

In selecting validation procedures, one commonly wishes to know how much of the test variance is attributable to each of a number of constructs, including both the intended constructs and impurities. Factor analysis, often used to explore construct validity, leads to such a report. Since the factors are uncorrelated, the squared loadings can be interpreted directly as a percentage of the test variance (Cronbach, 1971).

It is important in determining validity to address the problem of representativeness (generalizability), that is, the extent to which the sample of lessons represents the interaction taking place in the activity classes concerned.

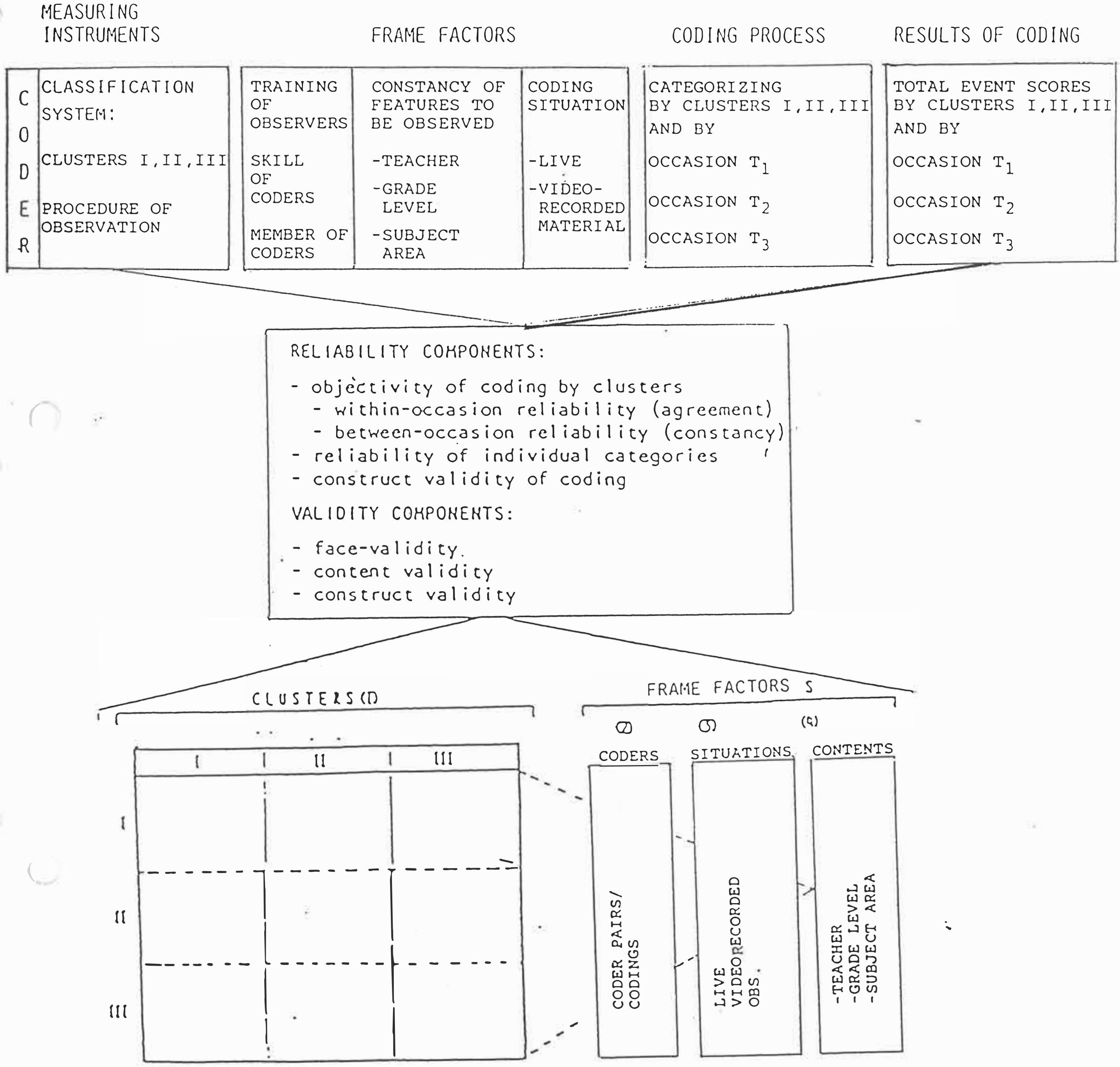


Figure 11. Research model: Determination of validity and reliability of observation.

Of interest from the point of view of validity and sensitivity are (1) how the variables describe the structure of a given group of P.E. classes as classified by, e.g., (a) sex of teacher, (b) age level of pupils, and (c) P.E subject area, and (2) what instructional characteristics are found when one and the same set of data is analyzed by employing a variety of techniques.

A major problem in developing an observation instrument is how to get adequate information for refining the classification system, and especially the rules guiding the observers so that theoretically important concepts can be measured objectively (cf. Komulainen, 1970). In evaluating the usefulness of a measuring instrument, attention must be paid both to the quality of the information available and to the way in which it is used in the coding process. The value of the results of observational studies depends crucially on the manner in which the instrument has been used in the coding process. For this reason, the present study concentrated on the objectivity of coding. In this context it signifies the degree of independence between the final results of coding and the coder himself (Komulainen, 1970; 1973).

Data Collection and Analysis

Several different procedures were used to collect the data for determining the construct validity, sensitivity, objectivity, and reliability of coding of the observation instrument. Each of these procedures was designed to insure a controlled environment for data collection and to satisfy the requirements of a particular phase of instrument testing.

Because the study was not a laboratory experiment, nor simply an experiment in natural surroundings, the variables such as activity lessons were not chosen by means of random sampling. They were selected on the basis of theoretical considerations in an attempt to obtain a sample which would ensure that the variables would vary in a natural way. The sampling contains the activity lessons of two teachers of different sex, with three grade levels and in four subject areas. The coding occasions included both the live situation and videorecorded observation.

The activity lesson material contained different types of structures composed of the categorical elements in the three clusters of the measuring instrument, PEIAC/LH-75. These structures were intended to be either (1) common to all lessons, (2) common to a group of lessons, or (3) unique to a single lesson.

The data (Table 4) were gathered in the Faculty of Physical and Health Education at the University of Jyväskylä, in the physical education teacher training classes taught in the autumn term of 1973. The sample consisted of boys' and girls' P.E. classes at three different grade levels, covering four different subject areas for a total of 24 hours.

Table 4. Research Data

Teacher	N	Level	N	Subject Area	Coding Occasion T_1	T_2	T_3
		Lower	4	Gymnastics	3	3	3
				Apparatus	3	3	3
Man	12	Middle	4	Rhythmic movement-expression	3	3	3
		Upper	4	Ball games	3	3	3
		Lower	4	Gymnastics	3	3	3
				Apparatus	3	3	3
Woman	12	Middle	4	Rhythmic movement-expression	3	3	3
		Upper	4	Ball games	3	3	3
Number of lessons observed					24	24	24
Number of 6-second time units					4800	4800	4800
Number of six coders total time units					28800	28800	28800

Grade levels: Lower=Grades 1-3; Middle=Grades 4-6; Upper=Grades 7-9

The observed classroom activity was recorded using the Faculty's closed-circuit television system (see Appendix B). Visual recording took place with a general-purpose camera manipulated from a control room and with a manually controlled camera in the gymnasium. Audio recording took place with a general microphone and a throat microphone. This arrangement was intended to make the recorded material resemble the live situation as closely as possible.

The six trained coders observed the activity independently three times. They first observed the live situation (T_1), which was at the same time recorded on videotape. Then, one month later they coded from the videotaped material (T_2), and once again in another month's time from the videotapes (T_3). The time order of recorded material was randomized. Each lesson was observed for 20 minutes by the six coders, with the coding beginning five minutes after the start of each lesson. Triple coding was performed by entering four numbers on the answer sheet at six-second intervals.

Data Analysis

The material was processed at the University of Jyväskylä Computer Center in 1974 and 1975 using the Honeywell H 1644 Time sharing system and the UNIVAC 1108/HYLPS programs D.P. and D.F. Scott's coefficients were computed with a special "Scott's" computer program designed for the purpose. The data representing the sequence of events from the six coders' coding sheets (20 coding sheets per coder for each 20-minute observation period), was recorded on computer cards. A detailed discussion of the data is presented in Chapter 6.

To determine the objectivity of the instrument, 8424 Scott's coefficients by coder pair were computed individually by cluster (I, II, III). To determine reliability, mean coefficient values and standard deviations were measured by coding occasions (T_1, T_2, T_3) for inter-coder agreement, within-coder constancy, and between-coder constancy. The variation of these component means and standard deviations was calculated by the different content situations of physical education (teacher, grade level and subject area).

The t-test was used to test the statistical significance between coder pair agreement and constancy coefficients and the same test was used to test the significance of differences between mean agreement and constancy values by cluster and by coding occasion (T_1, T_2, T_3). A one way analysis of variance and a t-test were used to test the statistical significance of differences between mean values of inter-coder agreement, within-coder constancy, and between-coder constancy and the targets of observation (teachers, grade levels and subject areas).

The inter-coder agreement was assessed for various individual categories of the three clusters of the PEIAC/LH-75 by using the Kendall coefficient of concordance, W (Siegel, 1956). In the statistical processing of the material, the sub-program FORTRAN NMCC was applied. To determine the inter-coder agreement, the total percentage of frequencies, per category and per observer, and summed over the sample of 24 lessons, was ranked separately by categories of the three clusters and by occasions T_1, T_2 and T_3. A chi-square test was used for estimating the degree of the statistical significance of the coefficients.

The intra-class correlation coefficient, based on the mean squares obtained from the six observers percentage per category, by cluster, over a sample of 24 lessons (28 800 time units), was used to calculate the reliability of the various individual categories separately on occasions T_1, T_2 and T_3.

The starting point for a discriminant analysis by analysing construct validity of coding were the six observers' score distributions of categories of the 24 lesson data (T_2), as well as the 27 categories of the three clusters of the category system.

For construct validity and sensitivity, the data of every category and cluster were analysed by analysis of variance (ANOVA), in which teacher, grade level and subject area effects were analysed in terms of differences in component variance.

The scores used in calculating indices for each group were obtained from 24 lesson data (T_2) of the six observers' material (T_2) from composite matrices showing the total frequencies and average percentages of marginal frequencies. The significance of differences in means of PEIAC/LH-75 indices between frame factors (teachers, grade levels and subject areas) was computed by using the Mann-Whitney U-test.

A cumulalive multivariate analysis of the factoral structure of instructional situations and a discriminant analysis of the groups formed with the factor scores was used by analysing construct validity and sensitivity of the observation instrument.

CHAPTER VI
RESULTS

Chapter Overview

The fundamental purpose of this research project was to test the reliability and validity of the observation instrument (PEIAC/LH-75) developed by the author for the description of interaction processes in physical education classes.

In this chapter the procedures for instrument testing and the results of each phase of testing are reported and discussed in three parts. The report will begin with a descriptive analysis of the characteristics of the observation instrument in Part I. The reliability and objectivity of coding are discussed in Part II. Part III will report on the construct validity and sensitivity of the observation instrument.

PART I
A DESCRIPTIVE ANALYSIS OF THE OBSERVATION INSTRUMENT PEIAC/LH-75

In this section the characteristics of the measuring instrument and the statistical procedures used in processing the data are presented. The starting point for these analyses was the score distributions and sequence of the categories of the three clusters across class time for 24 lessons as coded by six trained observers on three separate occasions (T_1, T_2, T_3). In addition, certain frame factors such as coding situations, teachers, grade levels and P.E. subject areas. The total coded class time for the sample was 28,800 six-second time units.

The main criterion for assessing the results was: How well does the PEIAC/LH-75 system work? The approach adopted for this study is descriptive. The data should essentially speak for itself, and is presented as directly and simply as possible. Furthermore, the discussion of the results is directed primarily at providing insights into the subtleties of the system and its application and into the limitations of the data. The results will be presented in terms of the following four major components:

1. Describing the use of PEIAC/LH-75 in live and in videorecorded observations, assessed by analyzing the variation of means by categories of the three clusters as a function of the coding situation and as a control repetition coding from videorecorded material (T_1-T_2, T_1-T_3, T_1-T_2).

2. Describing the instructional process by means of the categories of PEIAC/LH-75. Analysis is further divided with respect to variation as a function of teachers, grade levels and P.E. subject areas.

3. Describing the instructional process with PEIAC/LH-75 by using matrix analysis to determine general aspects of sequence and variety in the interaction process across class time by mean measures. Analysis is divided further with respect to variation as a function of teacher, grade level and P.E. subject area.

4. Describing the instructional process by means of major PEIAC/LH-75 parameters and indices, presented in percentages and ratios. Analysis is divided further with respect to variation as a function of teachers, grade levels and P.E. subject areas.

Describing the Use of PEIAC/LH-75 in Live and Videorecorded Observations

Table 5 presents the mean measures and variability for the categories of the three clusters with respect to variation as a function of the coding situations (T_1, T_2, T_3). The data were analyzed by using analysis of variance (ANOVA) in terms of differences in component variance.

The results of this analysis indicated that some categories, especially those which occurred often, were somewhat similar when coded in different situations, while the means of other categories which occurred infrequently were somewhat different for the live situation than for the videorecorded observation. The variation of the means of categories number I/01 (teacher accepts, praises, encourages) and I/03 (teacher uses, develops ideas, movement, tasks, suggested by pupils), was different as a function of the coding situation (4.5-3.1% and 0.8-0.3%) and these differences in means between live and videorecorded observations were statistically significant. This may be due in part to technical problems because a wireless throat microphone was not used to

TABLE 5. Means, standard deviations and percentages of the classtime by categories of three clusters of PEIAC/LH-75. Significance of differences in means estimated between coding occasions: T_1-T_2, T_1-T_3 and T_2-T_3 separately by clusters. N=24 lessons, 28654 time units.

Cluster	Categories	T_1 (live situation) N=24 $\bar{X}$	s	%	T_2 (videorec.obs. 1) N=24 $\bar{X}$	s	%	T_3 (videorec. obs.2) N=24 $\bar{X}$	s	%	difference df=46 T_1-T_2 t	T_1-T_3 t	T_2-T_3 t	Total N=72 X	s	df 2 df 69 F
I	Teachers' talk, movement; pupils' talk; other															
Teacher	01. Accepts, praises, encourages	53.9	34.2	4.5	36.8	22.6	3.1	36.9	21.1	3.1	-2.03^{x}	-2.06^{x}	.02	42.5	27.5	3.25^{x}
	02. Gives corrective feedback, urges	61.1	40.0	5.1	67.3	44.4	5.6	53.0	36.7	4.4	.51	-.73	-1.22	60.5	40.3	.75
	03. Uses, developes ideas, movement, tasks suggested by pupils	9.0	7.2	0.8	3.8	3.9	0.3	4.3	5.2	0.4	-3.10^{xx}	-2.56^{x}	.41	5.7	6.0	6.32^{xxx}
	04. Asks, initiates and terminates activity	98.2	49.3	8.2	80.8	56.4	6.7	86.2	58.2	7.2	-1.14	-.77	.33	88.4	54.5	.64
	05. Presents information, organizes	451.1	122.8	37.6	475.6	107.1	39.6	505.3	118.0	42.1	.73	1.56	.91	477.4	116.6	1.31
	06. Gives directions, commands during activity	51.9	42.8	4.3	46.1	53.4	3.8	37.8	44.9	3.1	-.41	-1.11	-.58	45.3	47.0	.53
	07. Criticizes pupils behaviour	15.0	18.0	1.2	9.3	12.3	0.8	9.0	12.2	0.8	-1.27	-1.36	-.11	11.1	14.5	1.31
Pupil	08. Answers question/clarifies, demonstrates	10.1	9.3	0.8	7.1	9.5	0.6	9.1	10.0	0.8	-1.14	-.39	.71	8.7	9.6	.64
	09. Pupil speaks spontaneously, initiates	23.1	20.5	1.9	20.0	17.0	1.7	16.2	15.3	1.3	-.58	-1.33	-.81	19.8	17.7	.92
Teacher	10. Teacher follows pupils' activity, silent guidance	337.0	159.0	28.1	370.8	155.1	30.9	360.0	161.0	30.0	.75	.50	-.24	355.9	156.7	.28
	11. Silent participation in movement activity	73.3	112.1	6.1	69.8	102.8	5.8	69.8	104.1	5.8	-.11	-.11	.00	70.9	104.9	.87
Other	12. Confused situation	16.3	12.6	1.4	12.6	1.5	1.1	12.4	1.1	1.0	-1.45	-1.50	-.33	13.8	7.5	2.14
		1200		100.0	1200		100.0	1200		100.0				1200		
II	Pupils' collective movement activity/passivity and social access															
Activity	1. Inter-pupil contacts and movement, space, time, energy restricted; range of ideas controlled	177.1	208.0	14.8	136.8	199.8	11.4	125.2	192.3	10.5	-.69	-.90	-.20	146.3	198.6	.44
	2. Inter-pupil contacts and/or movement free; range of ideas controlled	452.3	270.2	37.7	488.0	285.4	40.7	507.7	279.6	42.3	.44	.70	.24	482.7	275.5	.24
	3. Inter-pupil contacts free; range of ideas open	118.6	208.5	9.9	97.0	193.9	8.1	95.6	187.4	8.0	-.37	-.40	-.03	103.7	194.3	.10
	4. Pupils' spontaneous activity	7.1	18.9	0.6	5.8	18.6	0.1	4.0	9.7	0.3	-.25	-.72	-.42	5.6	16.1	.22
Passivity	5. Pupils follow instruction, demonstration	810.7	131.3	25.9	326.3	130.6	27.2	334.9	139.2	27.9	.41	.62	.22	324.0	132.3	.20
	6. Pupils organize themselves, assist in organization	107.2	53.2	8.9	125.6	63.4	10.5	114.3	63.4	9.5	1.09	.43	-.64	115.7	58.7	.59
	7. Pupils wait for turn	12.7	20.4	1.0	7.7	8.9	0.6	5.3	5.8	0.4	-1.10	-1.76	-1.10	8.6	13.4	1.94
Other	8. Confused situation	14.3	5.0	1.2	12.8	1.6	1.1	13.0	3.0	1.1	-1.36	-1.01	.36	13.4	3.5	1.19
		1200		100.0	1200		100.0	1200		100.0				1200		
III	Social form															
Situation	1. Complete class, uniform task	375.0	333.4	31.3	377.7	333.0	31.5	382.9	343.1	31.9	.03	.08	.05	378.5	331.8	.34
	2. Divided class, uniform task	327.4	390.2	27.3	336.0	412.0	28.0	340.3	386.2	29.1	.07	.19	.11	337.6	331.8	.18
	3. Divided class, differentiated tasks	281.0	350.3	23.4	271.5	338.1	22.6	269.5	343.2	22.4	-.10	-.11	-.02	274.0	339.1	.76
	4. Divided class, differentiated tasks distributed amongst groups & within group	107.3	177.4	8.9	107.8	185.1	9.0	100.6	180.4	8.4	.01	-.13	-.14	105.3	178.5	.12
	5. Individual work, uniform task	87.8	177.6	7.3	88.7	175.6	7.4	81.3	161.3	6.8	.02	-.13	-.15	85.9	169.2	.13
	6. Individual work, differentiated tasks	3.6	17.4	0.3	3.0	14.7	0.2	2.5	12.0	0.2	-.11	-.24	-.14	3.0	14.6	.30
	7. Other situation, confused situation	17.9	21.3	1.5	15.3	15.7	1.3	13.9	7.3	1.2	-.47	-.85	-.39	15.7	15.7	.38
		1200		100.0	1200		100.0	1200		100.0				1200		

6 observers
24 lessons
4800 6 second time units, tot. 28800 time units

x = $p \leq 0.05$
xx = $p \leq 0.01$
xxx = $p \leq 0.001$

record the teacher's voice and the voices of the pupils, as was done later (see Heinilä, 1977). In the live situation the aspects of teacher response behavior which are directed mostly to individuals may be easier to recognize.

Thus the systematic observation of physical education classes using the multidimensional category system PEIAC/LH-75 is possible with video-recorded material as well as more sensitive observations in live situations.

Describing the Instructional Process by Means of the Categories of PEIAC/LH-75

The data (the six observers' score distribution of every category of the three clusters for the 24 lessons, 28,654 six- second time units (T_2)) were analyzed with respect to variation as a function of teachers and frame factors by using analysis of variance (ANOVA) in which differences between scores were estimated in terms of component variance. Table 6 and Figure 12.

The score distribution clearly indicates that the teachers observed consistently emphasized their own verbal behaviors (60% of the class time) rather than nonverbal behaviors, and that most of the teacher talk was "initiation." The predominant teacher verbal behavior was "presenting information and organizing" (I/05). The variability of teacher verbal behavior, "silent guidance" (I/10) and "silent participation" (I/11) from class to class was high and related to pupil behavior and especially to the content of instruction, i.e., the P.E. subject area (Table 9). The variation of categories, e.g., the forms of verbal initiation behavior was related to teacher sex (Table 7). The woman teacher used more "initiation and termination of activity" (I/04) and "command during activity" (I/6), which is typical of the "command technique" of women's gymnastics. The interaction on the pupils' part was mostly nonverbal (99% of the class time) and differed somewhat from class to class. Interclass differences were to a considerable degree related to certain frame variables, notably pupil variables, such as sex and age of pupils (Table 8). Pupil speech behavior was mostly initiation.

With regard to pupil nonverbal participation, operational-ized as movement activity/passivity and social access, PEIAC/LH-75 categories

Table 6. Physical Education Interaction Process by Variables of the PEIAC/LH-75: Videorecorded Material (T_2), Means, Standard Deviations, Range, Percentage

Cluster			Categories	N = 24 1200 time units $\bar{X}$	S	N = 144 Md	N = 144 Max-Min	N = 24 %
I			Teachers' talk, movement; pupils' talk; other					
	Teacher talk	01.	Accepts, praises, encourages	36.8	22.6	5.00	19.00-0.00	3.1
		02.	Gives corrective feedback, urges	67.3	44.4	9.00	46.00-0.00	5.6
		03.	Uses, developes ideas, movement, tasks suggested by pupils	3.8	3.9	0.00	9.00-0.00	0.3
		04.	Asks, initiates and terminates activity	80.8	56.4	10.00	53.00-0.00	6.7
		05.	Presents information, organizes	475.6	107.1	78.00	126.00-27.00	39.6
		06.	Gives directions, commands during activity	46.1	53.4	5.00	43.00-0.00	3.8
		07.	Criticizes pupils behaviour	9.4	12.3	0.00	12.00-0.00	0.8
	Pupil talk	08.	Answers question/clarifies, demonstrates	7.0	9.5	0.00	10.00-0.00	0.6
		09.	Pupil speaks spontaneously, initiates	20.0	17.8	2.00	21.00-0.00	1.7
	Teacher silent	10.	Teacher follows pupils' activity, silent guidance	370.8	155.1	57.00	156.00-10.00	30.9
		11.	Silent participation in movement activity	69.8	102.8	5.00	66.00-0.00	5.8
	Other	12.	Confused situation	12.6	1.5	2.00	6.00-1.00	1.1
								100.0
II			Pupils' collective movement activity/passivity and social access					
	Activity	1.	Inter-pupil. contacts and movement, space, time, energy restricted; range of ideas controlled	136.8	199.8	1.00	142.00-0.00	11.4
		2.	Inter-pupil contacts and/or movement free; r. of id. contr.	488.0	285.4	76.00	167.00-0.00	40.7
		3.	Inter-pupil contacts free: range of ideas open	97.0	193.9	0.00	129.00-0.00	8.1
		4.	Pupils' spontaneous activity	5.8	18.6	0.00	18.00-0.00	0.4
	Passivity	5.	Pupils follow instruction, demonstration	326.3	130.6	56.00	105.00-7.00	27.2
		6.	Pupils organize themselves, assist in organization	125.6	63.4	18.00	56.00-2.00	10.5
		7.	Pupils wait for turn	7.7	8.9	0.00	9.00-0.00	0.6
	Other	8.	Confused situation	12.8	1.6	2.00	7.00-1.00	1.1
								100.0
III			Social form					
	Situation	1.	Complete class, uniform task	377.7	333.0	58.00	198.00-0.00	31.5
		2.	Divided class, uniform task	366.0	412.0	27.00	198.00-0.00	28.0
		3.	Divided class, differentiated tasks	271.5	338.1	3.00	161.00-0.00	22.6
		4.	Divided class, differentiated tasks distributed amongst groups & within group	107.8	185.1	0.00	107.00-0.00	9.0
		5.	Individual work, uniform task	88.7	175.6	0.00	90.00-0.00	7.4
		6.	Individual work, differentiated tasks	3.0	14.7	0.00	25.00-0.00	0.3
		7.	Other situation, confused situation	15.3	15.7	2.00	19.00-2.00	1.3
								100.0

6 observers
24 lessons (20 minutes) N = 144
28 800 6 second time units

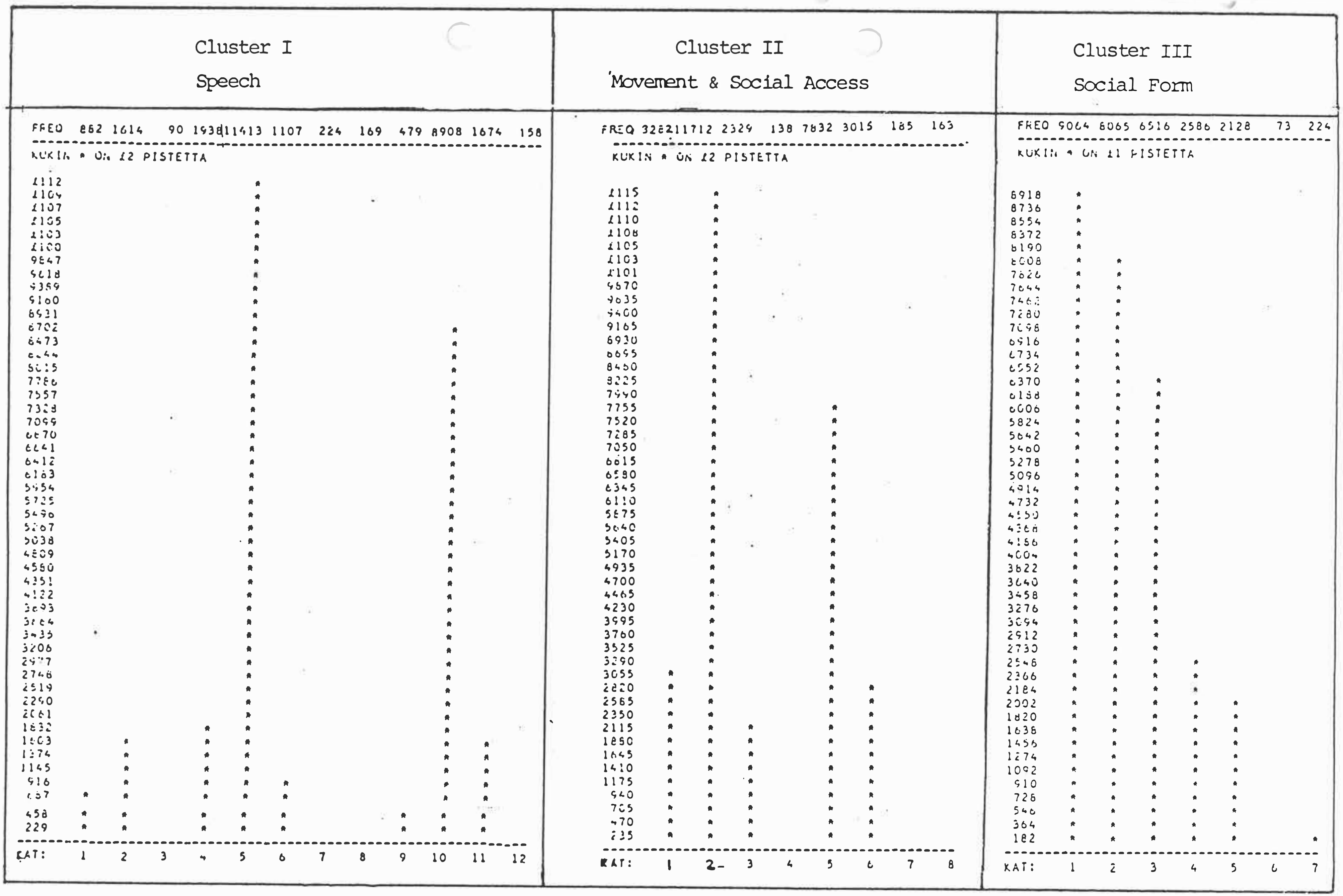


FIGURE 12. Frequence of Time Units for the Categories of Three Clusters of PEIAC/LH-75 Computed from Six Observers Videorecorded Material Observation 1 (T_2), N = 24 Lessons, 28656 Six Second Time Units.

TABLE 7.

Significance of Differences between Means Estimated for the Lessons of Two Teachers (Man-Woman) (T_2); t-test.

Cluster		Categories	Teacher 1.Man (N=12) $\bar{X}$	Teacher 1.Man (N=12) s	Teacher 2.Women (N=12) $\bar{X}$	Teacher 2.Women (N=12) s	t-test dif 1-2 df=22 t	tot. (N=24) $\bar{X}$	tot. (N=24) s	F-test df=1 df=22 F
I		Teachers' talk, movement; pupils' talk; other								
	Teacher	01. Accepts, praises, encourages	44.9	22.4	28.7	20.6	-1.85	36.8	22.6	3.42
		02. Gives corrective feedback, urges	69.6	38.1	64.9	51.6	-.25	67.3	44.4	.64
		03. Uses, developes ideas, movement, tasks suggested by pupils	3.9	3.8	3.6	4.1	-.21	3.8	3.9	.42
		04. Asks, initiates and terminates activity	54.6	42.4	106.9	57.9	2.52^{x}	80.8	56.4	6.37^{x}
		05. Presents information, organizes	484.0	119.9	467.1	97.2	-.38	475.6	107.1	.14
		06. Gives directions, commands during activity	24.3	14.4	68.0	68.7	2.16^{x}	46.1	53.4	4.66^{x}
		07. Criticizes pupils behaviour	13.3	15.6	5.4	6.4	-1.61	9.4	12.3	2.59
	Pupil	08. Answers question/clarifies, demonstrates	6.1	10.7	8.0	8.6	.48	7.0	9.5	.23
		09. Pupil speaks spontaneously, initiates	26.5	17.1	13.4	14.8	-2.00	20.0	17.8	4.00
	Teacher	10. Teacher follows pupils' activity, silent guidance	389.8	176.8	351.8	135.0	-.59	370.8	155.1	.35
		11. Silent participation in movement activity	70.1	95.6	69.4	113.8	-.02	69.8	102.8	.24
	Other	12. Confused situation	12.3	0.5	12.8	2.0	.97	12.6	1.5	.94
			1200		1200					
II		Pupils' collective movement activity/passivity and social access								
	Activity	1. Inter-pupil contacts and movement, space, time, energy restricted; range of ideas controlled	59.7	95.4	213.8	247.7	2.01	136.8	199.8	4.05
		2. Inter-pupil contacts and/or movement free; range of ideas controlled	631.2	285.1	344.8	210.4	-2.80^{x}	488.0	285.4	7.84^{x}
		3. Inter-pupil contacts free; range of ideas open	95.4	232.8	98.7	156.2	.04	97.0	193.9	.16
		4. Pupils' spontaneous activity	10.0	25.7	1.5	4.9	-1.13	5.8	16.6	1.27
	Passivity	5. Pupils follow instruction, demonstration	253.3	135.1	399.3	76.1	3.26^{xx}	326.3	130.6	10.64^{xxx}
		6. Pupils organize themselves, assist in organization	131.9	59.4	119.3	69.3	-.48	125.6	63.4	.23
		7. Pupils wait for turn	5.4	5.9	10.0	10.9	1.28	7.7	3.9	1.63
	Other	8. Confused situation	13.1	2.1	12.5	0.8	-.90	12.8	1.6	.50
			1200		1200					
III		Social form								
	Situation	1. Complete class, uniform task	374.4	291.5	380.9	383.3	.05	377.7	333.0	.22
		2. Divided class, uniform task	270.0	335.1	402.1	432.8	.78	336.0	412.0	.61
		3. Divided class, differentiated tasks	241.8	360.5	301.2	327.3	.42	271.5	338.1	.15
		4. Divided class, differentiated tasks distributed amongst groups & within group	179.3	234.1	36.2	75.3	-2.02	107.3	185.1	4.07
		5. Individual work, uniform task	115.8	206.5	61.5	142.1	-.75	88.7	175.6	.56
		6. Individual work, differentiated tasks	0.1	0.3	6.0	20.8	.99	3.0	14.7	.97
		7. Other situation, confused situation	18.5	22.2	12.2	0.6	-.99	15.3	15.7	.98
			1200		1200					

6 observers
24 lessons
4800 6 second time units, tot. 28800 time units

x = $p \leq 0.05$
xx = $p \leq 0.01$
xxx = $p \leq 0.001$

TABLE 8.

Significance of Differences between Means Estimated for the Lessons of Three Grade Levels (T_2); t-, F-test.

Cluster	Categories	Grade-levels: 1. Low level N = 8 $\bar{x}$	s	2. Middle level N = 8 $\bar{x}$	s	3. Upper level N = 8 $\bar{x}$	s	t-test: Dif 1-2 df=14 t	Dif. 1-3 df=14 t	Dif 2-3 df=14 t	tot. N = 24 $\bar{x}$	s	F-test df= 2 df= 21 F
I	Teachers' talk, movement; pupils' talk; other												
Teacher	01. Accepts, praises, encourages	35.5	17.7	36.9	21.2	39.0	30.1	.24	.36	.16	36.8	22.6	.76
	02. Gives corrective feedback, urges	59.9	30.9	82.3	53.3	59.6	47.9	1.03	-.01	-.89	67.3	44.4	.66
	03. Uses, developes ideas, movement, tasks suggested by pupils	6.9	4.7	2.6	1.6	1.8	2.8	-2.41 x	-2.64 xx	-.78	3.8	3.9	5.93 x
	04. Asks, initiates and terminates activity	91.9	57.3	80.1	66.6	70.3	49.7	-.38	-.81	-.34	80.8	56.4	.28
	05. Presents information, organizes	525.6	109.1	456.3	89.6	444.7	115.6	-1.39	-1.44	-.22	475.6	107.1	1.38
	06. Gives directions, commands during activity	56.6	71.2	47.4	48.2	34.4	41.4	-.30	-.76 xx	-.58	46.2	53.4	.33
	07. Criticizes pupils behaviour	15.1	13.0	12.4	13.8	0.5	1.1	-.41 x	-3.16 xx	-2.43 x	9.4	12.3	4.01 x
Pupil	08. Answers question/clarifies, demonstrates	15.6	12.6	3.5	3.0	2.0	1.8	-2.64 x	-3.03 xx	-1.21	7.0	9.5	7.81 xxx
	09. Pupil speaks spontaneously, initiates	35.5	18.7	16.1	9.2	8.2	8.4	-2.63	-3.76	-1.78	20.0	17.0	9.35 xxx
Teacher	10. Teacher follows pupils' activity, silent guidance	307.2	142.0	396.7	91.4	408.5	208.6	1.50	1.14	.15	370.8	155.1	1.03
	11. Silent participation in movement activity	36.9	31.6	53.5	90.6	118.9	145.6	.49	1.56	1.08	69.8	102.8	1.48
Other	12. Confused situation	13.3	2.4	12.2	0.5	12.1	0.4	-1.14	-1.29	-.61	12.5	1.5	1.46
		1200		1200		1200					1200		
II	Pupils' collective movement activity/passivity and social access												
Activity	1. Inter-pupil contacts and movement, space, time, energy restricted range of ideas controlled	147.9	182.0	161.1	246.8	101.3	186.6	.12	-.51	-.55	136.8	199.8	.18
	2. Inter-pupil contacts and/or movement free; range of ideas controlled	421.2	244.9	518.4	283.7	524.4	345.7	.73	.69	.04	488.0	285.4	.31
	3. Inter-pupil contacts free; range of ideas open	29.8	31.7	110.4	172.3	151.0	287.2	1.27	1.18	.34	97.0	193.9	.80
	4. Pupils' spontaneous activity	13.0	31.6	4.1	6.2	0.1	0.4	-.78	-1.15	-1.82	5.8	18.6	1.00
Passivity	5. Pupils follow instruction, demonstration	388.1	115.2	274.5	98.4	316.4	159.9	-2.12	-1.03 xx	.63	326.3	130.6	1.63 xxx
	6. Pupils organize themselves, assist in organization	179.1	71.8	111.1	34.6	86.6	40.3	-2.41 x	-3.18	-1.30	125.6	63.4	6.91
	7. Pupils wait for turn	7.5	5.0	7.6	8.6	8.0	12.7	.04	-.10	.07	7.7	8.9	.62
Other	8. Confused situation	13.4	2.4	12.8	1.2	12.2	0.7	-.67	-1.28	-1.04	12.8	1.6	1.01
		1200		1200		1200					1200		
III	Social form												
Situation	1. Complete class, uniform task	315.1	255.1	485.0	431.5	332.9	305.2	.96	.13	-.81	377.7	333.0	.61
	2. Divided class, uniform task	447.2	420.2	286.8	420.9	274.1	427.4	-.76	-.82	-.06	336.0	412.0	.42
	3. Divided class, differentiated tasks	296.4	319.1	213.2	344.9	304.9	385.8	-.50	.05	.50	271.5	338.1	.17
	4. Divided class, differentiated tasks distributed amongst groups & within group	47.8	94.5	92.5	197.5	183.0	231.9	.58	1.53	.84	107.8	185.1	1.12
	5. Individual work, uniform task	62.9	165.4	110.0	204.1	93.1	176.0	.51	.35	-.18	88.7	175.6	.14
	6. Individual work, differentiated tasks	9.0	25.5	0.1	0.4	0.0	0.0	-.99	-1.00	-1.00	3.0	14.7	.99
	7. Other situation, confused situation	21.6	27.2	12.4	0.7	12.0	.0	-.96	-1.00	-1.43	15.2	15.7	.96
		1200		1200		1200					1200		

6 observers
24 lessons (20 minutes) N=144
4800 time units, tot. 28800 time units.

x = p ≥ 0.05
xx = p ≤ 0.01
xxx = p ≤ 0.001

TABLE 9.

Significance of Differences between Means Estimated for the Lessons of Four Subject Areas of P.E. (T_2); t-, F-test.

Cluster	Categories	1. Gymnastics N = 6 x̄	s	2. Apparatus N = 6 x̄	s	3. Rhytmic movement N = 6 x̄	s	4. Ball games N = 6 x̄	s	t-test dif. 1-2 df=10 t	dif. 1-3 df=10 t	dif. 1-4 df=10 t	dif. 2-3 df=10 t	dif. 2-4 df=10 t	dif. 3-4 df=10 t	tot. N = 24 x̄	s	F df= 3 df=20 F
:	Teachers' talk, movement; pupils' talk; other																	
Teacher	01. Accepts, praises, encourages	47.8	21.9	53.8	16.6	21.0	14.4	24.5	20.3	.53	-2.50^{x}	-1.19	-3.66^{xx}	-2.74^{x}	.34	36.8	22.6	4.72^{x}
	02. Gives corrective feedback, urges	55.0	16.3	127.7	42.2	39.3	23.9	47.0	21.7	3.93^{xx}	-1.32	-.72	-4.46^{xx}	-4.16^{xx}	.58	67.3	44.4	12.92^{xxx}
	03. Uses, developes ideas, movement, tasks suggested by pupils	3.7	3.8	3.2	3.1	6.2	5.7	2.0	1.1	-.25	.89	-1.02	1.12	-.86	-1.75	3.8	3.9	1.26
	04. Asks, initiates and terminates activity	140.9	35.4	35.5	14.5	81.7	65.5	65.0	42.3	-6.74^{xxx}	-1.95	-3.37^{xx}	1.68	1.62	-.52	80.8	56.4	6.26^{xxx}
	05. Presents information, organizes	488.8	90.6	546.1	75.6	420.3	102.6	446.9	132.1	1.19	-1.21	-.64	-2.12^{x}	-1.60	.39	475.5	107.1	1.73
	06. Gives directions, commands during activity	105.0	81.8	30.5	12.8	30.0	22.4	19.0	11.5	-2.20	-2.17	-2.55^{x}	-0.5	-1.63	-1.07	46.2	53.4	5.03^{xxx}
	07. Criticizes pupils behaviour	12.2	17.2	8.7	10.3	12.0	15.0	4.5	4.8	-.43	-.02	-1.05	.45	-.90	-1.16	9.3	12.3	.48
Pupil	08. Answers question/clarifies, demonstrates	11.8	16.1	3.8	2.5	6.5	9.9	6.0	3.7	-1.20	-.69	-.86	.64	1.18	-.12	7.1	9.5	.73
	09. Pupil speaks spontaneously, initiates	20.8	21.7	28.5	28.4	15.3	12.1	15.2	15.5	.66	-.54	-.52	-1.46	-1.36	-.02	20.0	17.0	.79
Teacher	10. Teacher follows pupils' activity, silent guidance	235.5	79.7	328.2	64.3	384.3	105.0	535.2	179.8	2.22^{x}	2.77^{x}	3.73^{xx}	1.12	2.65^{x}	1.77	370.8	155.1	7.03^{xxx}
	11. Silent participation in movement activity	66.0	42.9	21.8	29.5	171.2	163.8	20.0	31.0	-2.08	1.52	-2.13	2.20	-.10	-2.22^{x}	69.8	102.8	3.95^{x}
Other	12. Confused situation	12.5	0.8	12.2	0.4	12.2	0.4	13.4	2.8	-.88	-.88	.70	.00	1.01	1.01	12.6	1.5	.82
		1200		1200		1200		1200								1200		
II	Pupils' collective movement activity/passivity and social access																	
Activity	1. Inter-pupil contacts and movement, space, time, energy restricted; range of ideas controlled	311.5	283.5	6.2	11.1	43.0	85.2	186.3	158.7	-2.64^{x}	-2.22^{x}	-.94	1.05	2.77^{xx}	1.95	136.8	199.8	4.17^{x}
	2. Inter-pupil contacts and/or movement free; range of ideas controlled	368.9	299.8	648.0	201.8	388.8	287.3	546.3	312.8	1.89	.12	1.00	-1.81	-.67	.91	488.0	285.4	1.36
	3. Inter-pupil contacts free; range of ideas open	17.8	38.5	19.9	48.6	347.9	261.2	2.7	5.1	.08	3.06^{x}	-.96	3.02^{x}	-.86	-3.24^{xx}	97.0	193.9	9.32^{xxx}
	4. Pupils' spontaneous activity	1.3	2.4	15.3	37.1	5.8	6.8	0.5	0.8	.92	1.53	-.80	-.62	-.98	-1.91	5.8	18.6	.78
Passivity	5. Pupils follow instruction, demonstration	378.0	100.4	347.3	159.0	283.2	115.8	296.8	150.7	-.40	-1.52	-1.10	-.80	-.56	.18	326.3	130.6	.65
	6. Pupils organize themselves, assist in organization	103.2	49.4	148.0	89.5	102.2	54.5	149.2	50.6	1.07	-.03	1.59	-1.07	.03	1.55	125.6	63.4	1.06
	7. Pupils wait for turn	6.0	5.1	2.5	2.8	16.8	12.4	5.5	5.9	-1.47	1.98	-.16	2.76^{x}	1.13	-2.02	7.7	8.9	4.24^{x}
Other	8. Confused situation	13.3	2.8	12.8	1.0	12.3	0.8	12.7	1.2	-.41	-.84	-.53	-.96	-.26	.56	12.8	1.6	.38
		1200		1200		1200		1200								1200		
II. Situation	Social form																	
	1. Complete class, uniform task	699.8	444.0	327.5	125.7	407.7	207.3	75.7	127.0	-1.98	-1.46	-3.31^{xx}	.81	-3.45^{xx}	-3.34^{xx}	377.7	333.0	5.83^{xxx}
	2. Divided class, uniform task	315.0	345.0	42.4	92.0	179.5	206.6	807.3	459.0	-1.87	-.83	2.10	1.43	4.00^{xx}	3.06^{x}	336.0	412.0	7.00^{xxx}
	3. Divided class, differentiated tasks	52.2	127.8	760.9	164.1	174.5	244.7	98.5	183.0	8.67^{xxx}	1.09	.51	-4.97^{xxx}	-6.78^{xxx}	-.61	271.5	338.1	19.59^{xxx}
	4. Divided class, differentiated tasks distributed amongst groups & within group	120.8	134.2	41.8	75.3	64.2	157.2	204.2	300.6	-1.26	-.67	.62	.31	1.28	1.01	107.8	185.1	.90
	5. Individual work, uniform task	0.0	0.0	2.7	6.5	349.7	180.9	2.3	5.7	1.00	4.73^{xxx}	1.00	4.73^{xxx}	-.09	-4.70^{xxx}	88.7	175.6	22.15^{xxx}
	6. Individual work, differentiated tasks	0.0	0.0	0.0	0.0	12.1	29.3	0.0	0.0	0.0	1.02	.00	-1.02	0.0	-1.02	3.0	14.7	1.03
	7. Other situation, confused situation	12.2	0.4	24.8	31.4	12.3	0.9	12.0	0.0	.99	.45	-1.00	-.97	-1.00	-1.00	15.3	15.7	.97
		1200		1200		1200		1200								1200		

6 observers

24 lessons (20 minutes) N=114

4164 6 sec. time units. tot. 28800 time units.

x = p = 0.05

xx = p = 0.01

xxx = p = 0.001

clearly indicated that the interaction on the pupils' part was mostly "collective movement activity" (60% of the class time), or preparation for it by "following instruction" (II/5) or "organizing themselves" (II/6) (30% of the class time). Pupils' movement activities were response behavior, also characterized by teacher initiation as analysed by the social access categories. This was emphazised in movement activities where "inter-pupil contacts and/or movements are free but the range of ideas is controlled" (II/2) (40% of class time). The use of the pupils' own ideas in movement activity was strongly related to certain frame variables, such as the P.E. subject area. (Table 9.)

The variability of the social form, division of labor and responsibility, from class to class was typical (Table 6 and Figure 12). The predominant social form (39% of the class time) was "complete class, uniform task" (III/1), which was used, e.g., in situations where pupils are following instruction. However, the use of other social forms (e.g., divided class) was also very common, with a uniform task (28% of the class time) as well as with different tasks (22% of the class time). Individual work, especially with differentiated tasks, was used rarely. The distribution of the social forms was strongly related to the content of instruction, i.e., the P.E. subject area.

In describing the instructional process using the categories of PEIAC/LH-75, twenty-two statistically significant differences as a function of frame factors were found in the 27 categories: four between the two teachers observed, five between grade levels (related to pupil behavior), and thirteen between the different P.E. subject areas.

Of the four categories describing differences between the two teachers, two were in the area of "teacher's verbal/nonverbal behavior," and two in the area of "pupil collective movement activity/passivity." These variables appear to be related to teacher education, which is somewhat different for women than for men. They reflected the characteristics of teacher initiation behavior (i.e., command technique). The instructional process was very sensitive to different frame factors, such as pupil behavior. These differences were reflected both in teacher response and in initiation behavior, and most clearly in categories describing pupils' initiation and response behavior.

The subject area differences were statistically significant in half of the 27 categories. In most categories describing "division of labor and responsibility" and in half of the categories describing "verbal behavior," differences were statistically significant (Table 9). Also in three categories describing "pupils' collective activity/passivity," one finds statistically significant differences between mean scores of instructional process with different content.

These are structural characteristics of the instructional process described with the three aspects of PEIAC/LH-75. Mostly they describe general features. The results are not very reliable, however, because some of the variables were used infrequently and the number of scores was low. In the next step, an attempt was made to analyze the sequential tendencies of the instructional process.

Matrix Analysis of Sequence Patterns in the Instructional Process

The matrices of the three clusters computed from the same data are presented as absolute frequencies and percents in Table 10, and as millage matrices in Table 11. The millage matrices describing the interaction process from the perspective of two teachers, three grade levels and four P.E. subject areas are presented in Tables 12, 13 and 14 respectively.

In the intepretation of the results, a flow-card description (see Flanders, 1970, pp. 115-120) was drawn of the matrices and the cell frequencies were used to support theoretical speculations. In this context, <u>instructional process</u> means the transition of the system from one state to another as a function of time. <u>Transitions</u> are sequence pairs with different numbers, <u>steady states</u> are sequence pairs with the same number. The concept <u>variety</u> refers to the total number of different configurations which occur in a gymnasium. The concept <u>sequence</u> refers to how many different configurations occurred in a given time period. Decoding a matrix attempts to recreate those aspects of the original instruction which were encoded by building a description of process.

In a flow diagram, knowledge of the clockwise rotation of events and the differences between columns and rows are essential. The steps used in analyzing the three cluster matrices are as follows:

1. Search for the highest cell frequency as the starting point, and ...
2. ...locate the event which is most likely to flow (is located) by inspecting the row which is designated by the second number in the address of the starting cell.
3. Look in the row designated by the number in the address of the cell just marked.
4. Search for the next most frequent event what will be found, as before, in the row designated by the second number in the address of the present cell.

The flow diagram can be used to help clarify the sequence and to make the matrix display more understandable. Each cell of the interaction matrices and millage matrices indicates how many times in general the system has shifted from the state represented in the row to a state represented in the column in question.

These transition frequencies were denoted by decoding the matrices in terms of patterns. Of particular interest was the number of different configuration pairs which occurred in general in the 24 P.E. lessons, and the total number, or variety, of different configurations in the matrices of the three clusters.

There was a great variability between the clusters of transition cells and steady state cells. On the average, 50% of all sequence pairs in the diagonal in the first cluster were in the steady state cells, more than 80% in the second cluster, and 90% in the third cluster. Thus, the tempo of transition was quite different for these different aspects. The critical decisions made by the teacher are thus strongly related to the time factor. In the first cluster, the "teaching" (5-5) and "silent guidance" (10-10) categories contain the highest percent of scores, more than half of which are in the steady state cells. The transitions in the other categories are not so strongly centralized to these cells (see Tables 10and 11). In the second cluster, the most dominant steady state cells are "activity 2" (2-2) and "following instruction" (5-5), with more than 90% of the transitions in these categories found in these cells. Also in the third cluster, more than 95% of transitions are in the steady state cells.

In these situational settings, the critical teaching behavior is analyzed by observing critical transitions, i.e., sequence pairs with

Table 10. Matrices for Episodes by Category: Videorecorded Material (T_2)

Categories

Teachers talk, movement; pupils talk; other

Teacher	01.	Accepts, praises, encourages
	02.	Gives corrective feedback, urges
	03.	Uses, developes ideas, movement, tasks, suggested by pupils
	04.	Asks, initiates and determinates activity
	05.	Precents information, organizes
	06.	Gives directions, commands during activity
	07.	Criticizes pupils' behaviour
Pupil	08.	Answers question/clarifies, demonstrates
	09.	Pupil speaks spontaneously, initiates
Teacher	10.	Teacher follows pupils' activity, silent guidance
	11.	Silent participation in movement activity
Other	12.	Confused situation

CAT	1	2	3	4	5	6	7	8	9	10	11	12	TOT	%
1	45	88	7	100	327	23	8	1	25	238	18	2	867	3.08
2	64	275	6	107	373	47	4	0	27	401	69	11	1414	5.03
3	6	4	2	6	36	1	1	1	0	29	0	4	90	0.31
4	52	40	12	89	729	148	14	147	30	439	74	14	1938	6.76
5	243	363	13	875	7613	303	83	13	237	1342	157	71	11411	39.03
6	34	47	1	97	331	390	25	2	19	137	21	3	1107	3.88
7	2	3	2	20	91	19	26	0	13	44	2	0	224	0.78
8	16	10	10	27	75	2	5	2	6	11	2	1	164	0.59
9	34	59	31	23	225	7	8	0	18	60	7	2	474	1.67
10	323	698	6	525	1209	142	41	3	93	5790	45	33	8908	31.04
11	30	76	0	64	154	23	2	0	11	55	1251	8	1674	5.84
12	4	1	0	3	50	2	0	0	0	81	20	9	158	0.55
TOT	882	1614	90	1938	11413	1107	224	169	474	8908	1674	156	28658	100.00

Pupils' collective movement activity/passivity and social access

Activity	1.	Inter-pupils contacts and movement, space, time, energy restricted; range of ideas controlled
	2.	Inter-pupils contacts and/or movement free; r. of id. contr.
	3.	Inter-pupils contacts free; range of ideas open
	4.	Pupils' spontaneous activity
Passivity	5.	Pupils follow instruction, demonstrations
	6.	Pupils organize themselves, assist in organization
	7.	Pupils wait for turn
Other	8.	Confused situation

CAT	1	2	3	4	5	6	7	8	TOT	%
1	2794	30	1	0	395	49	8	5	3282	11.45
2	37	10692	11	7	621	255	25	64	11712	40.87
3	2	13	2183	2	75	39	1	14	2329	8.13
4	0	4	4	111	7	11	1	0	138	0.48
5	304	673	82	6	6125	470	56	56	7832	27.33
6	37	193	22	11	542	2180	18	12	3015	10.52
7	30	33	7	0	32	5	76	2	185	0.65
8	18	73	19	1	35	6	0	11	163	0.57
TOT	3282	11712	2329	138	7832	3015	185	163	28658	100.00

Social form

Situation	1.	Complete class, uniform task
	2.	Divided class, uniform task
	3.	Divided class, differentiated tasks
	4.	Divided class, differentiated tasks distributed amongst groups & within group
	5.	Individual work, uniform task
	6.	Individual work, differentiated tasks
	7.	Other situation, confused situation

CAT	1	2	3	4	5	6	7	TOT	%
1	8801	74	35	16	34	6	38	9064	31.63
2	35	7926	40	17	4	0	38	8065	28.14
3	24	23	6425	1	1	0	42	6516	22.74
4	32	10	2	2520	5	0	17	2586	9.02
5	31	9	2	3	2071	2	10	2128	7.43
6	1	0	0	0	2	65	5	73	0.25
7	79	23	12	29	6	0	75	224	0.78
TOT	9064	8065	6516	2586	2128	73	224	28658	100.00

Table 11. Millage Matrices for Episodes by Category with Transition Cells and Steady State Cells: Videorecorded Material (T_2)

Millage matrix - Cluster I

Categories

Teachers talk, movement; pupil talk; other
1. Accepts, praises, encourages
2. Gives corrective feedback, urges
3. Uses, develops ideas, movement, tasks, suggested by pupils
4. Asks, initiates and terminates activity
5. Presents information, organizes
6. Gives directions, commands during activity
7. Criticizes pupils behaviour
8. Answers question/clarifies, demonstrates
9. Pupil speaks spontaneously, initiates
10. Teacher follows pupils' activity, silent guidance
11. Silent participation in movement activity
12. Confused situation

cat	1	2	3	4	5	6	7	8	9	10	11	12
1	1	3	0	3	11	0	0	0	0	8	0	0
2	3	9	0	3	13	1	0	0	0	20	2	0
3	0	0	0	0	1	0	0	0	0	1	0	0
4	1	3	0	3	25	5	0	5	1	18	2	0
5	8	9	0	30	272	10	2	0	8	46	5	2
6	1	1	0	3	11	13	0	0	0	4	0	0
7	0	0	0	0	3	0	0	0	0	1	0	0
8	0	0	0	1	2	0	0	0	0	0	0	0
9	1	2	1	0	7	0	0	0	0	2	0	0
10	11	24	0	18	42	4	1	0	3	202	1	1
11	1	2	0	2	5	0	0	0	0	1	43	0
12	0	0	0	0	1	0	0	0	0	2	0	0
TOT	30	56	3	67	398	38	7	5	16	310	58	5

N=28656

Millage matrix - Cluster II

Pupils' collective movement activity/passivity and social access

1. Inter-pupil contacts and movement, space, time, energy restricted; range of ideas controlled
2. Inter-pupil contacts and/or movement free; r. of id. contr.
3. Inter-pupil contacts free: range of ideas open
4. Pupils' spontaneous activity
5. Pupils follow instruction, demonstrations
6. Pupils organize themselves, assist in organization
7. Pupils wait for turn
8. Confused situation

cat	1	2	3	4	5	6	7	8
1	97	1	0	0	13	1	0	0
2	1	373	0	0	21	8	0	2
3	0	0	76	0	2	1	0	0
4	0	0	0	3	0	0	0	0
5	12	23	2	0	213	16	1	1
6	1	6	0	0	18	76	0	0
7	1	1	0	0	1	0	2	0
8	0	2	0	0	1	0	0	0
TOT:	114	408	81	4	273	105	6	5

N=26656

Millage-matrix - Cluster III

Social form

1. Complete class, uniform task
2. Divided class, uniform task
3. Divided class, differentiated tasks
4. Divided class, differentiated tasks distributed amongst groups & within group
5. Individual work, uniform task
6. Individual work, differentiated tasks
7. Other situation, confused situation

cat	1	2	3	4	5	6	7
1	309	2	1	0	1	0	1
2	1	276	1	0	0	0	1
3	0	0	224	0	0	0	1
4	1	0	0	87	0	0	0
5	1	0	0	0	72	0	0
6	0	0	0	0	0	2	0
7	2	0	0	1	0	0	2
TOT:	310	281	227	90	74	2	7

N=26656

different numbers. It is probable that, in Cluster I, the most important decisions of the teacher occur in certain rows (nine and ten) and columns (one through seven). The tallies in these cells represent the first verbal reaction of the teacher at the moment when a student stops talking or moving. In Cluster II, the tallies in the cells formed by the intersection of rows three and four and columns five to eight represent the first collective passive behavior after pupils' collective activity in which pupils were initiative. In the third cluster, all tallies in the cells formed by rows three, four and five and column one represent the reaction of the teacher to direct the complete class and to make decisions connected to the next transition concerning division of labor and responsibility.

In the first cluster, one distinguishes four different patterns representing the teacher's verbal/nonverbal critical behavior. The most dominant pattern is the "silent guidance, a long teaching" pattern (10-10, 10-5, 5-5, 5-10). The second pattern is "silent guidance" and "stopping activity, teaching - starting activity, a short drill" (10-10, 10-4, 4-5, 5-5). In the third pattern, "command, teaching during activity" (6-6, 6-5, 5-5, 5-6) is found. The fourth critical sequence pattern is "silent guidance, corrective feedback, silent guidance" (10-10, 10-2, 2-2, 2-10). In general, teacher verbal initiation was a dominating characteristic, but one could also recognize the use of patterns describing teacher response behavior.

In the second cluster matrix, the variety of different configurations describing pupils' participation was not as great. In the clockwise flow, we can distinguish the most dominant pattern, a long "pupils' movement activity" period with "inter-pupil contacts and/or movement free, range of ideas in movement activity controlled, instruction following, pattern (2-2, 2-5, 5-5, 2-5). In the second orbit, a "pupils' movement activity with total control, instruction following, organizing" pattern (1-1, 1-5, 5-6, 6-6) is found. The third critical sequence pattern is "pupils' collective activity with inter-pupil contacts free and range of ideas open, instruction following" (3-3, 3-5, 5-5, 5-3), and "pupils' spontaneous activity, pupils organize themselves, pupils follow instruction" (4-4, 4-6, 6-6, 6-5).

In the matrix of the third cluster, describing the flow as different social forms used in classes, observations are centralized in the steady state cells (90%), and the variety of different configurations is low compared with the other clusters. The most dominant sequence pattern is the use of "complete class with uniform task, divided class with uniform task" (1-1, 1-2, 2-2, 2-1). The first critical sequence pattern is "differentiated tasks, complete class, uniform task" (3-3, 3-1, 1-1, 1-3); the second, 4-4, 4-1, 1-1, 1-4; the third, 5-5, 5-1, 1-1, 1-5; and the fourth, 5-5, 5-2, 2-2, 2-5. Thus the sequence patterns describe mostly teaching for all, then division of labor and responsibility in different forms. In describing the flow of critical sequence patterns, such as in the cells formed by rows 3-4 and columns 1 and 2, "divided class, differentiated tasks" are distributed amongst groups and within groups. In row 5, "individual work, uniform tasks," cell 5-5, the sequence, the number of different configuration pairs and the variabity seem to be higher than with other, more direct social forms. The situation is thus more variable and nondirective. However, in general, the critical teaching behavior described by the cell frequencies was characterised by directness in this sample.

<u>The Comparison of Sequence and Variety Across Class Time as a Function of the Teacher</u>

The millage matrices by clusters computed from the scores of 12 lessons for each of two teachers rated by six observers (T_2), each containing 14,328 six-second time units are presented in Table 11. Rows have been singled out representing categories in which significant differences in marginal frequencies between teachers were formed. The arrows are intended to help clarify differences in mean sequences in the three cluster matrices.

The dominant critical sequence pattern in the first cluster matrix for the male teacher is "silent guidance, present information, silent guidance" (10-10, 10-5, 5-5, 5-10), whereas for the female teacher it is a "silent guidance, terminates activity, present information, initiation of activity" pattern (10-4, 4-5, 5-5, 5-4). The second different critical pattern for the woman teacher is "teacher gives direction, commends during activity, gives information, follows pupils' activity, silent" (6-6, 6-5, 5-4, 4-10), and for the man "pupils' verbal intiation, teaching" (9-9, 9-5, 5-5, 5-9)and "silent participation, teaching" (11-11, 11-5, 5-5, 5-11) patterns. The variety of tran-

Table 12. Millage Matrices for Episodes by Two Teachers

Categories

Teachers talk, movement; pupils talk; other

Teacher	01.	Accepts, praises, encourages
	02.	Gives corrective feedback, urges
	03.	Uses, develops ideas, movement, tasks, suggested by pupils
	04.	Asks, initiates and determinates activity
	05.	Presents information, organizes
	06.	Gives directions, commands during activity
	07.	Criticizes pupils behaviour
Pupil	08.	Answers question/clarifies, demonstrates
	09.	Pupil speaks spontaneously, initiates
Teacher	10.	Teacher follows pupils' activity, silent guidance
	11.	Silent participation in movement activity
Other	12.	Confused situation

Pupils' collective movement activity/passivity and social access

Activity	1.	Inter-pupils contacts and movement, space, time, energy restrocted: range of ideas controlled
	2.	Inter-pupils contacts and/or movement free: r. of id. contr.
	3.	Inter-pupils contacts free: range of ideas open
	4.	Pupils' spontaneous activity
Passivity	5.	Pupils follow instruction, demonstrations
	6.	Pupils organize themselves, assist in organization
	7.	Pupils wait for turn
Other	8.	Confused situation

Social form

Situation	1.	Complete class, uniform task
	2.	Divided class, uniform task
	3.	Divided class, differentiated tasks
	4.	Divided class, differentiated tasks distributed amongst groups & within group
	5.	Individual work, uniform task
	6.	Individual work, differentiated tasks
	7.	Other situation, confused situation

Teacher 1 (Men, N=12)

Millage-matrix-cluster I

CAT	1	2	3	4	5	6	7	8	9	10	11	12
1	1	4	0	4	11	0	0	0	1	12	1	0
2	3	8	0	2	13	1	0	0	1	21	3	0
3	0	0	0	0	1	0	0	0	0	1	0	0
4	1	2	0	2	19	1	0	4	0	9	2	0
5	10	9	0	22	265	7	4	0	11	63	8	2
6	0	0	0	1	8	3	1	0	0	3	0	0
7	0	0	0	0	4	0	1	0	6	2	0	0
8	0	0	0	0	2	0	0	0	0	0	0	0
9	1	2	1	0	10	0	0	0	0	2	0	0
10	13	25	0	8	58	3	2	0	4	206	1	0
11	1	3	0	2	7	0	0	0	3	2	38	0
12	0	0	0	0	2	0	0	0	0	0	1	0
TOT	37	58	3	45	405	20	11	5	22	327	58	5

N=14328

Millage-matrix-Cluster II

CAT	1	2	3	4	5	6	7	8
1	44	0	0	0	4	0	0	0
2	8	498	8	0	21	10	1	3
3	0	0	76	0	7	1	0	0
4	0	0	0	8	0	0	0	0
5	3	24	2	0	163	14	1	1
6	0	7	1	0	17	81	0	0
7	0	1	0	0	0	0	1	0
8	0	2	0	0	1	0	0	0
TOT	49	528	79	8	212	110	4	5

N=14328

Millage-matrix-Cluster III

CAT	1	2	3	4	5	6	7
1	305	2	1	0	1	0	2
2	1	222	0	0	0	0	1
3	1	0	199	0	0	0	0
4	1	0	0	147	0	0	0
5	1	0	0	0	94	0	0
6	0	0	0	0	0	0	0
7	2	0	0	1	0	0	3
TOT	313	226	202	150	97	0	10

N=14328

Teacher 2 (Women, N=12)

Millage-matrix-Cluster I

CAT	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	0	2	11	0	0	0	0	4	0	0
2	2	9	0	3	12	2	0	0	0	20	1	0
3	0	0	0	0	1	0	0	0	0	0	0	0
4	2	4	0	4	31	8	0	5	1	28	2	0
5	4	9	0	38	270	13	1	0	5	30	2	2
6	1	2	0	5	14	23	0	0	0	8	1	0
7	0	0	0	0	1	0	0	0	0	0	0	0
8	0	0	0	1	2	0	0	0	0	0	0	0
9	0	1	0	0	4	0	0	0	0	1	0	0
10	7	23	0	28	25	4	0	0	2	197	1	1
11	0	1	0	1	3	1	0	0	0	0	48	0
12	0	0	0	0	1	0	0	0	0	3	0	0
TOT	23	54	3	89	391	56	4	4	11	294	58	5

N=14328

Millage-matrix-Cluster II

CAT	1	2	3	4	5	6	7	8
1	130	1	0	0	23	2	0	0
2	1	255	0	0	22	7	0	1
3	0	0	77	0	2	1	0	0
4	0	0	0	0	0	0	0	0
5	21	22	2	0	263	18	2	2
6	2	5	0	0	19	70	0	0
7	1	1	0	0	1	0	3	0
8	1	2	0	0	0	0	0	0
TOT	170	288	82	1	334	99	8	5

N=14328

Millage-matrix-Cluster III

CAT	1	2	3	4	5	6	7
1	313	2	1	0	0	0	0
2	1	331	2	0	0	0	1
3	0	1	248	0	0	0	2
4	0	0	0	28	0	0	0
5	0	0	0	0	50	0	0
6	0	0	0	0	0	4	0
7	3	1	0	0	0	0	0
TOT	319	336	252	30	51	5	5

N=14328

sitions and configurations was greater for the woman teacher as described in the sequence matrix of the first cluster.

In the second cluster, the dominating critical sequence pattern for the man teacher was "pupils' collective movement activity where inter-pupil contacts and/or movement free range of ideas controlled, pupils follow instruction" (2-2, 2-5, 5-5, 5-2). For the woman teacher the pattern was "pupils moving collectively, inter-pupils follow instruction, range of ideas controlled" (1-1, 1-5, 5-5, 5-1, 1-1, 1-5). The variety of configurations for the woman teacher was greater than for the man teacher.

In the third cluster, the most dominant sequence pattern for both teachers was "complete class, uniform task, divided class, uniform task" (1-1, 1-2, 2-2, 2-1). For the man teacher, the critical sequence pattern 4-4, 4-1, 1-1, 1-2 was common as was the pattern 2-2, 2-5, 5-5, 5-2. In general, the variety of social forms configurations and non-directiveness reflected through division of labor and responsibility were higher for the man teacher.

In general, the behavior of the two teachers of the sample was quite homogeneous. It was evident that they were rather flexible. The critical sequence pattern varied according to clusters. However, the differences in directiveness were discernible. The behavior of the man teacher was less directive than that of the woman teacher. These differences appear to be related to differences in teacher education.

The Comparison of Sequence and Variety Across Class Time as a Function of Grade Level

The millage matrices computed by clusters from 8 lessons of three grade levels rated by six observers (T_2), each containing 9,552 six-second time units, are presented in Table 13 Some rows representing categories in which significant differences were found between grade levels have been identified. The arrows are intended to help clarify differences in the critical sequence patterns of the three clusters.

In the Cluster I matrix, the lower grade level shows as dominating critical sequence patterns "silence, information, silence" (10-10, 10-5, 5-5, 5-10) and "silence, stop activity, information" (10-10, 10-4, 4-5, 5-5). The more specific critical patterns are "silence, command, silence" (10-10, 10-6, 6-6, 6-10); "pupil initiation, teacher informa-

Table 13. Millage Matrices for Episodes by Grade Level

Categories

Teachers' talk, movement; pupils' talk; other

01. Accepts, praises, encourages
02. Gives corrective feedback, urges
03. Uses, developes ideas, movement, tasks suggested by pupils
04. Asks, initiates and terminates activity
05. Presents information, organizes
06. Gives directions, commands during activity
07. Criticizes pupils behaviour
08. Answers question/clarifies, demonstrates
09. Pupil speaks spontaneously, initiates
10. Teacher follows pupils' activity, silent guidance
11. Silent participation in movement activity
12. Confused situation

Lower level (N=8)

KAT	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	0	3	11	1	0	0	1	3	0	0
2	2	0	0	6	12	1	0	0	1	16	2	0
3	0	0	0	0	2	0	0	0	0	1	0	0
4	1	3	1	4	23	4	1	11	1	19	1	0
5	0	7	0	33	310	17	4	0	16	33	3	2
6	2	2	0	3	16	15	1	0	1	3	0	0
7	0	0	0	1	4	0	1	0	1	2	0	0
8	1	0	0	2	3	0	0	0	0	0	0	0
9	3	3	1	1	13	0	0	0	1	3	0	0
10	4	18	0	18	31	4	1	0	3	166	2	0
11	0	2	0	1	3	0	0	0	0	2	17	0
12	0	0	0	0	1	0	0	0	0	2	1	0
TOT	20	50	5	76	440	67	12	13	29	256	30	0

N= 9552

Middle level (N=8)

KAT	1	2	3	4	5	6	7	8	9	10	11	12
1	0	4	0	2	10	0	0	0	0	10	0	0
2	3	10	0	1	16	1	0	0	1	29	1	0
3	0	0	0	0	0	0	0	0	0	0	0	0
4	3	2	0	2	26	4	0	2	1	23	2	0
5	7	11	0	32	244	18	3	0	4	57	3	2
6	0	1	0	3	11	13	1	0	0	4	1	0
7	0	0	0	0	4	1	1	0	0	2	0	0
8	0	0	0	0	1	0	0	0	0	0	0	0
9	0	1	0	0	6	0	0	0	0	1	0	0
10	12	33	0	16	53	7	2	0	2	196	1	1
11	0	1	0	1	4	1	0	0	0	1	32	0
12	0	0	0	0	1	0	0	0	0	2	0	0
TOT	30	80	2	67	362	39	10	2	13	332	44	3

N= 9552

Upper level (N=8)

KAT	1	2	3	4	5	6	7	8	9	10	11	12
1	1	2	0	4	12	0	0	0	0	[illegible]	1	0
2	3	0	0	2	0	1	0	0	6	19	3	0
3	0	0	0	0	0	0	0	0	0	0	6	0
4	0	2	0	2	20	5	0	1	0	12	4	0
5	0	7	0	25	262	4	0	6	3	48	7	2
6	0	0	0	3	4	12	0	4	0	4	3	0
7	0	0	0	0	0	0	0	0	0	0	6	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	1	0	0	3	0	0	0	0	4	0	0
10	14	20	0	16	40	3	0	0	1	241	0	1
11	2	3	0	3	3	0	0	0	0	2	11	0
12	0	0	0	0	1	0	0	0	0	1	1	0
TOT	32	40	1	50	372	28	6	1	[illegible]	342	44	3

N= 9552

Pupils' collective movement activity/passivity and social access

1. Inter-pupil contacts and movement, space, time, energy restricted; range of ideas controlled
2. Inter-pupil contacts and/or movement free; r. of id. contr.
3. Inter-pupil contacts free; range of ideas open
4. Pupils' spontaneous activity
5. Pupils follow instruction, demonstration
6. Pupils organize themselves, assist in organization
7. Pupils wait for turn
8. Confused situation

Lower level

KAT	1	2	3	4	5	6	7	8
1	100	1	0	0	12	1	0	0
2	1	315	0	0	22	5	0	1
3	0	0	20	0	2	1	0	0
4	0	0	0	9	0	0	0	0
5	10	25	2	0	259	21	1	2
6	1	4	0	0	25	115	1	1
7	1	1	0	0	1	0	2	0
8	0	3	0	0	1	0	0	0
TOT	123	352	24	10	325	150	6	6

N= 9552

Middle level

KAT	1	2	3	4	5	6	7	8
1	113	1	0	0	16	2	0	0
2	1	397	0	0	20	9	1	3
3	0	0	87	0	2	1	0	0
4	0	0	0	2	0	0	0	0
5	16	24	2	0	171	13	0	1
6	2	6	1	0	16	65	0	0
7	0	1	0	0	0	0	3	0
8	0	1	0	0	1	0	0	0
TOT	134	434	92	3	229	93	6	5

N= 9552

Upper level

KAT	1	2	3	4	5	6	7	8
1	70	0	0	0	12	0	0	0
2	0	406	0	0	21	8	0	2
3	0	0	120	0	2	1	0	1
4	0	0	0	0	0	0	0	0
5	10	20	3	0	210	14	3	1
6	0	9	0	0	14	47	0	0
7	1	0	0	0	1	0	2	0
8	0	2	1	0	1	0	0	0
TOT	84	439	126	0	264	72	6	5

N= 9552

Social form

1. Complete class, uniform task
2. Divided class, uniform task
3. Divided class, differentiated tasks
4. Divided class, differentiated tasks distributed amongst groups & within group
5. Individual work, uniform task
6. Individual work, differentiated tasks
7. Other situation, confused situation

Lower level

KAT	1	2	3	4	5	6	7
1	256	3	1	0	1	0	1
2	1	360	2	0	0	0	1
3	0	1	244	0	0	0	1
4	0	0	0	38	0	0	0
5	1	0	0	0	51	0	0
6	0	0	0	0	0	6	0
7	3	0	0	1	0	0	7
TOT	263	374	248	39	52	7	13

N= 9552

Middle level

KAT	1	2	3	4	5	6	7
1	399	2	0	0	1	0	2
2	1	235	0	0	0	0	1
3	1	0	176	0	0	0	0
4	0	0	0	76	0	0	0
5	0	0	0	0	90	0	0
6	0	0	0	0	0	0	0
7	2	1	0	0	0	0	0
TOT	406	240	178	77	92	0	5

N= 9552

Upper level

KAT	1	2	3	4	5	6	7
1	272	2	1	1	0	0	0
2	0	226	0	0	0	0	0
3	0	0	252	0	0	0	1
4	1	0	0	149	0	0	1
5	1	0	0	0	75	0	0
6	0	0	0	0	0	0	0
7	1	0	0	1	0	0	0
TOT	278	224	255	153	77	0	5

N= 9552

6 observers
24 lessons (20 minutes) N = 144
28 800 6 second time units

tion, pupil initiation" (9-9, 9-5, 5-5, 5-9); and "pupil initiation, teacher feedback, teacher information, pupil initiation" (9-9, 9-2, 2-5, 5-9).

In the middle grade level, the critical dominating sequence pattern was "silence, corrective information, silence" (10-10, 10-5, 5-5, 5-10). The more specific patterns were "silence, feedback, silence" (10-10, 10-2, 2-2, 2-10) and "teacher participation, information, teacher participation" (11-11, 11-5, 5-5, 5-11). Thus, there was more silent guidance, feedback, and teacher participation/information than in the lower grade level.

At the upper grade level, the most dominant critical patterns are the same as at middle grade level, i.e., 10-10, 10-5, 5-5, 5-10 and 10-10, 10-2, 2-2, 2-10. The more specific critical sequence patterns are "silence, teacher praises, silence" (10-10, 10-1, 1-1, 1-10); "teacher participation, information, teacher participation" (11-11, 11-5, 5-5, 5-11); and "teacher participation, pupil ideas, teacher feedback" (11-11, 11-3, 3-2, 2-2). Thus, again there was more silent guidance, more teacher participation, short feedback and use of pupils' movement ideas. At the upper grade level, the variety and the total number of different configurations used increased, which indicates a decrease of directiveness in verbal-nonverbal teaching behavior.

In Cluster II, the same characteristics of change were identified in the analysis of pupil collective activity/passivity sequence patterns. The dominant critical sequence pattern in all grade levels was "pupil collective activity in which inter-pupil contacts and/or movement are free, range of ideas are restricted, and pupils follow instruction" (2-2, 2-5, 5-5, 5-2). Typical at the lower grade level were the sequence patterns 2-2, 2-6, 6-6, 6-2 and 6-6, 6-5, 5-5, 5-6, indicating directiveness in activity and in preparations to activity. For the middle grade level, a specific critical sequence pattern was a "totally controlled movement activity, organizing" pattern (1-1, 1-6, 6-6, 6-2), indicating directiveness in different forms. A specific critical pattern at the upper grade level was "pupils collectively moving with free contacts, using open ideas, and following instruction" (3-3, 3-5, 5-5, 5-3).

In Cluster III, the dominating social form pattern was "divided class uniform task, complete class uniform task" (2-2, 2-1, 1-1, 1-2).

Specific critical sequence patterns were formed by grade level: the lower grade level, 2-2, 2-3, 3-3, 3-2; the middle grade level, 5-5, 5-1, 1-1, 1-5; and the upper grade level, 4-4, 4-1, 1-1, 1-4. The sequence and variety of the division of labor and responsibility increased as a function of grade level.

In summary, the sequence and variety increased as a function of grade level and were related to pupil behavior. In addition, the critical sequence patterns in all clusters changed and were characterized by directiveness.

The Comparison of Sequence and Variety Across Class Time as a Function of P.E. Subject Areas

The millage matrices computed by clusters from 6 lessons of four P.E. subject areas, rated by six observers (T_2), each containing 7,164 six-second time units, are presented in Tables 14a, b and d. The rows that are outlined represent the categories in which statistically significant differences were found between the four different subject areas. The category with the greatest difference is marked with a heavy line. The critical sequence patterns and the differences between them are marked with arrows.

These graphic tables are used to illustrate the next step in which the results were analyzed by using the major PEACH/LH-75 parameters compiled from these matrices. With the millage matrices, however, the critical sequence patterns are not discernible because there are only a limited number of time units and the information was computed from repeated measures. Therefore, the indices were also used to reduce and concentrate this information.

Summary

In each of the sequence patterns presented and discussed so far, decisions were required of the teacher for critical transitions, that is, sequence pairs with different numbers. In steady state cells sequence pairs have the same number.

The sequence and variety in the three cluster matrices were different, as expected. In Cluster I, more than one half of all sequence

Table 14a. Millage Matrices for Episodes by Four Subject Areas of Physical Education

A. Gymnastics (N=6 Lessons)

Cluster I: Teacher talk, movement, pupils talk, other

01. Accepts, praises
02. Gives corr. feedback
03. Uses ideas dev. by pup.
04. Asks, init., term. akt.
05. Presents inform., org.
06. Gives dir., comm.
07. Criticizes
08. Answers questions
09. Speaks spontan., init.
10. Silent guidance
11. Silent participation
12. Confused situation

CAT	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	0	6	16	1	0	0	1	6	2	0
2	2	5	0	6	10	2	0	0	0	13	2	0
3	0	0	0	0	1	0	0	0	0	0	0	0
4	3	5	0	6	42	13	0	8	1	27	6	1
5	9	8	0	55	259	18	4	1	7	31	9	2
6	3	3	0	9	18	41	0	0	0	8	1	0
7	0	0	0	1	4	0	1	0	1	0	0	0
8	0	0	0	2	3	0	0	0	0	0	0	0
9	1	1	0	1	8	0	0	0	0	1	0	0
10	12	15	0	21	32	7	0	0	2	102	0	0
11	3	3	0	5	9	1	0	0	0	1	28	[illegible]
12	0	0	0	0	1	0	0	0	0	1	2	0
TOT	40	46	3	117	409	87	10	9	17	197	55	5'

N= 7164

Cluster II: Pupils' collective movement activity/passivity and social access

1. Contacts, ideas cont.
2. Contacts free, ideas cont.
3. Contacts free, ideas open
4. Pupils' spont. activity
5. Pupils follow instruction
6. Pupils organization
7. Pupils wait for turn
8. Confused situation

CAT	1	2	3	4	5	6	7	8
1	212	2	0	0	41	3	0	0
2	3	267	0	0	25	8	0	1
3	0	0	11	0	1	1	0	0
4	0	0	0	0	0	0	0	0
5	37	29	2	0	227	14	1	3
6	3	4	0	0	18	58	0	0
7	2	0	0	0	0	0	1	0
8	1	3	0	0	0	0	0	0
TOT	260	308	14	1	316	86	5	6

N= 7164

Cluster III: Social form

1. Complete class, uniform task
2. Divided class, uniform task
3. Divided class, different task
4. Div. cl. diff. task within gr.
5. Individual work, unif. tasks
6. Individual work, diff. tasks
7. Other, conf. situation

CAT	1	2	3	4	5	6	7
1	579	3	0	0	0	0	1
2	1	259	0	0	0	0	0
3	0	0	42	0	0	0	0
4	1	0	0	97	[illegible]	0	1
5	6	0	0	0	0	0	0
6	0	0	0	0	0	0	0
7	3	0	0	1	0	0	0
TOT	580	263	43	101	0	0	5

N= 7164

Table 14b.(cont.)

B. Apparatus (N=6 Lessons)

Cluster I: Teacher talk, movement, pupils talk, other

	CAT	1	2	3	4	5	6	7	8	9	10	11	12
01. Accepts,praises	1	0	1	0	2	4	0	0	0	0	10	0	0
02. Gives corr.feedback	2	1	2	0	2	9	0	0	0	0	20	1	0
03. Uses ideas dev.by pup.	3	0	0	0	0	0	0	0	0	0	0	0	0
04. Asks,init.,term. akt.	4	1	2	0	1	22	0	0	4	0	19	0	0
05. Presents inform.,org.	5	0	4	0	23	265	8	1	0	8	50	1	2
06. Gives dir..,comm.	6	0	0	0	0	9	1	0	0	0	2	0	0
07. Criticizes	7	0	0	0	0	1	0	0	0	0	1	0	0
08. Answers questions	8	0	0	0	0	2	0	0	0	0	0	0	0
09. Speaks spontan.,init.	9	0	1	1	0	7	0	0	0	0	1	0	0
10. Silent guidance	10	9	23	0	22	44	3	0	0	1	339	1	2
11. Silent participation	11	0	1	0	0	3	0	0	0	0	1	9	0
12. Confused situation	12	0	0	0	0	2	0	0	0	0	1	0	0
	TOT	20	39	1	54	374	15	3	5	12	449	16	6

Cluster II: Pupils' collective movement activity/passivity and social access

	CAT	1	2	3	4	5	6	7	8
1. Contacts, ideas cont.	1	3	0	0	0	1	0	0	0
2. Contacts free,ideas cont.	2	0	505	0	0	16	13	0	3
3. Contacts free, ideas open	3	0	0	15	0	0	0	0	0
4. Pupils' spont. activity	4	0	0	0	11	0	0	0	0
5. Pupils follow instruction	5	0	20	0	0	251	16	0	1
6. Pupils organization	6	0	11	0	0	18	92	0	0
7. Pupils wait for turn	7	0	0	0	0	0	0	0	0
8. Confused situation	8	0	0	0	0	2	0	0	0
	TOT:	5	542	16	12	290	123	2	5

N= 7164

Cluster III: Social form

	CAT	1	2	3	4	5	6	7
1. Complete class, uniform task	1	266	1	1	0	0	0	2
2. Divided class, uniform task	2	0	33	2	0	0	0	0
3. Divided class, different task	3	1	0	631	0	0	0	3
4. Div. cl. diff. task within gr.	4	1	0	0	33	0	0	0
5. Individual work, unif. tasks	5	0	0	0	0	1	0	0
6. Individual work, diff. tasks	6	0	0	0	0	0	0	0
7. Other, conf. situation	7	2	0	1	1	0	0	10
	TOT:	274	35	637	35	2	0	15

N= 7164

Table 14c (cont.)

C. Rhythmic Movement Expression (N=6 Lessons)

Cluster I: Teacher talk, movement, pupils talk, other

01. Accepts, praises
02. Gives corr. feedback
03. Uses ideas dev. by pup.
04. Asks, init., term. akt.
05. Presents inform., org.
06. Gives dir., comm.
07. Criticizes
08. Answers questions
09. Speaks spontan., init.
10. Silent guidance
11. Silent participation
12. Confused situation

KAT.	1	2	3	4	5	6	7	8	9	10	11	12
1 .	0	0	0	1	8	0	0	0	0	4	0	0
2 .	1	5	0	2	8	0	0	0	0	10	3	0
3 .	0	0	0	0	1	0	0	0	0	2	0	0
4 .	1	2	1	3	22	3	1	5	1	22	2	0
5 .	3	3	0	30	244	6	2	0	6	46	5	2
6 .	0	[illegible]	0	2	8	7	1	0	0	2	1	0
7 .	0	0	0	1	4	0	0	0	0	2	0	0
8 .	0	0	0	1	1	0	0	0	0	0	0	0
9 .	0	0	1	1	5	0	0	0	1	1	0	0
10 .	[illegible]	15	0	20	40	4	2	0	2	222	3	0
11 .	0	3	0	2	5	0	0	0	0	2	126	0
12 .	0	0	0	0	0	0	0	0	0	3	0	0
TOT.	17	32	5	68	352	25	10	5	12	321	143	5

N= 7164

Cluster II: Pupils' collective movement activity/passivity and social access

1. Contacts, ideas cont.
2. Contacts free, ideas cont.
3. Contacts free, ideas open
4. Pupils' spont. activity
5. Pupils follow instruction
6. Pupils organization
7. Pupils wait for turn
8. Confused situation

KAT.	1	2	3	4	5	6	7	8
1 .	30	0	0	0	4	1	0	0
2 .	0	287	0	0	27	6	2	0
3 .	0	0	276	0	7	3	0	1
4 .	0	0	0	3	0	0	0	0
5 .	3	26	8	0	177	14	4	1
6 .	0	4	2	0	17	58	0	0
7 .	1	2	0	0	2	0	6	0
8 .	0	2	1	0	0	0	0	0
TOT:	36	325	291	4	237	85	14	5

N= 7164

Cluster III: Social form

1. Complete class, uniform task
2. Divided class, uniform task
3. Divided class, different task
4. Div. cl. diff. task within gr.
5. Individual work, unif. tasks
6. Individual work, diff. tasks
7. Other, conf. situation

KAT.	1	2	3	4	5	6	7
1 .	329	3	1	0	4	0	1
2 .	2	144	1	0	1	0	0
3 .	0	1	142	0	0	0	0
4 .	0	0	0	52	0	0	0
5 .	4	1	0	0	285	0	1
6 .	0	0	0	0	0	9	0
7 .	4	0	0	0	0	0	0
TOT:	341	150	146	53	292	10	5

N= 7164

Table 14d (cont.)

D. Ball Games (N=6 Lessons)

Cluster I: Teacher talk, movement, pupils talk, other

CAT	1	2	3	4	5	6	7	8	9	10	11	12
01. Accepts, praises	0	1	0	2	4	0	0	0	0	10	6	0
02. Gives corr. feedback	1	2	0	2	9	0	0	0	0	20	1	0
03. Uses ideas dev. by pup.	0	0	0	0	0	0	0	0	0	0	0	0
04. Asks, init., term. akt.	1	2	0	1	22	0	0	4	0	19	0	0
05. Presents inform., org.	0	4	0	23	265	8	1	0	8	50	1	2
06. Gives dir., comm.	0	0	0	0	9	1	0	0	0	2	0	0
07. Criticizes	0	0	0	0	1	0	0	0	0	1	0	0
08. Answers questions	0	0	0	0	2	0	0	0	0	0	0	0
09. Speaks spontan., init.	0	1	1	0	7	0	0	0	0	1	0	0
10. Silent guidance	9	23	0	22	44	3	0	0	1	339	1	2
11. Silent participation	0	1	0	0	3	0	0	0	0	1	9	0
12. Confused situation	0	0	0	0	2	0	0	0	0	1	0	0
TOT	20	39	1	54	374	15	3	5	12	449	16	6

N= 7164

Cluster II: Pupils' collective movement activity/passivity and social access

CAT	1	2	3	4	5	6	7	8
1. Contacts, ideas cont.	143	1	0	0	8	1	0	0
2. Contacts free, ideas cont.	1	429	0	0	16	7	0	2
3. Contacts free, ideas open	0	0	1	0	0	0	0	0
4. Pupils' spont. activity	0	0	0	0	0	0	0	0
5. Pupils follow instruction	8	17	0	0	199	20	0	1
6. Pupils organization	1	5	0	0	21	94	0	0
7. Pupils wait for turn	0	0	0	0	1	0	2	0
8. Confused situation	0	3	0	0	1	0	0	0
TOT:	156	457	2	0	248	124	4	5

N= 7164

Cluster III: Social form

CAT	1	2	3	4	5	6	7
1. Complete class, uniform task	59	1	1	0	0	0	0
2. Divided class, uniform task	0	669	0	1	0	0	3
3. Divided class, different task	0	0	80	0	0	0	0
4. Div. cl. diff. task within gr.	1	0	0	166	0	0	0
5. Individual work, unif. tasks	0	0	0	0	1	0	0
6. Individual work, diff. tasks	0	0	0	0	0	0	0
7. Other, conf. situation	0	3	0	0	0	0	0
TOT:	63	676	82	170	1	0	5

N= 7164

pairs were in steady state cells, in Cluster II more than 80%, and in Cluster III more than 90%. The critical decisions concerning social form, division of labor and responsibility, and the forms of pupil collective activity/passivity were the general dominating aspects when teaching behavior was analyzed.

The two teachers were quite homogeneous but flexible. The sequence and variety were related to different aspects. However, the man teacher was in general less directive. Changes in critical teaching behavior appeared as functions of grade level. The directiveness of the teacher decreased as the age of pupils increased. At the same time, the teacher's silent guidance, participation, use of pupils' ideas, and pupils' responsibility increased, as did the variety of critical sequence patterns.

The interpretation and comparison of matrices describing the instructional process in different content areas of physical education are made in the next step. The results of the major PEIAC/LH-75 parameters, computed from the primary and secondary information of these matrices, are presented and discussed. The displays presented in the four parts of Table 14a, b, c and d are used to enhance and clarify this discription.

Describing the Instructional Process with the Major PEIAC/LH-75 Parameters and Indices

Further analysis included a comparison of the means of each interaction process across class time with PEIAC/LH-75 parameters (Table 3,p. 79). The indices are based on unit coding and the statistical procedures are based on category frequencies, percents and ratios. These are computed separately from matrices of the three clusters. The significance of the differences between the means of PEIAC/LH-75 indices as a function of frame factors (teacher, grade level and P.E. subject areas) was estimated by using the Mann-Whitney U-test and the rank order was determined by functions of variability. The results are presented in Tables 14, 15 and 16, and the statistical differences of the means of PEIAC/LH-75 indices by frame factor are summarized in Table 18.

The indices were used to reduce the data and to give a concentrated picture of the elements of this category system grouped into three clusters.

Variation of the Means of PEIAC/LH-75 Indices as a Function of the Teacher

The significance of differences between indices estimated for the 12 lessons of the two teachers, rated by six coders (T_2), and containing 14,328 six-second time units, are presented in Table 15.

The differences of the two teachers' initiation/response behavior were reflected in pupil behavior. The "pupil verbal initiation ratio" (PVIR), "nonverbal initation ratio" (PIR), and percent of "pupil collective movement activity" were higher for the male teacher than for the female teacher. The differences in the means of these indices were statistically significant. On the other hand, the "teacher question and activity initiation/termination ratio" (TQAR) and the "pupil collective following instruction, organizing ratio" (PIOP) were higher for the female teacher. The differences in the means of these indices were also statistically significant. The "teacher response ratio" (TRR), based on verbal behavior, was only slightly higher for the man teacher than for the woman teacher.

Variation of the Means of PEIAC/LH-75 Indices as a Function of Grade Level

The significance of differences between indices as estimated for the 8 lessons of three grade levels, rated by six coders (T_2), and containing 9,552 six-second time units, are presented in Table 16.

The differences between the instructional processes of the three grade levels were clearly recognized in the parameters describing the general features of the use of time, such as the indices describing pupil verbal/nonverbal behavior and pupil collective movement activity/passivity. The percent of class time used for "pupil talk" (PT) decreased at the middle and upper grade levels (from 4% to 0.86%), whereas the amount of "teacher's silent guidance and participation" (TSPGR) increased (from 23% to 44%). At the middle and upper grade levels, the "teacher verbal praise ratio" (TPR) increased. The differences in these indices were statistically significant. The percent of "pupil collective activity" (PCA) increased at the middle grade level (51% to 66%), whereas the ratio describing pupil collective passivity, in which the "pupils follow instruction, organize themselves" (PIOR), decreased (47% to 32%). The differences between these indices were

Table 1[5]. Significance of Differences between PEIAC/LH-75 Indices Estimated for Two Teachers (Man-Woman) (T_2), Mann-Whitney U-test

No	Symbol	Name of Index	Teacher 1. (N = 12 h)	Rank	Teacher 2. (N = 12 h)	Rank	Differences: Mann-Whitney U-test 1. - 2. z
1	TT	Percent teacher talk	58.17	2.	62.35	1.	-0.98
2	PT	Percent pupil talk	2.73	1.	1.79	2.	-1.36
3	TSAR	Teacher sustained activity ratio	53.04	2.	56.62	1.	-0.64
4	TSGPR	Teacher silent guidance and participation ratio	39.88	1.	36.13	2.	-0.58
5	TRR	Teacher response ratio	30.57	1.	28.14	2.	-0.17
6	TQAR	Teacher question and activity initiation-termination ratio	10.13	2.	18.63	1.	-2.37[xx]
7	CCR	Content emphasis ratio	45.11	2.	48.07	1.	-0.35
8	PVIR	Pupil verbal initiation ratio	81.33	1.	62.65	2.	-3.33[xxx]
9	PIR	Pupil initiation ratio (verbal and nonverbal)	94.57	1.	77.85	2.	-1.85[x]
10	TPR	Teacher praise ratio	77.22	2.	84.07	1.	-0.61
11	PCA	Percent pupil collective activity	66.69	1.	55.18	2.	-2.54[xx]
12	PSAR	Pupil sustained activity ratio	86.48	1.	82.22	2.	-1.56
13	PSAR	Pupil social access ratio	13.24	2.	15.20	1.	-0.29
14	PIOR	Pupil collective following instruction,organizing ratio	32.27	2.	43.43	1.	-2.48[xx]
15	SGWR	Pupil social group work ratio	35.65	1.	6.26	2.	-0.46
16	PIWR	Pupil individual work ratio	9.81	1.	5.71	2.	-0.34
17	SFVR	Social form variability ratio	7	1.	7	2.	-0.06
18	SSFR	Sustained social form ratio	97.38	2.	97.65	1.	-0.96

6 observers

N - 14328 6 second time units

Levels of significance

x = $p < 0.05$

xx = $p < 0.01$

xxx = $p < 0.001$

Table 16. Significance of Differences between PEIAC/LH-75 Indices Estimated for Three Grade Levels (T_2), Mann-Whitney U-test

No	Symbol	Name of Index	Lower level 1. (N = 8 h)	Rank	Middle level 2. (N = 8 h)	Rank	Upper level 3. (N = 8 h)	Rank	Mann-Whitney U-test Differences: 1. - 2. z	1. - 3. z	2. - 3. z
1	TT	Percent teacher talk	66.19	1.	60.13	2.	54.46	3.	-1.26	-1.58	-0.42
2	PT	Percent pupil talk	4.28	1.	1.64	2.	0.86	3.	-2.63[xx]	-3.05[xxx]	-1.68[x]
3	TSAR	Teacher sustained activity ratio	53.06	2.	50.23	3.	61.20	1.	-0.42	-1.37	-1.79[x]
4	TSGPR	Teacher silent guidance and participation ratio	23.90	3.	40.96	2.	44.78	1.	-1.68[x]	-1.79[x]	-0.74
5	TRR	Teacher response ratio	26.64	3.	27.74	2.	33.09	1.	-0.11	-1.47	-0.74
6	TQAR	Teacher question and activity initiation-temination ratio	14.88	2.	14.94	1.	13.64	3.	-0.53	-0.53	-0.21
7	CCR	Content emphasis ratio	51.71	1.	44.92	2.	43.13	3.	-1.37	-1.58	-0.42
8	PVIR	Pupil verbal initiation ratio	69.44	2.	17.83	3.	80.49	1.	-0.79	-0.74	-0.32
9	PIR	Pupil initiation ratio (verbal and nonverbal)	76.43	2.	32.25	3.	99.95	1.	-1.16	-1.26	-0.11
10	TPR	Teacher praise ratio	48.25	3.	74.87	2.	98.73	1.	-1.47	-2.80[xx]	-2.70[xx]
11	PCA	Percent pupil collective activity	51.24	3.	66.49	1.	65.06	2.	-2.31[xx]	-1.89[x]	-0.37
12	PSAR	Pupil sustained activity ratio	83.11	3.	84.10	2.	85.85	1.	-0.42	-0.84	-0.21
13	PSAR	Pupil social access ratio	6.99	3.	14.42	2.	19.46	1.	-0.74	-1.00	-0.54
14	PIOR	Pupil collective following instruction organizing ratio	47.51	1.	32.30	3.	33.76	2.	-2.31[xx]	-1.89[x]	-0.32
15	SGWR	Pupil social group work ratio	29.20	3.	38.51	2.	41.07	1.	-0.53	-0.63	-1.47
16	PIWR	Pupil individual work ratio	6.10	3.	9.27	1.	7.84	2.	-0.72	-0.72	-0.14
17	SFVR	Social form variability ratio	7	1.	7	1.	6	3.	-0.49	-0.50	-0.00
18	SSFR	Sustained social form ratio	97.25	3.	97.69	1.	97.60	2.	-1.27	-0.32	-0.32

6 observers

N = 9552 6 second time units

Levels of significance

x = $p < 0.05$

xx = $p < 0.01$

xxx = $p < 0.001$

statistically significant. The "pupil individual work ratio" (PIWR) was at its highest at the middle grade level, but the differences in this variable between grade levels were not statistically significant.

Variation of the Means of PEIAC/LH-75 Indices as a Function of the Content of P.E. Subject Areas

The significance of differences between indices as estimated for 6 lessons of four P.E. subject area, rated by six coders (T_2), and containing 7,164 six-second time units, are presented in Table 16. Differences in indices were strongly related to the content of the subject areas. Fourteen of the eighteen indices produced statistically significant differences. These will be presented by referring to the rank order of the indices.

The percent of class time devoted to "teacher talk" (TT) was highest (71%) in gymnastics and lowest in rhythmic movement expression (50%) and ball games (51%). Both the "teacher sustained activity ratio" (TSAR) and the "teacher silent guidance and participation ratio" (TSGPR) were highest in ball games and rhythmic movement expression and lowest in gymnastics. The "teacher response ratio" (TRR), which was adapted from Flanders' ID-ratio, was highest in gymnastics, second highest in apparatus and lowest in ball games.

Typical of gymnastics was a high percentage for the "teacher question, activity initiation and termination ratio" (TQAR). This ratio was second highest in rhythmic movement expression and lowest in apparatus. The "content emphasis ratio" (CCR) was highest in gymnastics and second highest in apparatus and lowest in rhythmic movement expression.

The "pupil verbal initiation ratio" (PVIR) was highest in apparatus and lowest in ball games. The variability of "pupil verbal and nonverbal initiation ratio" (PIR) was great. It was highest in rhythmic movement expression and lowest in gymnastics and ball games. The "pupil sustained movement activity ratio" (PSAR) was highest in apparatus, second highest in ball games and lowest in gymnastics.

The "pupil social access ratio" (PSAR), measured with "pupil movement activity," was strongly related to the subject area. It was highest in rhythmic movement expression and lowest in ball games. It should be noted also that the differences between the indices describing the

Table 17. Significance of Differences between PEIAC/LH-75 Indices Estimated for Four Subject Areas (T_2), Mann-Whitney U-test

No	Symbol	Name of Index	Gymnastics 1. (N = 6 h)	Rank	Apparatus 2. (N = 6 h)	Rank	Rhythmic movement express 3. (N = 6 h)	Rank	Ball games 4. (N = 6 h)	Rank	Mann-Whitney U-test Differences: 1. - 2. z	1. - 3. z	1. - 4. z	2. - 3. z	2. - 4. z	3. - 4. z
1	TT	Percent teacher talk	71.48	1	67.46	2	51.13	3	50.98	4	-0.80	-2.56^{xx}	-2.72^{xx}	-2.08^{x}	-2.56^{xx}	0.00
2	PT	Percent pupil talk	2.73	1	2.71	2	1.82	3	1.77	4	-0.40	0.00	-0.16	-0.96	-1.28	-0.08
3	TSAR	Teacher sustained activity ratio	44.81	4	50.95	3	61.18	2	62.38	1	-1.76^{x}	-2.40^{xx}	-2.56^{xx}	-1.44	-1.92^{x}	-0.56
4	TSGPR	Teacher silent guidance and participation ratio	26.11	4	30.29	3	47.64	2	47.77	1	-0.80	-2.24^{xx}	-2.40^{xx}	-2.24^{xx}	-2.56^{xx}	-0.32
5	TRR	Teacher response ratio	38.25	1	35.99	2	35.79	3	19.28	4	-0.32	-0.16	-2.08^{x}	0.00	-2.56^{xx}	-1.76^{x}
6	TQAR	Teacher question and activity initiation-termination ratio	22.37	1	6.10	4	16.27	2	12.70	3	-2.88^{xx}	-1.28	-1.44	-2.24^{xx}	-1.28	-0.80
7	CER	Content emphasis ratio	52.74	1	48.71	2	42.04	4	42.86	3	-1.44	-1.92^{x}	-2.40^{xx}	-1.12	-0.80	-0.32
8	PVIR	Pupil verbal initiation ratio	73.78	2	85.74	1	70.23	3	61.65	4	-1.36	-1.13	-0.16	-0.32	-1.92^{x}	-0.97
9	PIR	Pupil initiation ratio (verbal and nonverbal)	66.25	4	93.24	2	115.25	1	72.08	3	-1.76^{x}	-2.08^{x}	-0.48	-1.44	-2.42^{xx}	-2.24^{xx}
10	TPR	Teacher praise ratio	79.72	3	86.13	1	10.93	4	84.39	2	-0.16	-0.49	-0.65	-0.48	-0.56	-0.16
11	PCA	Percent pupil collective activity	58.58	3	57.73	4	65.79	1	61.63	2	-0.00	-0.96	0.00	-0.96	-0.32	-0.48
12	PSAR	Pupil sustained activity ratio	77.85	4	88.47	1	83.98	3	87.12	2	-2.88^{xx}	-1.92^{x}	-2.56^{xx}	-1.44	-0.48	-0.96
13	PSAR	Pupil social access ratio	2.74	3	5.10	2	45.02	1	0.43	4	-0.33	-1.93^{x}	-0.82	-2.26^{xx}	-0.08	-2.41^{xx}
14	PIOR	Pupil collective following instruction, organizing ratio	40.30	2	41.49	1	32.28	4	37.35	3	0.00	-1.44	-0.00	-0.96	-0.32	-0.80
15	SGWR	Pupil social group work ratio	14.56	4	68.30	1	20.10	3	25.48	2	-2.89^{xx}	-0.32	-0.32	-2.72^{xx}	-1.92^{x}	-0.00
16	PIWR	Pupil individual work ratio	0.00	4	0.23	2	30.47	1	0.20	3	-1.00	-3.08^{xxx}	-1.00	-2.99^{xxx}	-0.12	-2.99^{xxx}
17	SFVR	Social form variability ratio	5	4	6	2	7	1	6	2	-0.70	-2.00^{x}	-1.55	-1.24	-0.58	-1.08
18	SSFR	Sustained social form ratio	97.96	2	97.77	3	96.33	4	97.99	1	-0.08	-2.72^{xx}	-0.08	-2.40^{xx}	-0.32	-2.72^{xx}

N = 7164 6 seconds time units
6 observers

Levels of significance
x = $p < 0,05$
xx = $p < 0.01$
xxx = $p < 0.001$

division of labor and responsibility were clearly related to the content of the subject area.

The "pupil social group work ratio" (SGWR) was highest in apparatus and lowest in gymnastics, whereas the "individual work ratio" (PIWR) was high only in rhythmic movement expression and could not be estimated for gymnastics with this data. The "sustained social form ratio" (SSFR) was highest in ball games and lowest in rhythmic movement expression.

In only four of the eighteen indices were the differences between subject areas not statistically significant. These were such general characteristics as "pupil talk" (PT), "pupil collective activity ratio" (PCA), and "pupil collective following instruction, organizing ratio" (PIOR). As stated earlier, these characteristics are all related to pupil behavior, and thus to grade level.

<u>Summary</u>

In all 18 main parameters of the PEIAC/LH-75 system, statistically significant differences were found as a function of the identified frame factors, teacher, grade level and P.E. subject areas. Five of these differences were related to teachers, six to grade level, and fourteen to the subject areas in physical education (Table 18).

The teaching behavior of the man and the woman teacher in this study was quite homogeneous in many different contexts and they were rather flexible in their behavior. However, a difference between the teachers' initiation response behavior was discernible. It was related to pupil behavior and appeared to reflect the training background of the teachers. Within the teacher response behavior parameter, the praise ratio increased due to pupil behavior and content-centeredness diminished mainly in response to pupil behavior.

The influence of subject specific content on the instructional process was dominant and was reflected in the different aspects indicative of initiation/response behavior. The main point was thus, not the subject matter of physical education as such, but the kinds of content it consisted of, and how the instructional processes were arranged to accommodate them. The temporal basis of the instructional process, described, e.g., by analyzing the amount of teacher talk, silent guidance and participation, as well as teacher sustained activity, pupil

Table 18 Summary. Significance of Differences between PEIAC/LH-75 Indices Estimated for Two Teachers, Three Grade Levels and Four Subject Areas (T_2), Mann-Whitney U-test

No	Symbol	Name of Index	Teachers N = 12	Grade levels N = 6			Subject areas N = 6					
			Man-Woman	L-M	L-U	M-U	G-A	G-R	G-B	A-R	A-B	R-B
			z	z	z	z	z	z	z	z	z	z
1	TT	Percent teacher talk						XX	XX	X	XX	
2	PT	Percent pupil talk		XX	XXX	X						
3	TSAR	Teacher sustained activity ratio				X	X	XX	XX		X	
4	TSGPR	Teacher silent guidance and participation ratio		X	X			XX	XX	XX	XX	
5	TRR	Teacher response ratio							X		XX	X
6	TQAR	Teacher question and activity initiation-termination ratio	XX				XX			XX		
7	CCR	Content emphasis ratio						X	XX			
8	PVIR	Pupil verbal initiation ratio	XXX								X	
9	PIR	Pupil initiation ratio (verbal and nonverbal)	X				X	X			XX	XX
10	TPR	Teacher praise ratio			XX	XX						
11	PCA	Percent pupil collective activity	XX	XX	X							
12	PSAR	Pupil sustained activity ratio					XX	X	XX			
13	PSAR	Pupil social access ratio						X		XX		XX
14	PIOR	Pupil collective following instruction, organizing ratio	XX	XX	X							
15	SGWR	Pupil social group work ratio					XX			XX	X	
16	PIWR	Pupil individual work ratio						XXX		XXX		XXX
17	SFVR	Social form variability ratio						X				
18	SSFR	Sustained social form ratio						XX		XX		XX

Levels of significance
X = p 0.05
XX = p 0.01
XXX = p 0.001

L = Lower level
M = Middle level
U = Upper level

A = Apparatus
B = Ball games
G = Gymnastic
R = Rhythmic movement express

sustained activity and sustained social form ratios, was clearly related to the content of the subject area.

The social forms, division of labor and responsibility between teacher and pupils, and among pupils, were strongly related to the content and quality of the subject matter. The pre-interactive decisions strongly determined the environment of the instructional process and its progress across time.

It can be concluded that the major PEIAC/LH-75 parameters were able to provide concentrated information about the directiveness/nondirectiveness of teacher behavior and about how the frame factors used in this study influenced the teaching process in the gymnasium. The importance of preserving the sequence when categorizing these three aspects of teaching was emphasized in this study.

PART II
RELIABILITY AND OBJECTIVITY OF CODING

This section of the report of the study results deals with the problem of the reliability of coding attached to the use of the observation system PEIAC/LH-75 which is intended to measure the interaction process of physical education classes. In research work using observational systems, the testing of hypotheses requires that the observation system employed possesses sufficiently high reliability. Therefore, in developing and constructing a measuring instrument it is crucial to provide data pertaining to reliability, as well as to discuss which reliability measures were selected and why. The question of the reliability of observation systems is a complicated one because the classification system and coder together constitute the measuring instrument. Therefore, in evaluating its usefulness attention must be paid both to the quality of the information utilized and to the way in which it is used in the coding process. Because the value of results in observational studies depends crucially on the manner in which the instrument has been used in the coding process, an effort is made in the present study to concentrate on these aspects of evidence associated with reliability, that is, on the objectivity of coding. In this context it signifies the degree of independence between the final results of coding and the coder himself (Komulainen, 1970; 1973).

In examining the overall reliability of this observation instrument, the customary profile method, or total-events-approach, of Scott (1955) was applied. It was also considered appropriate to apply a method used in non-parametric measurement, the coefficient of concordance (W) elaborated by Kendall, to examine the reliability of various individual categories and to determine the applicability of various methods in examining the problem of objectivity of coding. Because this is a multidimensional classification system, every dimension had to be studied both separately and in conjunction with other clusters.

The purpose of this portion of the study, then, was (1) to determine (a) the within-occasion reliability (agreement) and (b) between-occasion reliability (constancy) (i) by cluster, (ii) by coder pair,

(iii) by situation, and (iv) by content of lessons; (2) to examine the reliability of the various individual categories (a) by category and by cluster, (b) between clusters, and (c) by coding occasion; and (3) to examine the applicability of the different methods used for assessing the reliability of a multidimensional observation instrument.

Results Concerning Overall Reliability

The reliability components, within-occasion reliability (agreement) and between-occasion reliability (constancy), were examined by clusters, by coder pairs, by different coding circumstances and by different content situations of physical education classes (teacher, grade level and subject area). The final results give some idea of the experimental use of the observation instrument and of the variation in the level of mean values for different reliability components in the three clusters.

In evaluating the results it must be remembered that the number of categories in the three clusters is not equal, but 12, 8 and 7, respectively. The estimated role of chance, which is subtracted in Scott's pi, decreases as the number of categories increases. Thus, the probable role of chance was the least in the Verbal Cluster I. The relative frequency of occurence of the categories is also taken into consideration by using Scott's coefficient. The mean values were highly sensitive to extreme variations and the range of variation of the six coders' coefficients by pair was shown to be remarkably wide.

A total of 8424 Scott's coefficients were computed. The differences of the means of coefficients were examined with the use of a t-test, and in some cases with both a t-test and a one-way analysis of variance (ANOVA). This method was chosen because the groups to be compared were usually more than two in number. A total of 1252 t-tests and 63 ANOVAs were computed. In interpreting the t-test, the effect of overlapping classifications at the risk-level limit was taken into consideration and thus the chosen risk-level of t values for $p>.01$ was not regarded as significant.

Clusters (I,II,III)

The differences between clusters are presented in (Table 19). The average level of mean coefficient values by cluster was ratherlow (.61, .65 and .69), and varied greatly between the different reliability components. An inter-coder agreement of .65, a within-coder constancy of .69 and a between-coder constancy of .60 were indicated in the scores of the video-recorded observations.

Table 19. Analysis by Cluster: Inter-coder Agreement, Within-coder Constancy and Between-coder Constancy. Mean Values and Standard Deviations of Scott's Pi Coefficients by Cluster (I,II,III) and by Occasion (T_1,T_2,T_3)

	CLUSTER I (Verbal)		CLUSTER II (Movement & Social Access)		CLUSTER III (Social Form)	
	X	SD	X	SD	X	SD
INTER-CODER AGREEMENT (N=360)						
Live Situation (T_1)	.57	.17	.61	.26	.75	.28
Videotape Recording 1 (T_2)	.61	.18	.71	.22	.77	.36
Videotape Recording 2 (T_3)	.61	.19	.59	.36	.60	.59
WITHIN-CODER CONSTANCY (N=144)						
T_1-T_2	.66	.15	.59	.28	.62	.48
T_2-T_3	.71	.13	.66	.31	.69	.47
BETWEEN-CODER CONSTANCY (N=720)						
T_1-T_2	.54	.18	.56	.30	.61	.47
T_2-T_3	.59	.19	.62	.32	.62	.54

Examining the mean Scott's pi coefficient values of the coding system and the corresponding standard deviations for the videotaped observations, systematic differences in inter-coder agreement between clusters may be noted (Table 20). The mean coefficient values of Cluster I (Verbal) were the lowest and their standard deviations the smallest. There was no difference between the mean coefficient values in the videotaped material coding occasions T_2 and T_3. In Cluster II (Movement and Social Access), the mean values were slightly higher than those in the previous cluster and the range of standard deviations was larger. A great mean value variation (.71-.59) and statistically significant difference was noted in this cluster between the two videotape coding occasions T_2 and T_3. In Cluster III (Social Form), the mean coefficient values were the highest and the range of standard deviations the greatest.

Table 20. Analysis by Cluster: Differences in Means of Scott's Pi Coefficients Computed Separately by Cluster (I,II,III) and by Occasion (T_2 and T_3) (p=<.01)

	CLUSTER I (Verbal)		CLUSTER II (Movement & Social Access)		CLUSTER III (Social Form)		DIFFERENCES		
							I-II	I-III	II-III
	X	SD	X	SD	X	SD	t	t	t
Intercoder agreement									
T_2	.61	.18	.71	.22	.77	.36	-6.67	-7.53	-2.70
T_3	.61	.19	.59	.36	.60	.59	0.93	0.31	-0.27
N=360, df=718									
Within-coder constancy									
T_2-T_3	.71	.13	.66	.31	.69	.47	1.78	0.49	-0.64
N=144, df=286									
Between-coder constancy									
T_2-T_3	.59	.19	.62	.32	.62	.54	-2.16	-1.41	0.00
N=720, df=1438									

Differences in inter-coder agreement between all clusters were found to be statistically significant in the first videorecorded observation (T_2), but in the second videotape recording no statistically significant differences were found. The main difference between clusters was thus the constancy of variation in the inter-coder agreement coefficient level between coding occasions. This variation was smallest in Cluster I and greatest in Cluster II.

The comparison of the mean Scott's pi coefficient values showed that the values for within-coder constancy were higher than for inter-coder agreement and between-coder constancy in all clusters. The differences were quite noticeable in the Verbal Cluster I, where the level of the reliability coefficients as a whole was highest (.71). Also, the mean standard deviations of the Scott's pi coefficients varied noticeably between clusters (.13, .31, .47). However, statistically significant differences were not found between the mean coefficient values in the different clusters.

The level of between-coder constancy coefficients was found to be lower than the other reliability components in all clusters, and was the lowest (.59) in the verbal cluster. The differences between clusters were not found to be statistically significant.

In view of the results, it can be stated that the coding of the verbal cluster deviated from the other two clusters, among other things, in the systematic character of the between-coder constancy variation. The observers' coding of verbal events was more constant, but the differentiation between coders increased. Because this differentiation was not, however, reflected in a decreasing level of inter-coder agreement (T_3), it was apparent that the differences between coders were somehow compensated for in this cluster. In the other clusters the differences in between-coder constancy coefficients were minor, and differentiation was reflected in the decreasing level of inter-coder agreement (T_3). This differentiation of coders was, however, fairly random in character.

The structure of the coding system, the coders' behavior, and the characteristics of the coded phenomena were reflected in the results. The observation of verbal, logical communication was apparently more familiar to the coders and the interpretation of its features more constant than the observation of the other features of communication (non-verbal). The quality of the target of observation, such as tempo variation, was reflected in the results. The possible coding differences were more outstanding when a slowly changing phenomenon, such as the social form, was in question. This was found to be the case, for instance, in the considerable variation of the mean value levels within clusters.

Coder Pairs and Coders

Comparing the inter-coder agreement of coder pairs, statistically significant differences were found in the coding of verbal behavior (Tables 21 and 22). The sixth coder deviated systematically from the other five. Apparently he had internalized the concepts differently from the others, and the way in which he used the metalanguage of the coding system was unique. This lowered the mean level of agreement appreciably.

In addition to the previous findings, the six coders deviated noticeably from each other both in the average reliability coefficient level and in standard deviations, as well as in their occurence in different clusters (Table 23). The range of the Scott's pi coefficients for the different coders was greatest in Cluster III (.49-.82), second

Table 21. Analysis by Coder Pairs: Inter-coder Agreement. Mean Values and Standard Deviations of Scott's Pi Coefficients for the Videotaped Material by Cluster (I,II,III) and by Occasion (T_2,T_3)(N=24)

CODER PAIRS	CLUSTER I (Verbal)		CLUSTER II (Movement & social access)		CLUSTER III (Social form)	
	X	SD	X	SD	X	SD
Occasion T_2						
1,2	.66	.16	.76	.16	.79	.28
1,3	.66	.14	.73	.24	.88	.23
1,4	.62	.21	.70	.21	.68	.47
1,5	.63	.15	.71	.19	.86	.24
1,6	.46	.19	.61	.27	.73	.39
2,3	.66	.16	.73	.20	.88	.20
2,4	.63	.20	.76	.17	.68	.46
2,5	.60	.18	.74	.14	.85	.20
2,6	.48	.16	.65	.23	.73	.36
3,4	.68	.16	.73	.23	.74	.45
3,5	.73	.10	.72	.24	.92	.60
3,6	.55	.19	.63	.29	.76	.34
4,5	.68	.15	.78	.15	.73	.45
4,6	.55	.14	.70	.23	.60	.15
5,6	.56	.18	.72	.20	.77	.34
	.61	.18	.71	.22	.77	.36
Occasion T_3						
1,2	.72	.11	.68	.27	.66	.40
1,3	.68	.12	.74	.18	.76	.38
1,4	.66	.14	.61	.34	.50	.65
1,5	.72	.12	.59	.35	.63	.44
1,6	.44	.22	.49	.58	.44	.74
2,3	.67	.13	.75	.19	.19	.32
2,4	.64	.17	.62	.32	.59	.62
2,5	.68	.16	.59	.29	.64	.36
2,6	.48	.20	.51	.14	.48	.75
3,4	.70	.13	.62	.32	.60	.63
3,5	.68	.11	.59	.13	.74	.29
3,6	.43	.20	.49	.44	.60	.70
4,5	.70	.10	.73	.20	.63	.63
4,6	.47	.18	.42	.15	.42	.82
5,6	.51	.20	.44	.48	.51	.72
	.61	.19	.59	.36	.60	.59

TABLE 22A. SIGNIFICANCE OF THE DIFFERENCES BETWEEN CODER PAIR MEANS OF SCOTT'S PI COEFFICENTS;T-TEST.

CODING OCCASION	I CLUSTER - Teacher talk - Pupil talk - Silent Teacher Activity

coder pairs

T_1

	1,2	1,3	1,4	1,5	1,6	2,3	2,4	2,5	2,6	3,4	3,5	3,6	4,5	4,6	5,6
1,3	1.40														
1,4	.79	-.66													
1,5	.80	-.86	-.09												
1,6	-1.90	**-3.80**	**-3.02**	**-3.39**											
2,3	-.91	-2.47	-1.81	-1.95	.93										
2,4	-.03	-1.61	-.92	-.95	2.09	.97									
2,5	-1.12	-2.64	-2.00	-2.15	.65	-.22	-1.19								
2,6	**-2.76**	**-4.77**	**-3.96**	**-4.47**	-1.00	-1.85	**-3.05**	-1.52							
3,4	.53	.94	-.29	-.25	**2.79**	1.54	.62	1.73	**3.65**						
3,5	2.09	.74	1.40	1.76	**4.74**	3.23	2.41	**3.37**	**5.75**	1.71					
3,6	-1.27	**-2.74**	-2.13	-2.27	.40	-.43	-1.35	-.20	1.23	-1.87	**-3.43**				
4,5	.93	-.31	.25	.35	2.84	1.83	1.04	2.00	**3.66**	.50	-.92	2.12			
4,6	-.80	-2.45	-1.75	-1.91	1.17	.16	-.56	.40	2.12	-1.46	**-1.27**	.59	-1.76		
5,6	-.44	-1.69	-1.27	-1.34	1.42	.46	-.45	.68	2.28	-1.03	-2.60	.85	-1.35	.33	

T_2

	1,2	1,3	1,4	1,5	1,6	2,3	2,4	2,5	2,6	3,4	3,5	3,6	4,5	4,6	5,6
1,3	.05														
1,4	-.69	-.77													
1,5	-.56	-.65	.22												
1,6	**-3.83**	-4.08	**-2.76**	**-3.44**											
2,3	-.02	-.07	.68	.55	**3.16**										
2,4	-.51	-.58	.18	-.02	**3.01**	-.50									
2,5	-1.05	-1.15	-.27	-.55	2.67	-1.04	-.47								
2,6	**-3.93**	**-4.23**	**-2.70**	**-3.50**	.31	**-3.96**	**-2.58**	-2.61							
3,4	.41	.38	1.05	1.00	**4.26**	.32	.88	1.45	**4.41**						
3,5	1.92	2.04	2.37	**2.73**	**6.12**	1.98	2.23	**2.98**	**6.69**	1.48					
3,6	-2.55	**-2.72**	-1.62	-2.12	1.17	**-2.56**	-1.85	-1.96	.97	**-2.95**	**-4.62**				
4,5	.62	.61	1.25	1.24	**4.36**	.64	1.07	1.67	**4.77**	.20	-1.32	**3.22**			
4,6	-2.46	-2.69	-1.39	-1.[illegible]6	1.79	-2.48	1.64	-1.19	1.65	**-2.59**	**-5.09**	**.46**	**-3.27**		
5,6	-2.00	-2.16	-1.07	-1.53	1.86	-2.01	-1.31	-.68	1.73	-2.42	**-4.15**	.63	2.69	.24	

T_3

	1,2	1,3	1,4	1,5	1,6	2,3	2,4	2,5	2,6	3,4	3,5	3,6	4,5	4,6	5,6
1,3	-1.28														
1,4	-1.75	-.54													
1,5	-.00	1.24	1.71												
1,6	**-5.70**	**-4.72**	**-4.19**	**-5.63**											
2,3	-1.59	-.33	.21	-1.54	**4.42**										
2,4	-2.08	-.98	-.47	-2.04	**3.59**	-.68									
2,5	-1.01	.07	.54	-.99	**4.43**	.36	.94								
2,6	**-5.24**	**-4.09**	**-3.53**	**-5.06**	**.78**	**-3.77**	**-2.92**	**-3.81**							
3,4	-.46	.73	1.21	-.45	**5.36**	1.03	1.58	.57	**4.56**						
3,5	-1.35	-.01	.54	-1.33	4.80	.33	1.00	-.09	**4.17**	-.77					
3,6	**-6.26**	**-5.18**	**-4.58**	**-4.16**	-.09	**-4.83**	**-3.91**	**-4.21**	-.91	**-5.63**	**-5.23**				
4,5	-.72	.64	1.16	-.69	**5.31**	.57	1.56	.46	4.72	-.18	.63	**5.85**			
4,6	**-5.50**	**-4.74**	**-4.10**	**-5.79**	.51	**-4.19**	**-3.41**	**-4.36**	-.31	**-5.22**	**-4.85**	.64	**-5.46**		
5,6	**-4.46**	**-3.45**	**-2.92**	**-4.35**	1.22	**-3.15**	-2.36	**-3.21**	.48	**-3.93**	**-3.15**	1.30	**-4.04**	.80	

N = 24
df = 46

____ = p o.o l (underlined values, shown in bold)

T_1 = direct observation
T_2 = videotape observation 1.
T_3 = videotape observation 2.

TABLE 22 B. SIGNIFICANCE OF THE DIFFERENCES BETWEEN THE CODER PAIR MEANS OF SCOTT'S PI COEFFICENT; T-TEST.

CODING OCCASION	II CLUSTER – Social Access – (pupils collective movement activity/passivity)
T_1, T_2, T_3	coder pairs

T_1

	1,2	1,3	1,4	1,5	1,6	2,3	2,4	2,5	2,6	3,4	3,5	3,6	4,5	4,5	5,6
1,3	.20														
1,4	-.70	-.68													
1,5	.05	-.14	.70												
1,6	.07	-.11	.70	.02											
2,3	.25	.04	.94	.18	.15										
2,4	-.21	-.37	.39	-.24	-.25	-.41									
2,5	1.00	.81	1.58	.89	.84	-.60	1.02								
2,6	-.59	-.77	.11	-.60	-.60	-.83	-.29	-1.49							
3,4	-.94	-1.11	-.24	-.93	-.93	-1.18	-.60	-1.80	-.35						
3,5	.69	.49	1.31	.59	.55	.47	.76	-.33	1.22	1.54					
3,6	-.00	-.21	.71	-.05	-.07	-.26	.21	-1.03	.60	.96	-.72				
4,5	-.07	-.25	.58	-.11	-.13	-.30	.14	-.98	.48	.81	-.70	-.07			
4,6	-.98	-1.15	-.33	-.99	-.98	-1.21	-.67	-1.79	-.44	-.11	-1.54	-1.00	-.87		
5,6	-.44	-.63	.26	-.46	-.47	-.69	-.17	-1.38	.15	.53	-1.09	-.45	-.35	.58	

T_2

	1,2	1,3	1,4	1,5	1,6	2,3	2,4	2,5	2,6	3,4	3,5	3,6	4,5	4,6	5,6
1,3	-.45														
1,4	-1.07	-.89													
1,5	-.85	-.27	.25												
1,6	-2.32	1.68	-1.31	-1.58											
2,3	-.45	.04	.58	.35	1.83										
2,4	.15	.56	1.18	.96	2.39	.57									
2,5	-.33	.22	.66	.61	2.19	.19	-.47								
2,6	-1.18	-1.17	-.74	-1.02	.60	-1.30	-1.90	-1.65							
3,4	-.37	.07	.57	.35	1.75	.63	-.48	-.13	1.24						
3,5	-.68	-.20	.28	.05	1.49	-.26	-.79	-.46	.56	-.27					
3,6	-1.83	-1.26	-.83	-1.13	.31	-1.38	-1.88	-1.65	-0.24	-1.33	-1.09				
4,5	.45	.81	1.49	1.29	2.69	.66	.28	.82	2.21	.74	1.05	2.13			
4,6	-.56	-.43	.05	-.22	1.31	-.51	-1.06	-.75	.76	-.53	-.22	.90	-1.35		
5,6	-.65	-.13	.40	.15	1.66	-.19	-.77	-.41	1.12	-.21	.69	1.22	-1.36	.33	

T_3

	1,2	1,3	1,4	1,5	1,6	2,3	2,4	2,5	2,6	3,4	3,5	3,6	4,5	4,6	5,6
1,3	.90														
1,4	-.83	-1.71													
1,5	-1.06	-1.93	-.23												
1,6	-1.67	-2.37	-.95	-.75											
2,3	1.09	.25	1.05	2.06	2.48										
2,4	-.70	-1.61	.15	.38	1.69	-1.76									
2,5	-1.11	-2.16	-.18	.07	.85	-2.31	-.35								
2,6	-1.73	-2.57	-.92	-.70	.11	-2.68	-1.08	-.62							
3,4	-.78	-1.70	.07	.31	1.03	-1.65	-.07	.27	1.01						
3,5	-1.14	-2.12	-.24	.00	.70	-2.27	-.41	-.07	.73	-.33					
3,6	-1.79	-2.56	-1.02	-.80	.02	-2.65	-1.17	-.97	-.13	-1.31	-.84				
4,5	.65	-.27	1.47	1.70	2.19	-.50	1.37	1.87	2.15	1.45	1.86	2.37			
4,6	-2.26	-2.69	-1.49	-1.28	-.51	2.99	-1.63	-1.41	-.65	-1.23	-1.33	-.51	-2.72		
5,6	-2.12	[illegible]	-1.39	-1.17	-.31	2.96	-1.53	-1.31	-.41	-1.22	-1.22	[illegible]	-2.67	.14	

N = 24
d f = 46

____ = p 0.01

T_1 = direct observation
T_2 = videotape recording 1
T_3 = videotape recording 2

TABLE 22 C. SIGNIFICANCE OF THE DIFFERENCES BETWEEN THE CODER PAIR MEANS OF SCOTT'S PI COEFFICENT; T-TEST.

CODING OCCASION	III CLUSTER - Social Form (Division of labour and responsibility)
	coder pairs

T_1

	1,2	1,3	1,4	1,5	1,6	2,3	2,4	2,5	2,6	3,4	3,5	3,6	4,5	4,6	5,6
1,3	.10														
1,4	-.97	-1.13													
1,5	.67	.60	1.71												
1,6	-1.00	-1.10	-.29	-1.50											
2,3	.07	-.02	.99	-.55	1.02										
2,4	-1.36	-1.52	-.43	-2.07	-.01	-1.35									
2,5	-.11	-.20	.73	-.68	.83	-.17	1.08								
2,6	-1.37	-1.47	-.71	-1.85	-.37	-1.38	-.42	-1.19							
3,4	-.40	-.54	.64	1.15	.74	-.45	1.05	-.23	1.15						
3,5	.53	.44	1.66	-.22	1.43	.40	2.05	.55	1.79	1.05					
3,6	-.77	-.87	.00	-1.32	.25	-.60	.35	-.61	.63	-.49	-1.24				
4,5	-.27	-.39	.68	-.95	.78	-.33	1.07	-.13	1.17	.10	-.83	.54			
4,6	-1.10	-1.22	-.32	-1.60	.09	-1.12	.04	-.91	.40	-.83	-1.62	-.27	-.67		
5,6	-.75	-.81	-.04	-1.22	.17	-.78	.23	-.61	.52	-.50	-1.14	-.07	-.55	.10	

T_2

	1,2	1,3	1,4	1,5	1,6	2,3	2,4	2,5	2,6	3,4	3,5	3,6	4,5	4,6	5,6
1,3	1.14														
1,4	-1.00	-1.64													
1,5	.93	-.22	1.69												
1,6	-.62	-1.59	.41	-1.42											
2,3	1.21	.22	1.89	.24	1.65										
2,4	-1.02	-1.68	-.00	-1.73	-.42	-1.93									
2,5	.08	-.46	1.61	-.22	1.32	-.51	1.65								
2,6	-.63	-1.65	.44	-1.47	.02	-1.72	.45	-1.37							
3,4	-.45	-1.30	.47	-1.15	.09	-1.34	.48	-1.05	.08						
3,5	2.24	.97	2.52	1.24	2.42	1.07	2.58	1.79	2.58	1.96					
3,6	-.31	-1.35	.71	-1.16	.31	-1.41	.72	-1.05	.36	.18	-2.27				
4,5	-.52	-1.37	.42	-1.23	.03	-1.42	.43	-1.13	.01	-.65	-1.04	-.25			
4,6	-1.59	-2.40	-.55	-2.26	-.55	-2.45	-.55	-2.19	-1.03	-1.01	-3.06	-1.29	-.96		
5,6	-.27	-1.13	.74	-1.13	.34	-1.37	.75	-1.01	.34	.21	-2.24	.04	.26	1.32	

T_3

	1,2	1.3	1,4	1.5	1,6	2,3	2,4	2,5	2,6	3,4	3,5	3,6	4,5	4,6	5,6
1,3	.89														
1,4	-1.03	-1.70													
1,5	-.22	-1.07	.84												
1,6	-1.31	-1.92	-.32	-1.13											
2,3	1.20	.26	1.94	1.37	2.13										
2,4	-.47	-1.15	.50	-.28	.79	-1.39									
2,5	-.16	-1.11	.93	.08	1.24	-1.46	.36								
2,6	1.02	-1.62	-.07	-.85	.23	-1.82	-.54	-.94							
3,4	-.40	-1.07	.54	-.22	.83	-1.29	.05	-.29	.58						
3,5	.76	-.24	1.63	.95	1.26	-.56	1.05	.93	1.55	.96					
3,6	-.33	-1.01	.51	-.21	.78	-1.21	.04	-.28	.55	-.61	-.90				
4,5	-.22	-.90	.69	-.05	.97	-1.12	.20	-.11	.72	.15	-1.73	-.81	-.98		
4,6	-1.28	-1.84	-.37	-1.12	-.06	-2.03	-.61	-1.21	-.28	-.85	-1.76	-.81	-.98		
5,6	-.83	-1.51	.03	-.72	.35	-1.72	-.41	-.61	.12	-.46	-1.43	-.43	-.60	.40	

N = 24
df = 46

____ = p<o.ol

T_1 = live occasion
T_2 = videotape recording 1
T_3 = videotape recording 2

TABLE 23. Within-coder constancy of coding situation T_1-T_2 . Significance of the differences of coders' mean Scott's Pi coefficients by clusters, t-test.

T_1-T_2

I CLUSTER (speech)

Coders:	1	2	3	4	5
2	.74				
3	.85	.11			
4	-.00	-.94	-1.09		
5	.73	-.14	-.28	1.05	
6	1.06	.23	.11	1.49	.49

N=24
df=46
__=p < .01

II CLUSTER (movement)

Coders:	1	2	3	4	5
2	1.61				
3	.34	-1.21			
4	.49	-1.17	.12		
5	.31	-1.15	-.02	-.13	
6	2.07	.13	1.57	1.59	1.45

N=24
df=46
__=p < .01

III CLUSTER (social form)

Coders:	1	2	3	4	5
2	-.21				
3	.16	.36			
4	-.63	-.35	-.83		
5	.50	.68	.35	1.25	
6	-.76	-.50	-.94	-.23	-1.31

N=24
df=46
__=p < .01

TABLE 24. Within-coder constancy (T_2-T_3). Significance of the differences of coders' mean Scott's Pi coefficients by clusters, t-test.

T_2-T_3

I CLUSTER (speech)

Coders:	1	2	3	4	5
2	-.58				
3	-.26	.33			
4	-.95	-.58	-.77		
5	1.63	<u>2.92</u>	2.14	<u>2.71</u>	
6	-.61	-.21	-.42	.26	-2.10

N=24
df=46
__=p < .01

II CLUSTER (movement)

Coders:	1	2	3	4	5
2	.29				
3	-.6[illegible]	-.99			
4	-1.06	-1.32	-.49		
5	-1.61	-1.87	-1.04	-.51	
6	-1.67	-1.89	-1.18	-.70	-.22

N=24
df=46
__=p < .01

III CLUSTER (social form)

Coders:	1	2	3	4	5
2	-.21				
3	.16	.36			
4	-.63	-.35	-.83		
5	.50	.68	.35	1.25	
6	-.76	-.50	-.94	-.23	-1.31

N=24
df=46
__=p < .01

greatest in Cluster II (.59-.74), and smallest in Cluster I (.69-.78). The range of the standard deviations was similar in the different clusters. In testing the differences between the coders' mean reliability coefficient values (T_2-T_3), statistically significant differences were found in Cluster I between the most constant coder and two other coders (Table 24). It must be noted that the coder who deviated considerably from the others in inter-coder agreement did not differ significantly in within-coder constancy from the rest of the group. However, the within-occasion variation of his reliability coefficients, especially in the verbal cluster, was noticeably high (.17).

The statistical significance of the differences in mean reliability coefficient values for between-coder constancy of coder pairs was not tested because of the large data base. However, the examination of mean values and standard deviations was enough to show that differences did exist, especially in Cluster I and Cluster III (Table 25). The coder pair range of mean coefficient values in the videotaped material (T_2-T_3) was greatest in Cluster III (.37-.77) and second greatest in Cluster I (.41-.70). As was noted earlier, the between-coder agreement level was lowest in Cluster I. These differences between clusters were not equally great when a comparison was made between the reliability coefficients of the live and the videotaped coding situations (T_1-T_2).

It is apparent that the same pairs that were found to differ in the examination of inter-coder agreement differed significantly also in Cluster I and Cluster III and especially on the last coding occasion (T_3). The differences between coder pair reliability coefficients increased in recoding (T_3), and quite unsystematically so in Cluster II and Cluster III. The within-coder differences appeared especially in Cluster I.

Comparisons of Live Observation and Videotape Recording

A comparison of the inter-coder agreement coefficients by occasion show them to be the lowest in the coding of the live situation (T_1) in all clusters when compared with the first videotaped observation (T_2) (Table 26). A statistically significant difference was found between the lowest cluster (Cluster I) and the other clusters. Statistically significant differences were also in evidence between the live and the videotaped observations in Clusters I and II. The greatest change occurred in

Table 25. Analysis by Coder Pairs: Between-coder Constancy. Mean Values and Standard Deviations of Scott's Pi Coefficients by Cluster (I,II,III) and by Occasion (T_1-T_2,T_2-T_3) (N=24)

CODER PAIRS	CLUSTER I				CLUSTER II				CLUSTER III			
	T_1-T_2		T_2-T_3		T_1-T_2		T_2-T_3		T_1-T_2		T_2-T_3	
	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD
1,2	.55	.22	.64	.13	.53	.30	.70	.20	.66	.49	.70	.37
1,3	.60	.13	.59	.18	.48	.38	.71	.18	.68	.50	.75	.32
1,4	.63	.11	.64	.16	.55	.28	.60	.35	.61	.34	.48	.64
1,5	.63	.12	.62	.13	.52	.30	.56	.34	.66	.49	.68	.39
1,6	.53	.17	.41	.20	.63	.28	.44	.45	.62	.46	.46	.71
2,1	.60	.18	.67	.11	.57	.31	.69	.21	.58	.53	.65	.42
2,3	.53	.15	.57	.15	.59	.33	.70	.18	.63	.52	.77	.31
2,4	.58	.17	.57	.19	.64	.26	.63	.37	.59	.37	.48	.67
2,5	.55	.16	.62	.18	.60	.28	.59	.35	.61	.52	.68	.36
2,6	.49	.16	.42	.22	.66	.27	.49	.43	.62	.50	.51	.72
3,1	.55	.21	.70	.09	.50	.29	.69	.22	.61	.48	.73	.43
3,2	.52	.24	.65	.13	.54	.27	.66	.23	.62	.48	.74	.35
3,4	.62	.14	.67	.12	.57	.26	.66	.26	.62	.35	.54	.63
3,5	.66	.13	.70	.10	.55	.31	.64	.27	.63	.47	.74	.34
3,6	.54	.18	.46	.12	.63	.24	.47	.49	.64	.42	.53	.73
4,1	.55	.22	.63	.22	.49	.29	.71	.19	.54	.48	.57	.52
4,2	.56	.22	.63	.18	.56	.32	.70	.17	.54	.49	.60	.50
4,3	.56	.18	.59	.19	.49	.30	.71	.19	.58	.46	.63	.47
4,5	.58	.14	.66	.15	.50	.32	.58	.34	.58	.47	.59	.51
4,6	.52	.18	.52	.18	.52	.29	.47	.44	.58	.40	.37	.79
5,1	.51	.18	.69	.13	.55	.34	.69	.22	.67	.46	.71	.41
5,2	.48	.21	.66	.17	.61	.31	.71	.20	.68	.44	.72	.35
5,3	.62	.16	.67	.13	.53	.35	.71	.19	.71	.44	.81	.26
5,4	.61	.15	.72	.11	.59	.30	.63	.36	.64	.36	.54	.63
5,6	.58	.18	.53	.18	.63	.30	.53	.52	.67	.44	.55	.73
6,1	.38	.22	.49	.20	.48	.31	.64	.29	.50	.56	.60	.46
6,2	.37	.22	.51	.20	.56	.32	.64	.22	.53	.58	.62	.45
6,3	.41	.21	.46	.20	.52	.34	.63	.25	.55	.56	.72	.36
6,4	.46	.12	.48	.19	.57	.31	.54	.68	.54	.49	.52	.56
6,5	.44	.16	.54	.17	.60	.30	.50	.39	.55	.57	.62	.42
N=720	.54	.19	.59	.19	.56	.30	.62	.32	.61	.47	.62	.52

the coding of Cluster II. This may be partially due to the fact that the TV screen reduces and limits the perspective of these activities for all observers and, consequently, the inter-coder agreement was increased. Although the voices were also reduced in the recorded material, the recording may have had a more detrimental effect on visibility than on audibility.

Table 26. Analysis by Occasion: Inter-coder Agreement. Significance of Differences in Means of Scott's Pi Coefficients by Cluster (I,II,III) and by Occasion (T_1,T_2,T_3) (N=360,df=718,p<.01)

CLUSTER	T_1		T_2		T_3		DIFFERENCES		
	X	SD	X	SD	X	SD	T_1-T_2 t	T_1-T_3 t	T_2-T_3 t
Cluster I (Verbal)	.57	.17	.61	.18	.61	.19	<u>-3.06</u>	<u>-2.97</u>	-0.00
Cluster II (Movement & Social Access)	.61	.26	.71	.22	.59	.36	<u>-5.56</u>	0.85	<u>5.39</u>
Cluster III (Social Form)	.75	.28	.77	.36	.60	.59	-0.83	<u>4.35</u>	<u>4.66</u>

When comparing the mean Scott's <u>pi</u> coefficient values of <u>within-coder constancy</u> (Table 27) observed in the live situation and from the videotaped material (T_1-T_2) with the within-coder constancy coefficient mean values of the videotape recordings (T_2-T_3) it was noted that the latter constancy coefficients were higher in all clusters. This difference between the mean coefficient values was also found to be statistically significant in Cluster I. The within-cluster variation in the level of mean coefficient values was in accordance with the previous findings in that the coefficents were highest in Cluster I and lowest in Cluster II. Also the variation of standard deviations between clusters was found to be similar to the within-coder constancy variation in general (T_2-T_3) (.15, .28, .48). Obviously the same factors which influenced cluster variation in within-coder constancy (see Table 20) also influenced variation in between-situation constancy. However, the low level of the reliability coefficients in Cluster II is indicative of the fact that the observers coded the live situation differently than the videotaped one in which some of the 'live' elements were missing due to the nature of the recording. Apparently, the two data collecting methods, direct observation and coding of recorded material, did not always produce the same observations.

Table 27. Analysis by Occasion: Coder Constancy. Significance of Differencesin Means of Scott's Pi Coefficients by Cluster (I,II,III) and by Occasion (T_1-T_2,T_2-T_3) ($p<.01$)

	T_1-T_2 X	T_1-T_2 SD	T_2-T_3 X	T_2-T_3 SD	DIFFERENCES T_1-T_2 and T_2-T_3 t
Within-coder Constancy					
Cluster I	.66	.15	.71	.13	-3.01
Cluster II	.59	.28	.66	.31	-2.00
Cluster III	.62	.48	.69	.47	-1.25
(N=144,df=286)					
Between-coder Constancy					
Cluster I	.54	.18	.59	.19	-5.12
Cluster II	.56	.30	.62	.32	-3.67
Cluster III	.61	.47	.62	.54	-0.38
(N=720,df=1438)					

As before, it appeared that in different coding situations (T_1-T_2) between-coder constancy coefficients were lower than the other reliability coefficient values in all clusters. The variation of mean values between the clusters was noticeable (.54, .56, .61) and similar to the general character of between-coder constancy variation (T_1-T_2). Statistically significant differences were found in the mean Scott's *pi* coefficient values (T_2-T_3) between Cluster I and Cluster II.

An examination of the results indicates that, in spite of the circumstance variation, roughly the same general character of reliability coefficient variation was found within all the three clusters as well as between the clusters. This variation appeared to be most systematic in the coding results of Cluster I, and a result of the structure of the coding system, the observer's way of using it and the quality of the coding target. However, there is reason to assume that the coding situation partially influenced the low level of between-coder constancy coefficients in Cluster II (.56). It was apparent that the observers, when coding the videotaped material, were in fact observing a changed situation in which the 'live' elements were partially obliterated. Thus, the coding was carried out in greater agreement than in the live situation.

Coding Content Constancy

Coding content constancy was defined as the independence of the final results from the constancy of the coding target in different reliability components: inter-coder agreement, within-coder constancy and between-coder constancy.

In this study, the constancy variation was examined for the coding targets of two teachers, three grade levels and four physical education subject areas. The six coders' mean values and standard deviations are presented in the following tables by cluster and by reliability components with the results of the statistical significance test of the differences between the content mean values.

An overview of these results and their comparison with the previously presented general results show that the consistency of the observed phenomenon might have some systematic influence on the variation of the reliability component level in different clusters.

Teacher: When the lessons of two different teachers were the target of observation, the reliability coefficients differed systematically by reliability component and by cluster.

Inter-coder agreement varied from teacher to teacher in all clusters and in all coding occasions. The inter-coder agreement coefficient level varied according to the teacher so that in Cluster III the male teacher's coefficients were higher, but in Cluster I and in Cluster II the situation was reversed. The mean coefficient differences were found to be statistically significant (Table 28).

For within-coder constancy, in the coding of the videotaped material (T_2-T_3), no statistically significant differences were found between teachers (Table 29). However, in the coding of the live situation and the videotaped material (T_1-T_2), statistically significant differences appeared in all clusters. The same variation between teachers that was noted in inter-coder agreement appeared also in this reliability component.

Speech audibility may have varied for the two teachers between the live situation and the videotaped material. Also, the consistency of the observed features of behavior was reflected in the coding differences.

Table 28. Analysis by Content, Teacher: Inter-coder Agreement. Significance of Differences in Means of Scott's Pi Coefficients and ANOVA by Cluster (I,II,III) and by Occasion (T_1,T_2,T_3) (N=180,p<.01)

	TEACHER 1		TEACHER 2		TOTAL		DIFFERENCE df=358	ANOVA df=1/358
	X	SD	X	SD	X	SD	t	F
Occasion T_1								
Cluster I	.53	.17	.61	.16	.57	.17	4.58	21.00
Cluster II	.56	.26	.66	.24	.61	.26	3.59	12.89
Cluster III	.79	.19	.71	.34	.75	.28	-2.73	7.47
Occasion T_2								
Cluster I	.55	.18	.66	.16	.61	.18	5.99	35.89
Cluster II	.67	.25	.75	.16	.71	.22	3.44	11.87
Cluster III	.85	.14	.70	.48	.77	.36	-4.06	16.50
Occasion T_3								
Cluster I	.57	.21	.65	.16	.61	.19	3.93	15.47
Cluster II	.51	.42	.67	.28	.59	.36	4.35	18.91
Cluster III	.62	.60	.58	.58	.60	.59	-.68	.45

On the other hand, the coders might have learned to listen for and observe the reactions of the live target.

The mean coefficient differences in between-coder constancy were highly significant in all clusters, and these differences were greater when the coding circumstances varied (T_1-T_2). The differences in the level of mean coefficient values varied between teachers and by cluster, as in other reliability components, but in this case the variation was even more outstanding.

Consequently, two different teachers (a man and a woman) as the targets of observation seemed to cause systematic differences in reliability coefficients. The levels of within-occasion reliability (agreement) and between-occasion reliability (constancy) differed considerably, and the consistency of the observed behavior was reflected in a systematic way by cluster.

Grade Level: An overview of the mean Scott's *pi* coefficient values and standard deviations in the three clusters indicates systematic variation by grade level. Inter-coder agreement mean values (Table 30)

Table 29. Analysis by Content, Teacher: Coder Constancy. Significance of Differences in Means of Scott's Pi Coefficient and ANOVA by Cluster (I,II,III) and by Occasion (T_1-T_2,T_2-T_3) (p<.01)

	TEACHER 1		TEACHER 2		TOTAL		DIFFERENCE	ANOVA
	X	SD	X	SD	X	SD	t	F
Within-coder Constancy								
T_1-T_2								
Cluster I	.63	.14	.70	.14	.66	.15	3.18	10.12
Cluster II	.50	.31	.68	.22	.59	.28	4.03	16.25
Cluster III	.80	.18	.44	.60	.62	.48	-4.91	24.12
N=72							df=142	df=1/142
T_2-T_3								
Cluster I	.69	.14	.74	.12	.71	.13	2.39	5.72
Cluster II	.62	.35	.71	.26	.66	.31	1.80	3.24
Cluster III	.72	.47	.66	.48	.69	.47	-.78	.61
N=72							df=142	df=1/142
Between-coder Constancy								
T_1-T_2								
Cluster I	.42	.20	.60	.15	.54	.19	8.83	77.98
Cluster II	.46	.33	.65	.24	.56	.30	8.77	76.92
Cluster III	.79	.17	.43	.59	.61	.47	-11.22	125.80
N=360							df=718	df=1/718
T_2-T_3								
Cluster I	.55	.19	.64	.17	.59	.19	6.59	43.48
Cluster II	.57	.36	.67	.26	.62	.32	4.43	19.62
Cluster III	.70	.47	.55	.55	.62	.52	-3.91	15.32
N=360							df=718	df=1/718

Table 30. Analysis by Content, Grade Level: Inter-coder Agreement. Significance of Differences in Means of Scott's Pi Coefficients and ANOVA by Cluster (I,II,III) and by Occasion (T_1, T_2, T_3) (N=120,p<.01)

CLUSTER	LOWER		MIDDLE		UPPER		TOTAL		DIFFERENCES df=238			ANOVA df=2/357
									1-2	1-3	2-3	
	X	SD	X	SD	X	SD	X	SD	t	t	t	F
T_1												
I	.56	.16	.56	.18	.60	.17	.57	.17	.03	2.07	1.90	2.58
II	.58	.23	.58	.30	.67	.22	.61	.26	-.01	3.04	2.59	4.84
III	.69	.27	.78	.31	.79	.25	.75	.28	2.49	3.02	.25	4.96
T_2												
I	.58	.15	.59	.21	.65	.18	.61	.18	.06	3.33	2.76	6.01
II	.68	.18	.66	.29	.79	.13	.71	.22	-.67	5.46	4.51	13.44
III	.70	.44	.80	.28	.82	.33	.77	.36	2.25	2.46	.43	16.50
T_3												
I	.59	.21	.59	.19	.66	.15	.61	.19	-.12	2.82	3.11	5.28
II	.55	.37	.55	.40	.67	.20	.59	.36	-.15	2.78	2.80	4.77
III	.43	.77	.74	.36	.63	.52	.60	.59	4.02	2.53	-1.91	8.98

were noticeably higher in the coding of the upper level than in that of the middle and lower levels. In Cluster III, differences of mean values were noted between the lower and middle levels. The coefficients were again lowest in the live situation (T_1), and highest in the first coding occasion of the videotaped material (T_2).

Statistically significant differences of means of inter-coder agreement values were found between the three grade levels in Cluster I, between the upper level and other levels in Cluster II, and between the lower and the middle levels in Cluster III.

The differences in within-coder constancy (T_2-T_3) between the mean coefficient values of the lower, middle and upper levels were not found to be statistically significant in any cluster (Table 31). However, in the live situation and the first videotaped coding occasion (T_1-T_2), statistically significant differences were found between the lower and middle level mean coefficient values in Cluster I and again between the middle and upper levels in Cluster II.

Statistically significant differences were found in between-coder constancy in the live situation and in both videotaped coding occasions (T_1-T_2 and T_2-T_3). These differences existed between the lower and upper

Table 31. Analysis by Content, Grade Level: Coder Constancy. Significance of Differences in Means of Scott's Pi Coefficient and ANOVA by Cluster (I,II,III) and by Occasion (T_1-T_2, T_2-T_3) ($p<.01$)

CLUSTER	LOWER X	LOWER SD	MIDDLE X	MIDDLE SD	UPPER X	UPPER SD	TOTAL X	TOTAL SD	DIFFERENCES 1-2 t	DIFFERENCES 1-3 t	DIFFERENCES 2-3 t	ANOVA F
Within-coder Constancy												
T_1-T_2												
I	.69	.11	.62	.16	.68	.16	.66	.15	-2.69	-.55	1.78	3.48
II	.60	.20	.51	.34	.67	.21	.59	.28	-1.37	1.56	2.76	4.05
III	.51	.60	.75	.32	.60	.46	.62	.48	2.40	.81	-1.82	3.03
N=48										df=94		df=2/94
T_2-T_3												
I	.71	.13	.72	.13	.71	.14	.71	.13	.47	.18	-.29	.11
II	.61	.34	.66	.30	.72	.28	.66	.31	.69	1.76	1.13	1.61
III	.58	.61	.83	.27	.66	.45	.69	.47	2.54	.69	-2.25	3.49
N=48										df=94		df=2/94
Between-coder Constancy												
T_1-T_2												
I	.55	.16	.51	.22	.57	.16	.54	.19	-2.43	1.41	3.62	7.53
II	.55	.26	.48	.36	.65	.25	.56	.30	-2.46	4.47	6.17	21.39
III	.49	.56	.74	.31	.59	.47	.61	.47	6.15	2.01	-4.32	18.61
N=240										df=478		df=2/717
T_2-T_3												
I	.57	.18	.57	.20	.63	.18	.59	.19	-.01	3.46	3.30	7.45
II	.58	.32	.60	.34	.69	.28	.62	.32	.56	3.92	3.18	8.09
III	.52	.65	.75	.33	.59	.49	.62	.52	4.91	1.33	-4.18	13.01
N=240										df=478		df=2/717

levels as well as between the middle and upper levels in Cluster I and in Cluster II, and between the lower and middle as well as the middle and upper levels in Cluster III.

Subject Area: An overview of the mean Scott's pi coefficient values and standard deviations in the three clusters would appear to indicate systematic variation by subject area.

In Cluster I, the level of inter-coder agreement (Table 32) was lower in gymastics and apparatus than in rhythmic movement-expression and ball games, while in Cluster III the case was exactly the opposite. In Cluster II, the mean coefficient values were higher in apparatus and ball games than in gymnastics and rhythmic movement-expression. Statistically significant differences were found between these subject area mean values in all clusters (I,II,III) and in all coding occasions (T_1, T_2 and T_3), most frequently when ball games and gymnastics were compared with the other subject areas. These differences may be due in part to the constancy variations of the subject area. The differences were reflected in a systematic way, varying according to clusters. Variation, however, was least in Cluster I.

Within-coder constancy was not found to be so sensitive to subject area variation as inter-coder agreement. Statistically significant differences between the mean coefficient values were found in Cluster II between apparatus and gymnastics, and between apparatus and rhythmic movement expression (Table 33).

This was also true of the repeated coding occasions (T_2-T_3), which indicates the difficulty the coders had in interpreting and coding in a consistent manner, movement and social access variation in gymnastics and rhythmic movement-expression. Apparently, variations in activity/passivity and in the degree of pupils' freedom in social activity were smaller and more clearly defined in ball games and apparatus than in gymnastics and rhythmic movement-expression.

The level of the coefficients in apparatus was higher than in other subject areas. The same difference could be noted in the coding of the live situation and the videotaped material (T_1-T_2). In addition, in Cluster III the ball game mean values were found to be much lower than the mean values of other subject areas. These differences appeared to be statistically significant. Some features of the game situation, such as social form, were obscured in the recorded material.

Table 32. Analysis by Content, Subject Area: Inter-coder Agreement. Significance in Means of Scott's Pi Coeficients and ANOVA by Cluster (I, II, III) and by codin Occasion (N = 90, < .01)

CLUSTER	GYMNASTICS		APPARATUS		RHYT.M. EXPRESS		BALL GAMES		TOTAL		DIFFERENCES df=178						ANOVA df=3/356
											1-2	1-3	1-4	2-3	2-4	3-4	
	X	SD	X	SD	X	SD	X	SD	X	SD	t	t	t	t	t	t	F
T_1																	
I	.54	.17	.53	.18	.60	.18	.63	.13	.57	.17	-.33	2.28	3.84	2.56	4.10	1.26	7.13
II	.58	.26	.63	.30	.54	.25	.69	.20	.61	.26	1.23	-1.07	3.41	-2.24	1.73	4.70	6.46
III	.86	.19	.75	.31	.76	.14	.64	.38	.75	.28	-3.04	-4.22	-4.88	.35	-1.98	-2.70	9.76
T_2																	
I	.57	.19	.56	.15	.62	.19	.68	.17	.61	.18	-.41	1.67	4.15	2.24	5.04	2.36	8.91
II	.64	.24	.78	.17	.67	.23	.73	.19	.71	.22	4.63	.96	3.65	-3.63	-.93	2.67	9.43
III	.86	.16	.80	.31	.82	.15	.62	.58	.77	.36	-1.53	-1.77	3.68	.45	-2.53	-3.04	7.93
T_3																	
I	.58	.18	.58	.16	.62	.21	.66	.20	.61	.19	.09	1.46	2.64	1.47	2.73	1.07	3.30
II	.54	.40	.77	.15	.45	.39	.61	.37	.59	.36	5.24	-1.50	1.23	-7.40	-3.92	2.81	14.44
III	.86	.19	.65	.48	.56	.50	.34	.86	.60	.59	-3.91	-5.39	-5.60	-1.24	-2.98	-2.08	13.31

Table 33. Analysis of Content, Subject Area: Coder Constancy. Significance of Differences in Means of Scott´s Pi Coefficients and ANOVA by Cluster (I, II, III) and by Occasion (T_1-T_2, T_2-T_3) ($p < .01$)

CLUSTER	GYMNASTICS X	GYMNASTICS SD	APPARSTUS X	APPARSTUS SD	RHYT.M. EXPRESS. X	RHYT.M. EXPRESS. SD	BALL GAMES X	BALL GAMES SD	TOTAL X	TOTAL SD	DIFFERENCES 1-2 t	1-3 t	1-4 t	2-3 t	2-4 t	3-4 t	ANOVA F
Within-coder Constancy																	
T_1-T_2																	
I	.70	.11	.63	.17	.64	.18	.68	.11	.66	.15	-2.19	-1.72	-1.00	.26	1.40	.90	1.86
II	.53	.29	.68	.27	.52	.25	.65	.30	.59	.28	2.34	-.17	1.72	-2.69	-.52	2.01	3.26
III	.76	.31	.73	.41	.70	.19	.29	.70	.62	.48	-.27	-.98	-3.71	-.48	-3.33	-3.42	9.15
N=36														df=70			df=3/140
T_2-T_3																	
I	.72	.10	.68	.14	.71	.15	.74	.14	.71	.13	1.52	-.27	.76	1.00	1.97	.86	1
II	.57	.27	.79	.18	.56	.36	.72	.34	.66	.31	4.07	-.18	2.00	-3.50	-1.14	1.93	5.30
III	.74	.36	.75	.43	.64	.40	.61	.66	.69	.47	.10	1.11	1.06	-1.12	-1.09	-.25	.88
N=36														df=70			df=3/140
Between-coder Contancy																	
T_1-T_2																	
I	.53	.17	.49	.20	.55	.21	.59	.16	.54	.19	-.63	1.33	4.06	2.72	5.72	2.23	9.80
II	.49	.32	.64	.30	.47	.27	.63	.27	.56	.30	4.67	-.36	4.62	-5.41	-.30	5.43	16.82
III	.72	.31	.73	.36	.69	.18	.29	.70	.61	.47	.25	-1.17	-7.71	-1.31	-7.60	-7.57	44.85
N=180														df=358			df=4/716
T_2-T_3																	
I	.55	.18	.55	.16	.61	.20	.66	.18	.59	.19	-.16	2.80	5.58	3.11	6.12	2.45	14.73
II	.52	.35	.76	.16	.55	.35	.67	.32	.62	.32	8.44	.83	4.31	-7.43	-3.33	3.47	24.30
III	.72	.36	.69	.42	.60	.40	.47	.75	.62	.52	-.59	-2.91	-3.97	-2.12	-3.45	-2.03	8.66
N=180														df=358			df=4/716

Statistically highly significant differences were found in between-coder constancy both in coding occasions (T_1-T_2) and (T_2-T_3) in all clusters. The variation had the same characteristics as the variation in inter-coder agreement, but was even more pronounced.

The coding content constancy varied according to teacher, grade level and subject area, differing by cluster. In all the coding contents, the between occasions reliability (constancy) was higher than the within-occasion reliability (agreement). Thus, it was shown that the lack of reliability does not mean that the majority of classifications occurred by chance. The coders' individual and unique manner of interpreting the situation and using the metalanguage of the coding system might have been the main factors causing disagreement.

Reliabilities of Individual Categories

The inter-coder agreement was assessed for various individual categories of the three clusters of the PEIAC/LH-75 by using the Kendall coefficient of concordance, W (Siegel, 1956). In the statistical processing of the material, the sub-program FORTRAN NMCC was applied. The total percentage of frequencies, summed per category and per observer over a sample of 24 lessons, was ranked separately by the categories of the three clusters and by occasions T_1, T_2 and T_3. A Chi-square test was used for estimating the degree of the statistical significance of the coefficients (Table 34).

The intraclass correlation coefficient was also estimated for each category of observation from the variance between a sample of 24 lesson observations and the variance between the six observers' percentage per category, separately by cluster and by occasions.

The stability estimates were not computed in connection with these indices, but the range of variation of indices between coding occasions gave a rough description of the inter-coder agreement stability by individual categories and by the clusters of PEIAC/LH-75.

As can be seen in Table 34, the inter-coder agreement was rather high; 23 of the categories yielded a value of W statistically significant at the .01 level. Only the indices of one category with low frequencies (I/03), and the categories indicating a confused situation, also occurring infrequently (I/12, II/8 and III/7), were not statis-

Table 34. Inter-coder agreement: Kendalls' W, Interclass correlation and Chi square-test computed separately for the categories of the three clusters of the PEIAC/LH-75 and separately for different coding occasions T_1, T_2 and T_3

Cluster		Categories	T_1 (live situation) N=24 %	W	Intraclass correlation	x^2: df 23	T_2 (video rec. obs. 1.) N=24 %	W	Intraclass correlation	x^2: df 23	T_3 (video rec. obs. 2) N=24 %	W	Intraclass correlation	x^2: df 23
I		Teacher's talk, movement; pupils' talk; other												
	Teacher talk	01. Accepts, praises, encourages	4.5	.81	.78	112.2	3.1	.79	.75	109.4	3.1	.73	.68	100.9
		02. Gives corrective feedback, urges	5.1	.60	.52	82.8	5.6	.64	.56	87.6	4.4	.63	.56	86.9
		03. Uses, develops ideas, movement, tasks, suggested by pupils	0.8	.33[x]	.19	45.1	0.3	.36	.24	50.3	0.4	.38	.25	57.0
		04. Asks, initiates and terminates activity	8.6	.81	.77	112.0	6.7	.88	.86	122.0	7.2	.90	.88	123.9
		05. Presents information, organizes	37.6	.80	.76	110.1	39.6	.83	.79	113.9	42.1	.77	.72	105.9
		06. Gives directions, commands during activity	4.3	.55	.46	75.5	3.8	.71	.65	98.1	3.1	.71	.65	98.1
		07. Criticizes pupils' behaviour	1.3	.74	.69	102.4	0.8	.69	.63	95.1	0.8	.70	.64	96.4
	Pupil talk	08. Answers question/clarifies, demonstrates	0.8	.57	.48	78.7	0.6	.62	.54	85.7	0.8	.67	.60	92.1
		09. Pupil speaks spontaneously, initiates	1.9	.79	.74	108.4	1.7	.72	.67	99.6	1.3	.79	.74	108.5
	Teacher silent	10. Teacher follows pupils' activity, silent guidance	28.1	.86	.84	119.2	30.9	.88	.85	121.5	30.0	.88	.86	121.5
		11. Silent participation in movement activity	6.1	.94	.93	130.1	5.8	.96	.95	132.0	5.8	.96	.95	132.5
	Other	12. Confused situation	1.4	.35	.22	48.5	1.1	.08[x]	-.11	10.9	1.0	.26[x]	.11	35.2
II		Pupils' collective movement activity/passivity and social access												
	Activity	1. Inter-pupil: contacts and movement, space, time, energy restricted; range of ideas controlled	14.8	.88	.85	120.7	11.4	.90	.88	124.4	10.5	.73	.67	100.0
		2. Inter-pupil contacts and/or movement free; range of ideas controlled	37.7	.92	.90	126.2	40.7	.95	.95	131.6	42.3	.87	.84	119.3
		3. Inter-pupil: contacts free: range of ideas open	9.9	.96	.95	131.8	8.1	.93	.91	128.1	8.0	.85	.82	117.8
		4. Pupils' spontaneous activity	0.6	.48	.38	66.8	0.4	.51	.41	70.0	0.3	.37	.25	51.7
	Passivity	5. Pupils follow instruction, demonstrations	25.9	.89	.86	122.2	27.2	.91	.89	125.0	27.9	.89	.87	123.3
		6. Pupils organize themselves, assist in organization	8.9	.73	.67	100.6	10.5	.81	.77	111.4	9.5	.72	.66	98.8
		7. Pupils wait for turn	1.0	.41	.29	56.5	0.6	.50	.40	69.0	0.4	.36	.23	49.8
	Other	8. Confused situation	1.2	.46	.35	63.2	1.1	.16[x]	-.01	21.5	1.1	.37	.24	50.5
III		Social form												
	Situation	1. Complete class, uniform task	31.3	.97	.96	133.5	31.5	.96	.95	132.2	11.9	.96	.95	131.8
		2. Divided class, uniform task	27.3	.97	.96	133.2	28.0	.96	.95	132.2	29.1	.89	.87	122.4
		3. Divided class, differentiated tasks	23.4	.98	.98	135.8	22.6	.95	.94	131.6	22.4	.92	.90	126.0
		4. Divided class, differentiated tasks distributed amongst groups & within group	8.9	.96	.95	132.7	9.0	.97	.97	134.5	8.4	.92	.91	127.0
											6.8	.78	.74	106.2
		5. Individual work, uniform task	7.3	.93	.92	128.4	7.4	.93	.92	128.3				
		6. Individual work, differentiated tasks	0.3	-	-	138.0	0.3	.71	.65	98.0	0.2	.67	.60	92.0
		7. Other situation, confused situation	1.5	.58	.50	80.2	1.3	.63	.56	87.8	1.2	.19[x]	.03	26.1

all coefficients concordant to the level of significance 1 %
x = p > 0.05 significance and beyond.
6 observers, 24 lessons, 4800 time units., tot. 28800 time units

tically significant in all occasions. There was also one category which all observers did not use in the first coding occasion (III/6), and a *W* could not be computed in this case. The significant value of *W* means that the six independent observers were applying essentially the same standards in ranking the sample of 24 lessons by using most of the categories of the system. However, as cited earlier, a significant value of *W* does not mean that the rankings observed are correct. In this special case, because a relevant external criterion does not exist, the ranking of lessons by categories served more or less as an "objective one" (cf. Siegel, 1956).

The level of coefficient values varied between clusters in accordance with the level of overall reliability determined earlier by computing Scott's *pi* (see Table 18). Analyzing the values of videotaped material observation in occasions T_2 and T_3, it was noted that the general level of reliability of the individual categories was highest in the Social Form Cluster III, Md .95 and Md .89, second highest in the Movement and Social Access Cluster II, Md .86 and Md .73 and lowest in the Verbal Cluster I, Md .72 and Md .72. Inter-coder agreement also diminished with time, and most strongly in Cluster II, whereas in Cluster I it remained at the same level in both occasions.

In comparing the *W* values obtained in different situations, it was noted that inter-coder agreement was higher in the live situation than in the videotaped material observation in Cluster I, Md .76-.72 and in Cluster III, Md .96-.95. When the variation of means was tested by Scott's *pi*, the opposite situation was found to be true in Cluster I. It is possible that these differences of *pi* and *W* values reflect the role of chance agreement. As cited earlier, Scott's *pi* describes the average of observer agreement about the proportions of behaviors in the categories, corrected for chance agreement.

It can also be seen in Table 34, that the level of the intraclass correlation coefficient was in general rather high, but lower than the values of the coefficient of concordance, *W*. The variation of the level of this coefficient was also generally in accordance with the variation of *W*, and, in categories occurring frequently, the difference between indices was very small. Intraclass correlation possesses a known sampling distribution and, therefore, it may be given a standard psychometric interpretation. In this case, when the correlation coefficient was computed from mean squares obtained from the six observers' percentages per category by cluster, high values indicate that the variation between observers was small relative to the variation among observations in the sample of 24 lessons used in the study. The intraclass correlations were sensitive to variations of marginal frequencies, which was also noted in analyzing the variance of the means of Scott's pi coefficients for determining the level of the objectivity of coding.

Inter-coder agreement on the frequencies was satisfactory, although category I/03, with low frequencies, and the confused situation categories diminished the level of overall reliability decisively. Thus, it can be concluded that the three dimensional measuring instrument PEIAC/LH-75 was reliable when estimated by using a nonparametric coefficient of concordance, W. However, some revisions are needed. The question of inter-coder agreement is further examined in the following section using discriminant analysis techniques.

Discussion of Overall Reliability Results

In this section the general problems of reliability related to the procedures of categorization are discussed. The coefficients obtained in these analyses can be compared with reliabilities obtained in other studies. According to Flanders (1967b) a Scott's coefficient of .85 or higher is a reasonable level of performance. This statement is based on the analysis of errors of two observers during a four-month period, in which the original 10-category system was used. As a rule, however, in studies using instruments modified and expanded from the Flanders system, coefficients have failed to reach the limits suggested by Flanders (Hough & Ober, 1967). It was also noted by Flanders (1970) that by using a subdivided FIAC system the reliability checks produced inter-coder coefficients between .70 and .86, and during a "difficult" observation, .56.

When using multidimensional observation instruments modified from the Flanders system and constructed for the observation of physical education classes, the acceptable level of performance was lower. Gasson

(1972), in analyzing the verbal and nonverbal behavior of the teacher and pupils and the location of the teacher in relation to the class, noted that an inter-coder coefficient of at least .70 for each of three dimensions would be acceptable. Bookhout (1967), in his multidimensional observation instrument, accepted the level of .40 reliability in selecting variables to be submitted to factor analysis on the basis and stated that the higher the reliability cut-off point set, the fewer variables would be submitted and the greater the risk of throwing away valuable data.

However, Barrett (1971) recommended a level of 90% for determining the objectivity of coding by computing the percentage of inter-coder agreement among trained observers for a multidimensional system developed primarily as a research tool.

In the present study, in which a three-dimensional category system was used, the level of inter-coder agreement was rather low, md .65, varying between clusters as follows: Cluster I, .61; Cluster II, .65; and Cluster III, .69, e.g. in the observation of video-taped material (T_2). The reliability index used here, Scott's <u>pi</u>, took into consideration the estimated role of chance, and was roughly interpreted in this context to indicate the extent to which the codings of the six observers exceeded chance agreement divided by the amount that perfect agreement would exceed chance (cf. Scott, 1955, p. 323). However, chance seemed to have less significance as an error-causing factor than the coders, coding target and coding occasions. The general character of errors was found to be more systematic than random.

As was expected, within-occasion reliability (agreement) (md .61) was lower than between-occasion reliability (constancy) (md .64). In Cluster I, this difference (.61-.71) was found to be systematic. In a comparison of <u>pi</u> values, a wide variance was evident in inter-coder agreement by coder pair (T_1, .45-.65; T_2, .46-.73; and T_3, .43-.72) and still wider in coder consistency by coder pair (T_1-T_2, .37-.68; T_2-T_3, .41-.72). A similar range of variation was not evident in within-coder constancy which ranged between .64-.68 (T_1-T_2) and .69-.78 (T_2-T_3).

The coders' interpretations of the situations and use of the metalanguage of the coding system were unique. Regarding the coding occasions, inter-coder agreement diminished with time (T_2-T_3), except in the verbal cluster where the level of inter-coder agreement remained at the same level. It was apparent that the differences between coders were somehow compensated for in this cluster. The group of observers was heterogeneous with some demonstrating a higher level of agreement with themselves, whereas others agreed more consistently with other observers than with themselves. This kind of change phenomenon was also found by Barrett (1971), and has relevance to observer training as well as to a continuous estimation of reliability and objectivity. The checks of observer agreement carried out at the end of the training period or at given intervals were not enough to avoid systematic errors in coding. However, Komulainen (1970) pointed out in analyzing the overall reliability of an observation instrument modified from the Flanders category system, that "constancy control through time must also be resorted to" (p. 12).

Cluster I was modified and expanded from the Flanders category system and therefore the coding system proposed here uses the same principles of categorization and training procedures. Two of the ground rules given to trainees to increase consistency when choices occurred, need to be discussed here in more detail. First, the rule, "always maximize information by choosing the least frequently occuring category, when there is a choice," and second, "if the observer feels that the pattern at the moment is restrictive, he is cautious in the use of direct categories, but he remains alert to a shift in momentary patterns by remaining alert to the total social situation" (Flanders, 1967b, p. 159).

The results obtained in this study with a modified instrument and six-second time intervals seem to confirm that these ground rules are an invitation to biased observation. However, Flanders (1967b, p. 159) has stated that there is a theory of the "unbiased, biased observer," which contends that even if the observer's assessment appears to be biased, he is unbiased in that he remains open to all evidence of a changing situation. It is evident, too, that the time interval of six-seconds caused additional problems of choice in Cluster I. This error-causing effect was found to be present in the results, judging both by the level of the coefficients in different clusters and by the number of categories in these cluster comparisons. It is advisable to take this into account, however, as Flanders (1970) points out, by choosing time intervals as the unit of analysis: "When such time intervals are small, compared with the cycles or natural units which are of interest, then not too much error is introduced. This approach has the advantage that the observer does not have to make snap judgements about the beginning and end of natural units while he is observing" (p. 164). By using three-second time interval frequencies, as in the FIAC, compared with the six-second interval used in the PEIAC/LH-75, the frequencies are naturally higher and are also reflected in the level of reliability. In the other clusters, the range of variation of frequencies and also the pi coefficients were higher, and the role of random errors greater. The use of categorizing principles merits closer examination in connection with different time intervals.

One factor contributing to the unsystematic variation of reliability components in the Movement and Social Access Cluster II and in the Social Form Cluster III was related to the videotape recording and to the quality of the videotapes used. On several occasions, the video segment was less than adequate, with either teacher or student behavior obscured from view. It may be that the camera angle was not sufficiently thought out with the observation of these activities in mind. In general, the recording was found to have a more detrimental effect on visibility than on audibility. The rules guiding videotape recording and categorizing principles also merit a closer look.

The coding errors caused by the constancy and nature of the coding targets (teacher, grade level, subject area) were rather more systematic than random, and were reflected differently in different clusters. The reactions of "living instrument" to "living target," such as teachers, were clearly visible. When comparing grade level effects and teacher effects on the level of reliability, Tavecchio (1977) noted that the results obtained in a study using the generalizability of scores and profiles for reliability assessment, seemed to confirm the view that the former is "nested within teacher" (p. 95). This was found to be a general characteristic also of the present study because within-coder constancy variation was not statistically significant by grade level in any clusters as it was by teacher. As the coders became acquainted with the coding target, random errors became a systematic way of interpreting teaching behavior individually and uniquely, according to the coding system.

It was also evident that there were common elements and a certain degree of consistency in the interaction pattern in the condition of different P.E. subject areas. This consistency of variation seems to be reflected in the results of this examination, as well as in Cheffers' (1973) study where reliability was determined by submitting cell rankings to analysis using Kendall's coefficient of concordance, W.

One qualification is necessary here, however. There were various kinds of error by individual coders, although no attention was paid to the meaningfulness of errors in this study. An examination of the variance of coders would be a first step toward this kind of study.

Thus, it can be concluded that there was a high degree of consistency both in coding behavior and in the target observation. The results obtained suggest that the theory of the generalizability of scores and profiles presented by Cronbach et al. (1972), in which the question of reliability resolves into a question of the accuracy of generalizability, merits consideration in examining the multidimensional problems of reliability and validity in the construction of measuring instruments for the observation of physical education classes.

The consistency among samples of behavior challenges the investigators to work out concepts of variables to be measured as a part of instrument validation as well as a study of instrument precision (McGaw et al., 1972).

Summary of the Reliability and Objectivity of Coding

The aim of this investigation was to identify and describe the methodological problems involved in an observation instrument proposed for analyzing the directive/nondirective aspects of interaction in physical education teaching (Heinilä, 1976).

The overall reliability was determined by clusters using the scores of six trained observers, each having observed 24 P.E. lessons (20 minutes each) three times, in occasions randomly placed at one-month intervals, first in a live situation and then twice more in videotaped situations. The reliability of the different clusters was assessed by using the profile method, and was computed by using Scott's pi coefficient (Scott, 1955). A total of 8424 Scott's coefficients were computed.

The coefficients were examined by means of t-tests and a one-way analysis of variance (ANOVA). The variation of reliability coefficients was examined by analyzing the between-coder reliability (agreement) and the within-coder reliability (constancy). The contribution made to variation by the different components was analysed by means of a one-way analysis of variance. The reliability of individual categories was also determined by using the nonparametric Kendall coefficient of concordance, W and also by computing intraclass correlation coefficients.

Summarizing the main results, the average level of mean coefficient values was rather low and varied according to cluster (pi I/.61, pi II/.65, pi III/.69) and reliability component (inter-coder agreement .65, within-coder constancy .69, and between-coder constancy .60), indicated in results of the videotape recordings T_2 and T_3. The range of variation and dispersion of coefficients was high.

In Cluster I these "errors" were found to be more or less systematic in character. The reliability index used, Scott's pi coefficient, took into consideration the estimated role of chance in determining the level of reliability. However, in connection with the sample used in this study, chance seemed to have less significance as a reliability decreasing factor than that resulting from the coders, coding occasions and coding target. The chance phenomenon that was found to occur in the use of the categorizing principles of Cluster I judging both the between-coder and within-coder constancy comparisons seems to have relevance both to the development of the structure of the measuring instrument and to the improvement of the general rules guiding the coders and the training of observers.

The reliability operationalized as intercoder agreement and assessed by means of the Kendall coefficient of concordance W, was found to be rather high. Twenty-three of the 27 categories yielded a value of W significant at the .01 level (Chi-square test) but in all coding occasions, the coefficients of four categories of infrequent occurrence (I/03, I/12, II/8 and III/7) were not statistically significant, as was also evident by computing the coefficient of concordance. The categorizing principles need to be considered more closely.

In addition to the assessment of the objectivity of coding, the information concerning the "consensual ordering" of lessons by individual categories may be useful for refining the structure of the instrument and the rules of categorization to facilitate the measurement of theoretically important concepts. It can be concluded, after reviewing the results of this examination of both the overall reliability and the reliability of individual categories of the measuring instrument, that more information is needed about general factors causing errors in coding before the category system can be implemented to objectively measure these concepts.

Construct validity of coding
Background and Purpose

This sub-report will concentrate on the methodological problems associated with the development of the observational instrument and report on an experiment made to examine more closely the construct validity of coding by means of the multiple discriminant analysis technique.

As Dunkin and biddle (1974, 78) stated, when reviewing approximately 500 descriptive studies dealing with the observation of classroom interaction, "the terms reliability and validity have technical meanings when used to describe instruments for measurement of teaching.." and "to say that the instrument is reliable means that it provides the same score of measurement for repeated applications to the same teaching events", and "to say that an instrument is valid means that it measures what we think it is measuring."

In most cases the investigators constructing observational instruments consider only observer agreement and neglect the study of validity. This is common to researchers dealing with observation of physical education teaching and applying modified instruments already validitated, such as the most commonly used Flanders FIAC system (e.g. Dougherty, 1970, 47; Mancuso, 1973c, 84-85; Gasson, 1971, 38). But as we know, an instrument may be reliable without being valid but not vice versa, and thus it is appropriate to concern ourselves also with the crucial question of validity in connection with physical education studies.

Because in observation studies the observer and the classiciation system together form the measuring instrument, reliability is not to be regarded as a propertyt of an instrument but as that of measurement.

Similarly, an instrument itself is neither reliable nor unreliable, it is such only when it has been used to collect data nad data have been manipulated in some way to produce scores. Thus the observer becomes and additional source of errors of measurements. According to Komulainen (1973, 12) "the value of the final results depends crucially on the accurate use of the metalanguage of classification system in the coding process". Therefore the main problem in developing and observational instrument is how to get adequate information for refining the classification system and expecially the rules quiding the observers so that theoretically important concepts could be measured objectively (see Komulainen 1970b, 24).

Because there was no external criterion available to assess the validity of these codings it was decided to use multiple disriminant analysis for examining more closely the variability of coders, i.e. to describe common features of disagreement.

Research task

The purpose of the study was to determine the degree of variability in the codings of different observers when using the categories of the three cluster category system
PEIAC/LH-75 (Heinilä 1976a).

In this connection the aim was:

- to find those discriminant functions that best separate the observers from each other, in other words, maximize the between-observer variance, relative to the within-observer variance,
- to describe factors connected with the use of the category system that cause such differences and thus reduce the degree of agreement among coders,
- to examine the structure of the observer group in terms of the noted deviations and thereby attempt to describe the degree of validity in this "testing".

Discriminant Analysis of the Observational Data

Although discriminant analysis has rarely been used in observational studies, it is appropriate to explore its applicability as a method of assessing and describing factors predicting inter-coder agreement. This method is presented more comprehensively by Cooley and Lohnes (1971). Here the main features, tasks, assumptions and principles of the interpretation of results will be considered.

In this analysis the codings (scores) of different observers may be regarded as forming criterion groups (A-F), representing the universe of observers using the category system (27 categories) in the universe of coding situations (N=144). In discriminant analysis a linear function (II) is made from the predictors (categories of the classification system) so that this function maximally separates the groups (coding of observers). The residuals are treated in the same way. This may result in a new function (I), orthogonal to the former, which improves discrimination of the observer groups. If these functions should prove statistically significant, a curvilinear dependence exists between predictors and criteria. The geometric interpretation of discriminant analysis can be seen for the case of two groups (A and B) and two variates (distributions of two categories X and Y) with the assistance of Figure 13, in which the two sets of concentric ellipses represent the bivariate swarms for the two groups in idealized form. The two variates, X and Y, are moderately positively correlated. Cooley and Lohnes (1971) point out, "that this diagram depends upon the equality of the two group dispersions. If either the variances of X and Y or the X, Y covariance were different for the two groups the centroids for the two groups would not have the same shape and orientation, and the boundary (line II) would not be a straight line. The size of the two populations do not have to be the same, only the dispersion" (p. 245).

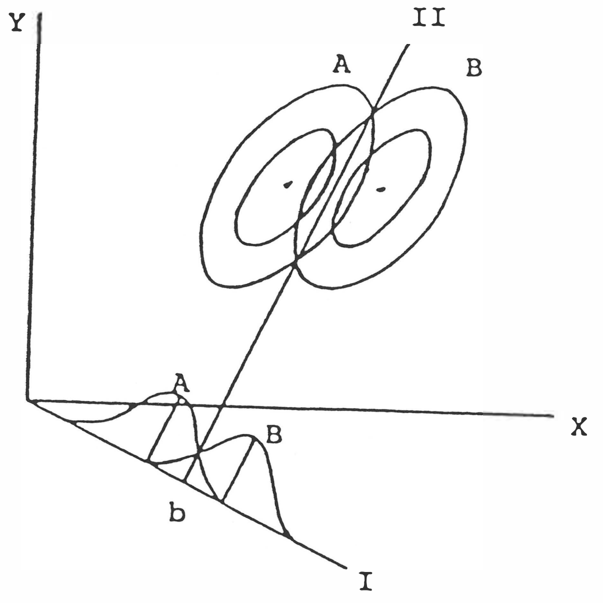


Figure 13. Example of discriminant analysis (Cooley & Lohnes, 1971, p. 245).

This model of analysis also makes it possible to classify observers by using discriminant functions in certain groups according to their scores. If we know, for example, the codings of observers A and B and we wish to place them into certain groups according to the set-up below, the discriminant analysis makes it possible to minimize the proportion of placement of observers into "wrong" cells.

Observer is classified into group A or B

Observer is a member of either group		A	B
	A	right	wrong
	B	wrong	right

The assumptions concerning the level of measurement also need to be considered. The discriminant analysis, like the analysis of variance and factor analysis, presupposes that the measurement fulfills the requirements of interval scales. Nevertheless, these methods of analysis have often been used with ordinal data. For example, Cooley and Lohnes (1971) presented a research example by using such data. Such methods have also been used in observational studies by such researchers as Medley and Mitzel (1958), Soar (1968), Heinilä (1970, 1971, 1980), Bookhout (1967) and Komulainen (1973). However, because of the nature of the measurement

Table 35. Means and standard deviations of six observer scores by using the PEIAC/LH-75 category system in video-recorded material observation 1. (T_2)

Cluster	Categories	A N=24 X	s	B N=24 X	s	C N=24 X	s	D N=24 X	s	E N=24 X	s	F N=24 X	s	Total N=144 X	s
I	Teachers' talk, movement; pupils' talk; other														
Teacher	01. Accepts, praises, encourages	5.46	4.06	4.92	3.53	5.17	4.41	5.63	4.36	5.42	3.75	10.21	5.76	6.13	4.68
	02. Gives corrective feedback, urges	6.33	6.32	12.54	5.80	10.67	10.10	16.58	13.48	11.08	7.73	10.04	8.33	11.21	9.37
	03. Uses, developes ideas, movement, tasks, suggested by pupils	.17	.48	.88	1.15	.21	.72	.25	.44	1.00	1.78	1.25	1.98	.63	1.19
	04. Asks, initiates and . terminates activity	11.00	9.16	12.67	10.45	10.79	8.50	11.83	7.43	13.42	9.84	21.04	13.77	13.56	10.49
	05. Presents information, organizes	79.13	19.37	72.67	19.07	83.92	21.72	79.45	18.03	88.13	22.46	72.25	17.07	79.26	20.19
	06. Gives directions, commands during activity	8.67	9.21	4.71	7.91	6.58	10.91	7.67	10.09	6.54	8.41	11.95	9.94	7.69	9.48
	07. Criticizes pupils behaviour	1.54	2.82	.54	1.28	1.96	2.49	2.38	3.05	1.04	1.52	1.87	2.91	1.56	2.48
Pupil	08. Answers question/clarifies, demonstrates	1.67	1.93	1.08	2.10	.54	1.28	.88	1.36	.92	1.44	2.45	2.67	1.17	1.93
	09. Pupil speaks spontaneously, initiates	1.92	2.70	1.67	2.18	3.58	3.64	3.04	3.54	1.96	2.29	7.79	5.82	3.33	4.11
Teacher	10. Teacher follows pupils' activity, silent guidance	71.75	25.51	73.21	26.33	63.00	27.00	58.00	29.55	57.12	28.24	48.08	25.14	61.68	27.93
	11. Silent participation in movement activity	10.71	14.59	13.08	18.30	11.58	18.07	12.25	18.77	11.38	18.34	10.75	17.03	11.63	17.28
Other	12. Confused situation	2.17	.87	2.04	.20	2.00	0.00	2.04	.46	2.00	0.00	2.29	.91	2.09	.55
II	Pupils' collective movement activity/passivity and social access														
Activity	1. Inter-pupils contacts and movement, space, time, energy restricted; range of ideas controlled	22.88	38.99	22.21	34.52	19.79	31.03	20.46	30.80	23.58	35.34	27.83	33.56	22.79	33.65
	2. Inter-pupil contacts and/or movement free; range of ideas controlle	85.00	50.03	80.96	47.78	87.08	47.10	81.54	49.69	79.83	51.49	73.58	45.64	81.33	48.00
	3. Inter-pupil contacts free; range of ideas open	16.63	33.86	16.12	33.01	13.79	30.90	18.46	34.45	16.29	33.54	15.75	31.53	16.17	32.35
	4. Pupils' spontaneous activity	1.67	3.67	.71	3.07	1.33	3.90	.63	2.86	.71	3.26	1.21	2.99	.96	3.27
Passivity	5. Pupils follow instruction, demonstration	49.63	19.97	56.25	22.89	50.96	22.02	56.25	24.08	56.96	23.97	56.29	23.79	54.39	22.62
	6. Pupils organize themselves, assist in organization	21.83	12.16	20.46	10.16	22.75	11.17	19.71	10.98	19.96	11.47	20.92	10.86	20.94	11.10
	7. Pupils wait for turn	9.92	1.95	1.25	1.94	2.29	2.63	.33	.96	.67	1.66	2.25	2.34	1.28	2.09
Other	8. Confused situation	1.96	.20	2.04	.20	2.00	0.00	2.63	1.24	2.00	0.00	2.17	.56	2.13	.61
III	Social form														
Situation	1. Complete class, uniform task	59.54	57.24	66.04	54.68	62.08	56.14	63.12	55.42	62.63	56.61	64.25	56.09	62.94	55.08
	2. Divided class, uniform task	56.79	69.99	55.45	70.66	57.00	70.60	50.96	64.61	58.04	70.37	57.79	67.75	56.01	67.86
	3. Divided class, differentiated tasks	46.12	57.08	43.46	54.48	45.96	57.43	47.14	56.72	45.25	56.13	43.54	57.90	45.25	55.65
	4. Divided class, differentiated tasks distributed amongst groups' within group	20.50	34.94	16.58	29.98	16.85	30.71	20.04	32.55	17.29	31.32	16.45	29.88	17.96	31.10
	5. Individual work, uniform task	13.83	27.78	15.33	30.86	14.83	29.85	15.58	28.72	14.08	28.34	15.00	30.46	14.78	28.85
	6. Individual work, differentiated tasks	.58	2.86	.46	2.24	.54	2.65	1.04	5.10	.04	.02	.37	1.84	.51	2.84
	7. Other situation, confused situation	2.63	3.06	2.67	3.06	2.71	3.48	2.08	.41	2.67	32.67	2.58	2.86	2.56	2.84

6 observers
24 lessons
4800 6 second time units, tot. 28800 time units

scale, the interpretation of the results remains largely tentative.

The starting point of the discriminant analysis in the present study was the marginal distributions of categories of the 24 lesson data (T_2) (Table 35) as coded by six trained observers, as well as the 27 categories of the three clusters of the classification system. The observation groups were structurally homogeneous and there were differences in the mean distributions of variables. The data fulfilled the requirements set on the number of criterion groups and variables. The use of discriminant analysis was not equally well justified with regard to the level of measurement. This will be taken into account in the interpretation of the data.

Interpretations of the Discriminant Analysis

The results of this analysis are presented in Table 36, in which are listed five discriminant functions, the maximum number possible since there were originally six groups. The table displays the structure and significance of discrimination. The Chi square test, computed from Wilks' lambda, indicated that of the five discriminant functions separating observers, the first two discriminations were statistically highly significant and the third almost significant. It was further established that the power of the discriminant functions to separate observers was great, since their canonical correlations were relatively high. The first discriminant function proved clearly more powerful than the other four with 58%, the second having only 21%, and the third, 11%. From the point of view of interpretation, the first three discriminant functions were the most clear and theoretically important.

The program selected 13 of the 27 classification categories and set them in sequence according to how much they increased the model's discriminating power. It is possible even on this basis to get an idea of the nature of the discrimination. The discrimination model included the nine categories of the Verbal Cluster I and four of the Movement and Social Access Cluster II. Both categories which occurred rarely and those occurring most frequently were represented. In previous studies (Heinilä, 1976), the former categories were found to possess low and the latter high reliability.

Table 36. Discriminant analysis on observers and process variables (PEIAC/LH-75)

Cluster	Categories	Observers A	B	C	D	E	F	Power of discrimination in categories JN	F	n_1	F_{1is}	I Discriminant function s	r	II Discriminant function s	r	III Discriminant function s	r	IV Discriminant function s	r	V Discriminant function s	r
I	01. Accepts, praises, encourages	5.46	4.92	5.17	5.62	5.42	10.21	6	5.07	30	4.84	-.77	-.51	-.28	.10	-.57	-.02	.30	.24	-.43	.31
	02. Gives corrective feedback, urges	6.33	12.54	10.67	16.58	11.03	10.24	5	5.27	25	6.99	-.18	.09	.88	.56	-.52	.07	-.50	.23	-.82	-.19
	03. Uses, develops ideas, movement, tasks suggested by pupils	.17	.08	.21	.25	1.09	1.25	8	4.59	40	2.61	-.17	-.31	.24	-.12	-.75	-.64	-.31	-.18	.37	.13
	04. Asks, initiates & terminates activity	11.00	12.67	10.79	11.83	13.42	21.04	3	6.40	15	16.32	-.99	-.42	.33	-.12	-.73	-.23	-.82	.10	-.22	.21
	05. Presents information, organizes	79.13	72.57	83.92	73.46	88.13	72.25	13	3.48	65	.77	-.44	.20	-.13	-.10	-1.13	.11	-1.19	.47	-.24	.54
	08. Pupil answers question	1.17	1.08	.54	.88	.92	2.46	10	4.01	50	1.62	.29	-.37	-.94	.04	-.36	-.20	.54	.38	-.25	.14
	09. Pupil speaks spontaneously	1.92	1.67	3.58	3.04	1.95	7.79	1	9.98	5	120.18	-1.11	-.68	.37	.15	.13	.28	-.36	.05	-.20	.21
	10. Teacher follows pupils' activity, silent guidance	71.75	73.20	63.00	53.00	57.13	48.03	7	4.83	35	3.51	-1.10	.29	.18	-.24	-1.77	-.09	-.59	.22	-1.47	-.58
	11. Teacher's silent participation in movement activity	10.71	13.08	11.58	12.25	11.38	10 75	12	3.61	60	1.01	-.41	.03	.55	.05	-.98	-.03	-.66	-.03	-.50	-.11
II	2. Inter-pupil contacts & movement free, range of ideas controlled	85.00	80.95	87.08	81.54	79.83	73.58	9	4.28	45	2.04	.28	.09	-.14	-.07	.65	.11	-.54	-.04	.38	-.07
	4. Pupils' spontaneous activity	1.17	.71	1.33	.63	.71.	1.21	11	3.77	55	1.30	.24	-.06	-.48	-.11	.44	.10	29	-.00	-.96	-.01
	7. Pupils wait for turn	.92	1.25	2.29	.33	.67	2.25	4	5.76	20	10.18	-.51	-.37	-.39	-.27	.38	.21	-.45	-.32	-.48	-.42
	8. Confused situation, uproar	1.96	2.04	2.00	2.63	2.00	2.17	2	7.25	10	32.04	.15	.03	.59	.63	.27	.33	.16	.24	.60	.11
								n_2 = 138				56,4 %		21,0 %		10,7 %		7,0 %		5,0 % Of total discrimination	
	Number of observations (144)	(24)	(24)	(24)	(24)	(24)	(24)														

Discriminant function		A	B	C	D	E	F	Discrimination	Test
DF I: Coding of verbal communication wide — versus narrow	m	-6.69	-7.01	-7.41	-6.62	-6.80	-9.59	F ≠ others	x^2_{17}=100.9 xxx, Rc = .73
	d	.64	.80	1.06	.88	.65	1.62		
DF II: Coding of pupils' activity as spontaneous with teachers' silent activity — versus ambiguous with teachers' corrective feedback	m	3.04	4.05	3.40	5.06	3.87	4.05	D ≠ others	x^2_{15} = 46.72 xxx, Rc = .54
	d	.81	.83	.84	1.62	.53	1.02		
DF III: Coding of the sequence of verbal and non-verbal communication by using infrequently occuring — versus frequently occuring categories	m	-9.69	-10.31	-9.16	-9.38	-10.33	-9.86	B and E ≠ C and A, D and F	x^2_{13} =25.74 x, Rc = .42
	d	.67	.94	.97	.89	.96	1.42		
DF IV: Matter-of-fact-centered coding of teacher talk — versus other	m	-8.07	-8.61	-9.15	-8.47	-9.04	-8.50	E and C ≠ others	x^2_{11}=17.42, Rc = .35
	d	.93	.83	.97	1.10	.86	1.25		
DF V: Silence-centered , coding — versus other	m	-5.59	-6.26	-5.90	-5.63	-5.27	-5.62	B ≠ others	x^2_9 = 12.64, Rc = .30
	d	.98	.97	1.06	.95	1.08	.96		

s = variables scaled in w-metrics, r = correlations, m = means on discriminant function, d = deviations on discriminant function

Content and Interpretation of Discriminant Functions

The following principles and sequence were used in the interpretation of the contents of the discrimination dimensions: First, note was made of the variables that had obtained high weights on scaled eigenvectors (s) and of their relative discriminating power. Second, it was ascertained how highly discriminant functions correlated (r) with variables selected into the model. Third, it was established how known groups (observers) were placed on the discriminant dimension on the basis of their means and standard deviations on these dimensions. Finally, their mutual placement in the discrimination plane, formed by two discrimination dimensions at a time, was studied.

From the structure of the coefficients, and the nature of the factors, the five functions extracted appear to measure the following variations in the coding behavior of observers A-F:

DF I: Coding of Teacher-Pupil Verbal Communication: Wide versus Narrow. The first and most important discriminant function distinguished the observers who had made a wide use of the categories of verbal communication from those who had used only some categories. The following categories, besides being highly related to discriminant functions, obtained high weights on scaled eigenvectors: pupil speaks spontaneously (I/09), teacher asks, initiates and terminates activity (I/04), teacher accepts, praises, encourages (I/01). On the basis of the placement of observers on the discrimination dimensions (Figure 14), observer F deviated clearly from the rest, most clearly from observers D and A, and was placed at a distance of over two standard deviations from the others. The observer in question was found to deviate significantly from the others also in the analysis of inter-coder agreement (Heinilä,

Figure 14. Placement of observer A-F group centroids on the discrimination plane formed by discriminative functions I and II.

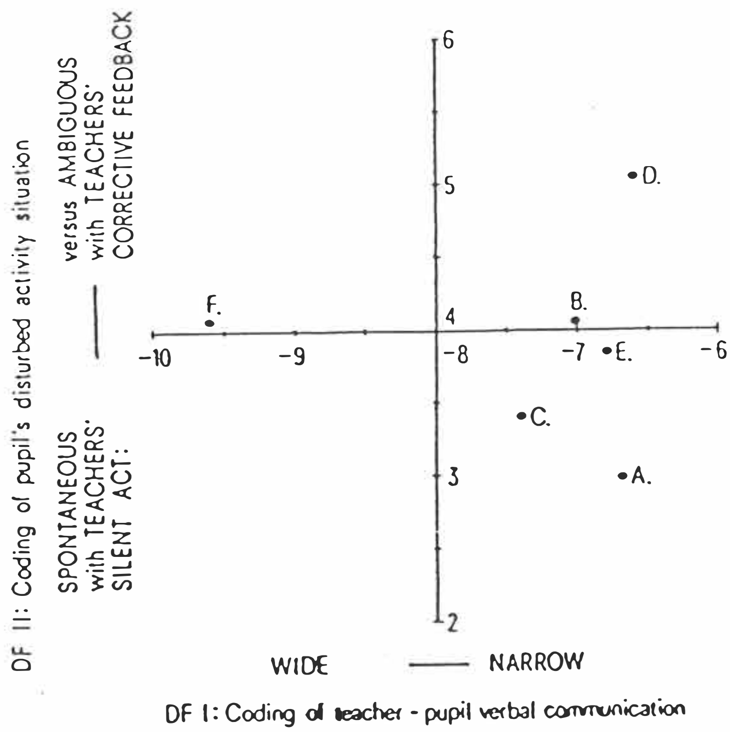
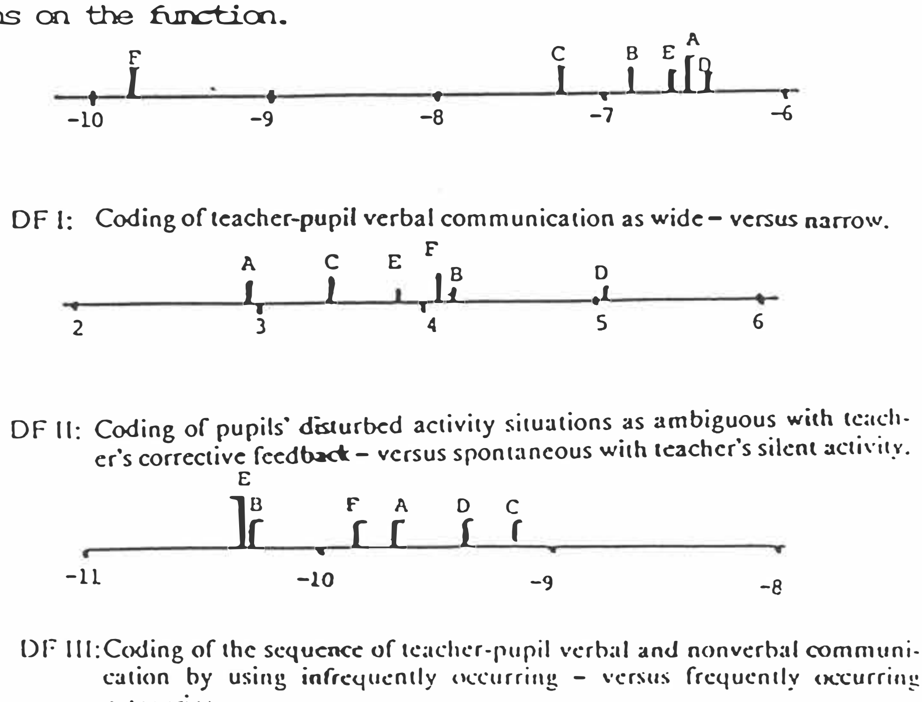


Figure 15. Placement of observers A-F centroids on the discriminant dimensions I, II and III on the basis of their means and standard deviations on the function.



1976). The nature of this factor was then examined closer, as was the shift phenomenon by coder which reduced the index of inter-coder agreement of the whole group. The way in which observer F used the classification system showed a tendency to code more frequently than the others the occurrence of "verbal communication, teacher and pupil initiative and response." The observer in question also attempted to take into account infrequent and more rapidly occurring events in order to describe the continuity of communication, whereas other observers were content with a less detailed coding of communication.

It is possible that the time interval of six seconds was reflected in these coding differences as well as Rule 4 (see Chapter 5, page 83).

DF II: Coding of Pupils' Collective Activity Situations: such as Ambiguous with Teachers Corrective Feedback versus Spontaneous with Teacher's Silent Activity. This discriminant function separated observers on the basis of how they coded ambiguous situations. An examination of the weights of scaled eigenvectors and of correlation coefficients indicates that the most important categories in this discrimination were the category discribing the ambiguity of pupil activity (II/8), teacher's corrective feedback (I/02) and teacher's silent participation in movement activity (I/11). When the placement of observers on the discrimination dimension was analyzed (Figure 14) it was seen that observer D differed clearly from the others, especially from observers A and C. Where observer D tended to code an ambiguous situation using the category "confused situation" (II/8), the rest, and particularly observer A, were more inclined to code it as "spontaneous pupil activity." Similarly, observer D coded the teacher's verbal behavior as "corrective feedback and teacher silent participation" more frequently, while the others used the category "teacher follows pupils' activity, guides silently" (I/10).

It appears that it was difficult to draw a line between confused and spontaneous pupil activity situations.

DF III: Coding of Verbal and Non-verbal Communication: Infrequently Occurring versus Frequently Occurring Categories. This third discrimination dimension was not as easy to interpret as the first two dimensions. It was, however, found to be statistically significant and quite interesting from the point of view of the theory and content validity of coding. The discrimination between observers was again related to coding differences in combining non-verbal and verbal communication. For interpretation, the most important discriminating categories proved to be the verbal category "teacher uses ideas, movement tasks suggested by pupils" (I/03) and the category indicative of teacher initiative "teacher asks questions, initiates and terminates activity" (I/04). Included in the model was the most frequently occurring pupil collective activity category "inter-pupil contacts and movement free, range of ideas controlled" (II/2), whose correlation with the mean of original variables was, however, low (.11). Also included was the category "pupils wait for turn" (II/7). On the basis of the placement of observer centroids on the discrimination dimension (see Figure 15), it was possible to establish that observers B and E deviated from the rest, most clearly from observer C and least from observer F, who, it will be remembered, represented a "wide coding of verbal communication" on the first dimension. Observers B and D tended more frequently than the others to use the categories "teacher initiates and terminates activity" (I/04), "teacher uses ideas, movement tasks suggested by pupils" (I/03) and "teacher participates silently in movement activity" (I/11). Observer C made exceptionally little use of these categories, but a frequent use of the categories "inter-pupil contacts free, range of ideas controlled" (II/2), "pupils wait for their turn" (II/7), and "pupils' spontaneous activity" (II/4). In general, observer C used a more reduced method of coding a combination of verbal and non-verbal communication than observer E. It would seem that combining verbal and non-verbal communication, which is the central feature of this classification system, requires special alertness and a certain attitude. At least half of the observers strived consciously to do so.

While the first three dimensions brought out differences in the coding of infrequent or rapidly occurring categories, confused situations and non-verbal communication, the situation was quite different with the last two dimensions. In them were distinguished coders who used frequently occurring categories in certain ways:

- Matter-of-fact-centered coding of teacher talk - versus other and silence-centred coding - versus other.

The difference between observers was not significant on the last two dimensions, even though it yielded a reasonable interpretation. It should be pointed out that, in general, the use of the most frequently occurring categories, such as I/05 and I/10 in ambiguous situations, is not recommended according to the instructions given in connection with this classification system or with the Flanders category system (see Rule 1, page 83). The discriminant analysis brought out this problem of reliability and construct validity of coding. Also, the shift phenomenon was highlighted in the interpretation of the last two dimensions.

Discussion of Results

Structure of the Observer Group

The discriminant functions that describe independent factors causing disagreement among coders were interesting from the point of view of theory.

Observers could be placed into a certain group which reflected their coding behavior. These discriminant functions were found to be associated with certain kinds of situations such as teacher-pupil verbal communication-centered, disturbed pupil activity situations or nonverbal communication-centered situations. The structure of the discriminant model reflected different coding decisions made in these situations and concerning the choice between infrequently (a) versus frequently (b) occurring categories:

DF I		DF II		DF III	
(a)	(b)	(a)	(b)	(a)	(b)

Obviously the crude coding was advantageous from the point of view of reliability but at the same time it reduced the construct validity, and thus the discriminant validity of the instrument, i.e., of coding. However, the central objective of the classification system was the identification of the sequence of teacher-pupil verbal and non-verbal communication, as well as the discrimination between directiveness and nondirectiveness of the teacher's interaction with pupils.

Naturally, it was more difficult to observe teacher activity in a noisy and confused situation, because audibility was bad. Such situations are not, however, very common in observation studies, but they should be taken into account in analyzing the reactions of different observers and in refining categories and coding instructions. The technical equipment and the methods used for voice recording obviously need to be examined more closely.

The structure of the group of six observers with a similar training background was quite heterogeneous when examined in the light of differences revealed in their individual manner of using the metalanguage of the classification system. Coder differences emerged clearly in three linear factor groups of different composition (see Table 35 and Figures 14 and 15). As is usual in discriminant analysis, the first linear function predictor of disagreement separated one group (observer F) from the rest, then the next one (D) from the rest and so on. Observer variability was great, especially on the first three dimensions and in the discrimination space defined by two discrimination dimensions at a time (Figure 14).

On the basis of the nature of coding decision differences it was possible to get a description of the problems of the construct validity of coding and of its level in connection with the "testing of the instruments." Roughly speaking, about half of the observers approached coding in a way considered valid in terms of the theory, which, however, in this context, often took place at the cost of reliability.

In this case at least, the heterogeneous group offered a good basis for the discrimination of systematic differences, the shift phenomenon and factors that reduced inter-coder agreement. Thus it can be noted that by using a team of observers in the study the universe of generalizability could be broadened. But, in which direction it should be broadened is a question that also merits consideration when the measuring instrument is being refined (McGaw et al., 1972).

Construct Validity of Coding

The empirical findings reported in this study established clearly that high frequencies of occurrence are not necessary prerequisites for the reliable measurement of behavior. Certainly, if a particular type of behavior is of sufficient interest, we should not be deterred from attempting to measure it solely on the grounds that its occurrence is relatively infrequent. Nor, on the other hand, can we assume that the accumulation of large numbers of observations of a particular type of behavior provides some kind of guarantee that we have achieved precision of measurement.

What really matters then is not the number of times that a particular type of behavior has been observed, but whether the subjects of the observation have differed consistently in the extent to which they display that behavior. This cannot be inferred from considerations of frequency alone, but need to be determined by an analysis of inter-coder agreement and between-coder agreement of the type described earlier or those reported by Heinilä (1976b), Bockhout (1967) or Komulainen (1973).

The construct of discriminators found in this study describe patterns of teacher and pupil behavior which in Bockhout's (1967) study were found to be related to the social emotional climate. The quantity of positive emotive expressions of teacher and pupil talk (DF I), and the sequence of verbal and non-verbal interaction (DF II) also distinguished situations where teacher and pupils were moving and teacher was participating in movement activity (DF III), causing disagreement among coders. Decisions concerning the level of different forms of pupils' collective activity, operationalized as social access, were also reflected in results describing variation between coders. Also the results obtained by Tavecchio (1977) suggest similar difficulties in coding interaction processes in physical education classes objectively.

In the present study, the inverse character of reliability and validity was highlighted, which had already been pointed out by Flanders (1970) in his analysis concerning the training of observers and reliability problems.

Suggestions for Further Study and Improvement of the Observation Instrument PEIAC/LH-75

Although the results of the discriminant analysis can only be regarded as tentative on account of the nature of the level of the measurement scale, they yielded quite useful information for the development of the instrument.

Further research should be conducted with the instrument created. It should include replications of the exploratory study. Further as the observation instrument is still crude, its potential for refinement should be capitalized upon for research and teacher training purposes. The results obtained suggest that more attention should be paid to the following questions:

(1) development of rules for coding more decisively teacher-pupil verbal and non-verbal communication and their sequence.

(2) the use of the six-second time interval needs to be considered more closely. May be a three second interval would be better in coding the first speech cluster, and one minute in connection with the other clusters,

(3) developing rules for coding more decisively pupils' collective movement activity and the forms of social access (categories II/3, II/4),

(4) also the rules for coding "pupils collective passivity" (II/7) "waiting for turn" must be refined,

(5) the rules quiding videotape recording must be determined moree exactly and so that the whole situation is taken into account,

(6) the technique for voice-recording must be implemented by using the wireless throat microphones. Recording of pupils voices also needs to be considered more closely,

(7) the training of coders as well as the treatment of material to be used in this connection need to be examined. A sample of material with different contents should be added to the observer training programme,

(8) agreement controls carried out only at the beginning of the study period are not enough to avoid systematic errors in coding: recurring constancy control must be resorted to,

(9) the selection of coders need to be considered more closely and by taking into account the cognitive as well as affective characteristics of the candidates.

The technique of multiple discriminant analysis outlined previously can be applied

- for refining the classification system so that reliability and validity problems can be examined simultaneously
- for implementation of observer training programmes so that aspects important from the theoretical point of view can be emphasized
- for studying and assessing "inter-investigation agreement" by having the same classification system used by different investigators.

PART III

INVESTIGATION OF THE CONSTRUCT VALIDITY AND SENSITIVITY OF THE OBSERVATION INSTRUMENT PEIAC/LH-75

This section will report on an investigation of the construct validity and sensitivity of the observation instrument PEIAC/LH-75 using a cumulative multivariate analysis of the factoral structure of instructional situations and a discriminant analysis of the groups formed with the factor scores.

The general principles underlying factor analysis and its various phases are well known. Only some special problems will be considered in this connection, after which the specific areas of multiple grouping analysis and multiple discriminant analysis will be discussed.

A great deal of correlational research on validity employs factor analysis which reorganizes a table of correlations to emphasize convergence. Reducing the central core of this information to a compact table of factor loadings often has a clarifying effect (cf. Cronbach, 1971; Medley, 1982).

In this connection, an attempt will be made (a) to use factor analysis as a means of reducing the dimensionality of the set of three cluster variables by taking advantage of their intercorrelations, and (b) to find ways of identifying fundamentally meaningful dimensions of the multivariate construct under study. This kind of evaluative research may be termed a method of controlled correlation to highlight the central roles of correlation coefficients as a primary index of the strength of relation, explanation, or prediction. Regarding kinds of possible conclusions, they will be probablistic in nature, reducing uncertainty, but not completly eliminating it (cf. Cooley & Lohnes, 1976).

In this study the greatest interest centers on correlations between the original variables and factors. The matrix of scores of the categories of the three-cluster correlations is called a factor structure. This matrix will be used here primarily as an interpretative device, just as it is in any multivariate analysis which results in a factoring of a measurement battery. Here the same factor matrix is regarded as expressing both the theoretical composition of a measurement, thus "explaining" the measurement, and the correlations of the factor with the measurement "explaining" the factor (Cooley & Lohnes, 1971).

When working with ipsative nominal scales, it is necessary to interpret the two poles of each factor separately. This situation is in general attributable to the use of taxonomies. "As the system is always in some state, an increase in any one form of behaviour leads to a decrease in the other forms" (Komulainen, 1971a, p. 16).

By using a three-cluster category system the variables are tied to ipsativity in more than one way: between the categories within each cluster and between the categories of different clusters. Thus, we can discuss inter-cluster ipsativity and between-cluster ipsativity. A factor analysis will be employed in this context as a means of exploring ipsativeness on the construct under study.

The set of data analyzed here was recorded on videotape during the autumn term of 1973. The data were gathered by six trained observers coding each situation three times: first in the live classroom and then twice more with the videotape at one month intervals. The data set includes 24 P.E. lessons with a total of 28,800 six-second time units.

Aims of the Factor Analysis

This analysis will explore, from the point of view of the validity of Flanders' theory, the interaction in 24 P.E. lessons by considering the systematic variance among scores when using the PEIAC/LH-75 three-cluster category system on the construct under investigation.

In this phase of the study, the aims were:

1. to examine interaction in physical education classes by means of the factor analytical r-technique
   - to identify the structural dimensions of interaction,
   - to consider whether they correspond to logical dimensions or to the theoretical framework, and
   - to consider the behavior of the emerging factors (factor scores) in combination with certain other variables (frame factors) as classified in accordance with the sex of the teacher, grade level and physical education subject area;

2. to explore the formation of homogenous groups of lessons in grouping analysis based on factor scores; and
3. to explore the formation of the factors predicting variability and grouping of lessons, "known" to be different. In this connection the aim was:
   - to find those discriminant functions that best separate the criterion groups from each other, in other words to maximize the between-group variance, relative to the within-group variance;
   - to describe factors connected with the use of the category system and predicting the grouping of lessons, and thus to describe the ability of the instrument to distinguish between groups known to behave differently on the construct under study;
   - to examine the structure of the groups formed by the grouping analysis in terms of noted deviations; and
   - to describe the sensitivity of the instrument, i.e., the ability of the instrument to make the discriminations required for research problems.

Procedures

Selection of Variables

The establishment of a minimum acceptable reliability for variables to be submitted to factor analysis was based on the following principles: Since there were no previous studies using this observation instrument, reliability of the data could not be presumed. The higher the reliability cut-off point set, the fewer variables would be submitted, and the greater the risk of throwing away valuable data. On the other hand, the lower the cut-off point, the greater the risk of diluting the factor analysis with so much worthless data that a great many poorly defined factors would be required to account for total variance. For this quasi experiment, the intention was to submit to factor analysis those variables which might contribute significant loadings to factors. Estimating reliability by using the Kendall coefficient of concordance (W), 23 of the 27 categories were significant at the 0.01 level. The remaining four were categories with low frequencies and/or indicating a confused situation (I/03, I/12, II/8, III/7). In the light of this criterion, a total of 27 variables were submitted for analysis. The results reported here are based on a video-recorded observation (T_2), in which the level of reliability was the highest of the three rating times. The means of Scotts pi, computed from the scores of the six trained observers, were, by clusters, .61, .71 and .77.

Factoring and Principles of Interpretation

The intercorrelation matrix was obtained by correlating the three-cluster category frequencies 27 x 27 computed from the six observers' scores (total 28,800 six-second time units) in the lessons (N=24). The data from three coding occasions (Appendix C.1) were subjected to factoring separately. The correlation matrices were factored by using the principal axis method, and the numerically highest correlations were used as estimates of h. Rotation was carried out by the varimax technique. This rotation method was chosen because, being octagonal, it was likely to yield a simple and clear-cut result useful at the initial stage of this "structure seeking" investigation.

The number of factors to be rotated was determined according to the principle that (1) it is preferable to include too many than too few factors, and (2) a description that is optimal both interpretationally and in terms of the simple structure rule should be sought with successive reductions of the primary base. Four, five, six, seven and eight factors were rotated with the varimax technique.

Seven factors proved to be the most interpretable and stable combination. The consistency of the structures of the seven factor varimax resolution was examined by analysing the factor structure computed from three data sets (coding occasions T_1, T_2, T_3) by means of Symmetric Transformation Analysis (Appendix C.2). Each factor extracted was interpreted as a structural dimension by studying the categories with appreciable loadings (-.30), synthesizing them, and naming the composite pattern.

The factor loadings of categories and the regression coefficients obtained by them (Appendix C.3) in the estimation of factor scores helped to identify the categories that were central in the construct of the factor in question. In addition, the lessons for which the factor scores were the highest were compared with those with the lowest factor scores.

Results of the Factor Analysis

Correlations between Categories of the Three Clusters

The correlation matrices between categories of different clusters (Table 37) express the interdependence of the categories of each cluster throughout the lessons observed. The figures are in general so low that categories may be considered sufficiently independent of each other to meet the requirements of independence imposed on observational methods. Using ipsative nominal scales, it is evident that there will be some high negative correlations, and as stated before, the process is always in some state. Therefore an increase in any one form of behavior leads necessarily to a decrease in the other forms. For instance, in the verbal cluster (I) the category indicating teacher's silent behavior (I/10) and the category indicating the most dominant teacher's verbal behavior (I/05) correlated negatively. Also it is understandable that there will be positive correlations between the categories of initiative behaviors and response behaviors. Categories of different clusters correlated with each other both positively and negatively. The highest positive correlation, .98-.97, was found between categories II/8 and III/7 of clusters two and three, both indicating a confused situation. These categories were always used together in beginning and finishing coding.

Results

Factor analysis yielded seven factors accounting for 68.6% of the total variance (Table 38). Factor scores estimated for every lesson in the seven factors are presented in Table 39. The results are illustrated in Figure 16, based on the means and dispersions of factor scores and demonstrating the location of each lesson in structural dimensions as classified according to the sex of the teacher, the grade level, and the physical education subject area.

It was found that the positive pole activities consisted mostly of the teacher's verbal activities. However, in the first factor a type of non-verbal form of teacher activity, participation in student activity (I/11), was evident (Figure 16A). The teacher's silent behavior as guidance (I/10), which is a common type of activity in ball games, was characteristic of the negative-pole activities. Two factors, IV and V,

Table 37. Categories of the Three Clusters on Correlation Matrix for Observation T_2. The Highest Correlation Coefficient of Each Variable is Placed on the Diagonal.

VARIABLES CLUS. No.	CAT. No.	I 01	02	03	04	05	06	07	08	09	10	11	12	II 1	2	3	4	5	6	7	8	III 1	2	3	4	5	6	7
I	01	55																										
	02	55	66																									
	03	-02	05	51																								
	04	-17	-35	-14	54																							
	05	10	32	-02	-19	61																						
	06	02	-10	09	54	-02	72																					
	07	-00	-02	32	06	29	03	52																				
	08	-04	-18	51	45	20	31	52	58																			
	09	-26	18	50	-23	48	05	43	49	58																		
	10	-24	-19	-09	-31	-60	-47	-16	-36	-35	-60																	
	11	-15	-37	-15	01	-31	-08	-34	-26	-30	-21	77																
	12	00	-08	-01	09	-02	12	04	06	-05	-04	-00	39															
II	1	-27	-25	-21	54	-11	72	-21	09	-22	-24	03	30	72														
	2	46	33	13	-44	-17	-51	07	-23	14	54	-38	-19	-65	-65													
	3	-21	-26	-06	-12	-25	-10	-04	-17	-29	-11	77	-08	-18	-41	77												
	4	06	22	17	-10	07	-12	04	-06	54	-14	06	05	-19	13	-03	98											
	5	-20	05	09	41	61	25	03	46	08	-57	-18	01	26	-55	-20	-16	61										
	6	-13	-02	04	-19	58	-14	38	31	58	-08	-45	08	-17	-02	-31	19	25	58									
	7	-44	-39	-08	14	-07	17	08	09	-19	-25	56	21	19	-57	59	-10	07	-06	59								
	8	21	-05	15	13	17	-14	50	58	47	-19	-12	05	-13	01	-15	19	24	26	-02	58							
III	1	07	07	15	36	-14	47	-12	-24	03	-32	18	-18	41	-16	-03	15	-08	-34	-08	-23	-59						
	2	-39	-42	-18	18	21	-12	13	32	08	11	-20	39	18	-20	-19	-21	34	27	07	22	-59	-59					
	3	41	66	06	-36	22	-10	04	02	20	-07	-27	-19	-41	34	-18	10	06	18	-17	-11	-19	-47	66				
	4	30	-06	-23	-23	-35	-21	-10	05	-10	38	03	-08	-12	25	08	-10	-38	-16	-15	25	-26	-06	-18	-38			
	5	-33	-38	19	-20	-29	-22	-09	-24	-10	11	62	-12	-25	-15	76	06	-35	-20	48	-13	13	-24	-31	-08	76		
	6	-17	03	51	40	-07	01	19	42	-06	04	-15	-08	-15	-02	04	-07	22	-00	16	-11	-08	-05	22	-13	-10	51	
	7	10	28	12	-11	10	-07	-00	-05	52	-16	00	07	-15	12	-10	98	-10	18	-15	17	09	-18	15	-05	-03	-05	98

Table 38. Varimax-rotated Factor Matrix.

Cluster	Cat.	1	2	3	4	5	6	7	h^2
I	01.	29	-03	-64	28	01	-13	-32	68
	02.	37	11	-66	-10	16	24	-12	69
	03.	-01	-07	-13	35	18	-06	65	60
	04.	04	-68	29	11	-13	02	29	67
	05.	18	14	-09	23	08	79	-11	75
	06.	06	-82	01	04	-05	16	05	70
	07.	01	12	04	61	03	12	27	47
	08.	12	-20	16	72	-10	23	36	79
	09.	17	10	-17	58	51	25	10	73
	10.	20	48	23	-29	-11	-62	08	81
	11.	-84	-13	00	16	03	-13	-21	81
	12.	06	-10	36	05	06	07	-15	17
II	1.	09	-76	40	18	-10	10	-17	82
	2.	44	51	-37	02	10	-47	05	82
	3.	-90	08	-05	-06	-09	-06	-02	83
	4.	-03	04	-06	08	98	00	-00	97
	5.	13	-25	16	16	-17	71	12	68
	6.	25	35	21	30	21	47	06	58
	7.	-67	-11	26	-00	-11	21	08	59
	8.	07	05	10	78	10	02	-13	65
III	1.	-10	-66	-29	-22	-22	-13	-07	64
	2.	18	16	75	17	-20	20	-09	74
	3.	25	28	-64	-10	04	30	15	67
	4.	05	17	-03	19	-14	-49	-33	44
	5.	-82	18	03	-06	09	-23	16	79
	6.	03	-04	-05	06	-13	-08	71	53
	7.	08	02	-07	05	95	05	-05	92
Eigenvalue		3.4	3.1	2.7	2.5	2.5	2.7	1.7	18.5
%		12.5	11.5	10.0	9.3	9.2	10.0	6.2	68.6

Table 39. Estimated Factor Scores.

Lesson No	Rec.date	Teacher	Grade level	Subject area	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
1	12.11.73	W	L	A	463	413	538	515	467	703	495
2	15.11.73	M	M	A	432	436	626	458	471	483	502
3	17.12.73	W	L	G	490	713	543	633	470	550	538
4	20.11.73	W	M	G	452	788	456	378	493	457	438
5	20.11.73	W	H	B	451	386	315	413	481	451	537
6	22.11.73	M	M	R	629	395	468	482	511	565	564
7	18.11.73	M	H	A	449	443	676	461	459	491	436
8	27.11.73	W	M	B	421	470	329	425	482	506	512
9	27.11.73	W	H	G	487	705	472	390	485	489	464
10	27.11.73	M	H	R	775	472	560	493	469	399	413
11	27.11.73	M	H	B	400	389	500	451	472	256	484
12	29.11.73	M	M	B	432	444	454	603	465	379	387
13	3.12.73	W	L	B	457	496	313	472	484	524	433
14	4.12.73	W	M	R	687	512	496	515	543	407	531
15	4.12.73	W	H	R	736	502	422	424	476	596	449
16	5.12.73	M	L	B	484	465	396	483	490	681	463
17	10.12.73	W	L	R	490	523	525	530	465	543	875
18	11.12.73	W	L	A	446	465	627	484	464	594	522
19	12.12.73	M	L	R	524	454	519	506	516	331	678
20	13.12.73	M	M	G	470	550	594	531	498	434	474
21	14.12.73	M	L	A	465	489	531	528	961	529	473
22	17.12.73	W	M	A	455	458	669	438	458	577	432
23	17.12.73	M	L	G	479	514	420	892	456	527	478
24	18.12.73	M	H	G	471	518	549	495	464	526	371

Teacher	Grade level	Subject area
M = man	L = low	G = gymnastics
W = woman	M = middle	A = apparatus
	H = high	R = rhytmic
		B = ball games

had high loadings only in the positive pole. The dispersion of factor scores was highest in the third factor 313-676.

The factors obtained are shown below. The first factor was clear-cut in content. Here, all the most important loadings were negative. The

Factor I:	Cluster/ Category	Positive pole	Cluster/ Category	Negative pole
	I/02	+.37	II/3	-.90
	II/2	+.44	I/11	-.84
			III/5	-.82
			II/7	-.67

loadings were spuriously high. The social access and the social form cluster categories (Cluster II and Cluster III) had high loadings on this structural dimension as well as teacher's silent participation in movement activities in situations where interpupil contacts were free and range of ideas open, work divided among groups or individuals. The positive pole activities consisted of the teacher's verbal positive reactions and corrective feedback to the pupils' activities. Comparing the different lessons by considering the factor scores estimated for them, the lesson of rhythmic movement expression showed the highest loadings in this factor. These variables are descriptive of the entire indirect influence area. This structural dimension was labelled "indirect nonverbal integrative idea generation -- teacher's verbal communication and motivation."

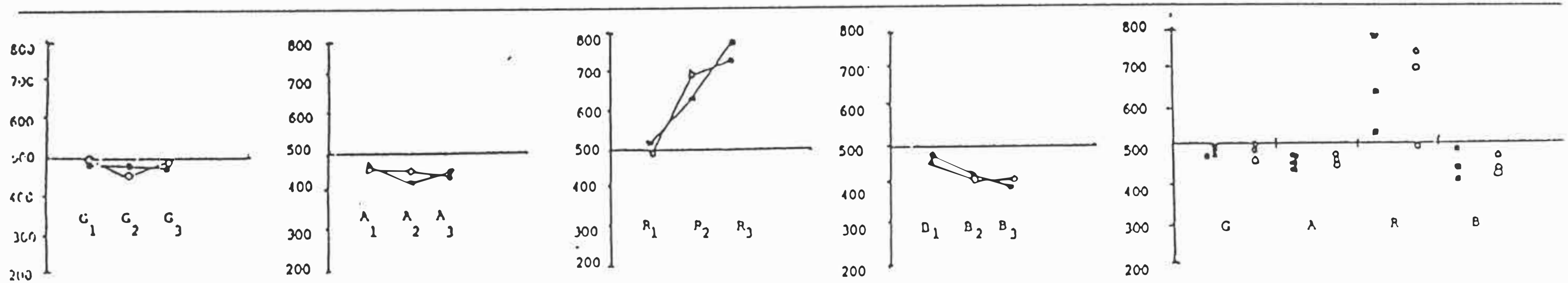
The content of Factor II also was clear. The negative pole concerned the teacher's verbal direct communication and its intensity in

Factor II:	Cluster/ Category	Positive pole	Cluster/ Category	Negative pole
	II/2	+.51	I/06	-.82
	I/10	+.48	II/1	-.76
	II/6	+.35	I/04	-.68

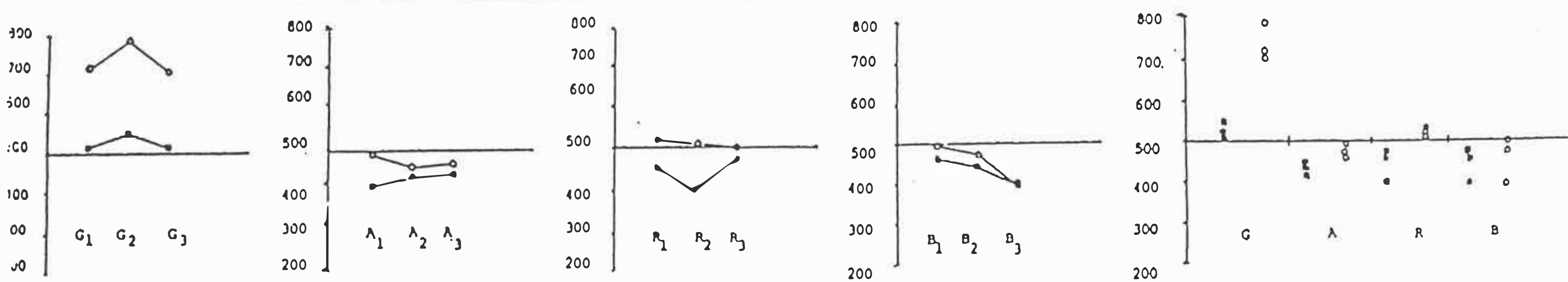
the situation in which inter-pupil contacts and movement activities were restricted and range of ideas controlled. The positive pole was associated with situations in which the teacher's silent guidance was predominant and in which inter-pupil contacts were free but the range of ideas was still controlled. All the woman teacher's gymnastic lessons

Figure 16. Location of each lesson in structural dimensions based on the means and disperions of factor scores

A. Factor I. IDEA GENERATION: Teacher's and pupil's non-verbal integrative idea generation (+)/Teacher's verbal idea generation and motivation (-)



B. Factor II. INTENSITY: Teacher's total, intensive guidance (+)/Teacher supervision and organization (-)



G = Gymnastics
A = Apparatur
R = Rhythmic movement-expression
B = Ball games

1 = lower level
2 = middle level
3 = upper level

• = Teacher 1 (man)
o = Teacher 2 (woman)

showed high loadings on this factor (Figure 16B). The structural dimension was descriptive of the entire direct influence area. It was labelled "intensity of teacher's verbal direct guidance."

Factor III consisted of categories from all three clusters. In the positive pole the highest loading was related to situations where the class was divided by uniform task, and the second highest variable

Factor III:	Cluster/ Category	Positive pole	Cluster/ Category	Negative pole
	III/2	+.75	I/02	-.66
	II/1	+.40	I/01	-.64
			III/3	-.37
			II/2	-.37

loading described the social situation in which inter-pupil contacts and movement activities were restricted and the range of ideas controlled. The dominating characteristics of the negative pole were the teacher's positive verbal reactions to pupil activities, specificity of supportive supervision in the situation in which the class was divided, the tasks differentiated, and the range of ideas controlled. In this factor, the apparatus and gymnastics lessons, especially of the male teacher, showed high loadings (Figure 16C). This structural dimension was labelled "uniformity of teacher's nonverbal guidance -- specificity of verbal supportive supervision." These aspects are descriptive of the entire direct/ indirect influence area.

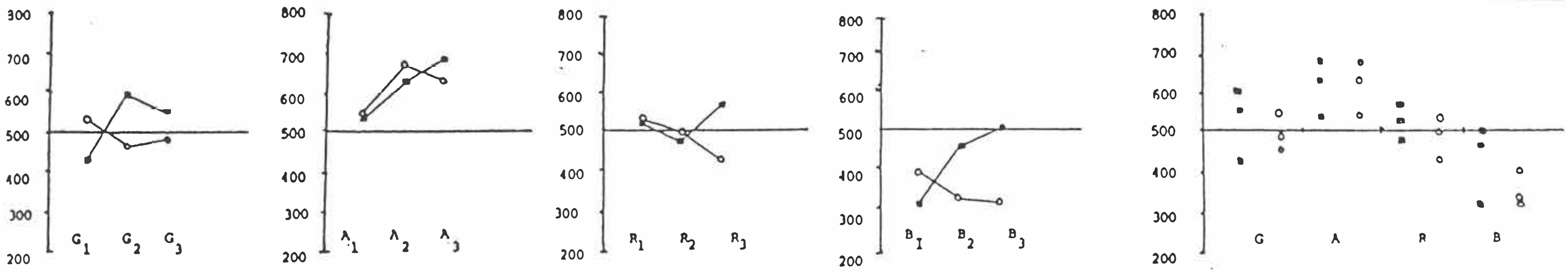
In Factor IV, all the most important loadings were positive. The fourth factor was related to confused situations where the dominant

Factor IV:	Cluster/ Category	Positive pole	Negative pole
	II/8	+.78	-
	I/01	+.72	
	I/07	+.61	
	I/09	+.58	
	I/03	+.35	
	I/06	+.30	

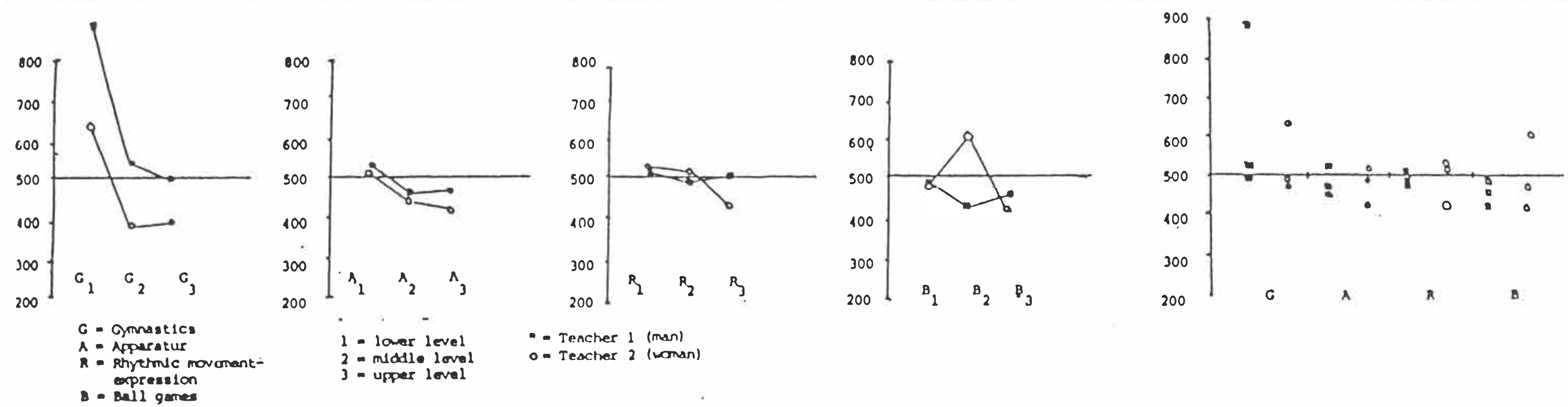
characteristic was pupil-teacher verbal communication, which consisted particularly of pupils' suggestive activity. The dimension was typified by the high loading of teacher's acceptance of pupils' spontaneous

Figure 16 (cont.)

C. Factor III. SPECIFICITY-UNIFORMITY OF GUIDANCE: Specificity of supportive supervision (+)/Uniformity of teacher guidance (-)



D. Factor IV. DIRECTING COMMUNICATION (+)/ACTIVITY (-)



activity as well as by the loading of teacher's criticism. One low level gymnastic lesson in particular had high loadings on this factor (Figure 16D). This dimension was labelled "directing communication."

Factor V was typified as non-structured situations in which the social form as well as social access were unclear. In this context pupils were asking for instructions and expressing their own ideas. Only one low level apparatus lesson had exceptionally high loadings of this factor (Figure 16E). The dimension was labelled "non-structured spontaneous pupil activity."

Factor V:	Cluster/ Category	Positive pole	Negative pole
	II/4	+.98	-
	III/7	+.95	
	I/09	+.51	

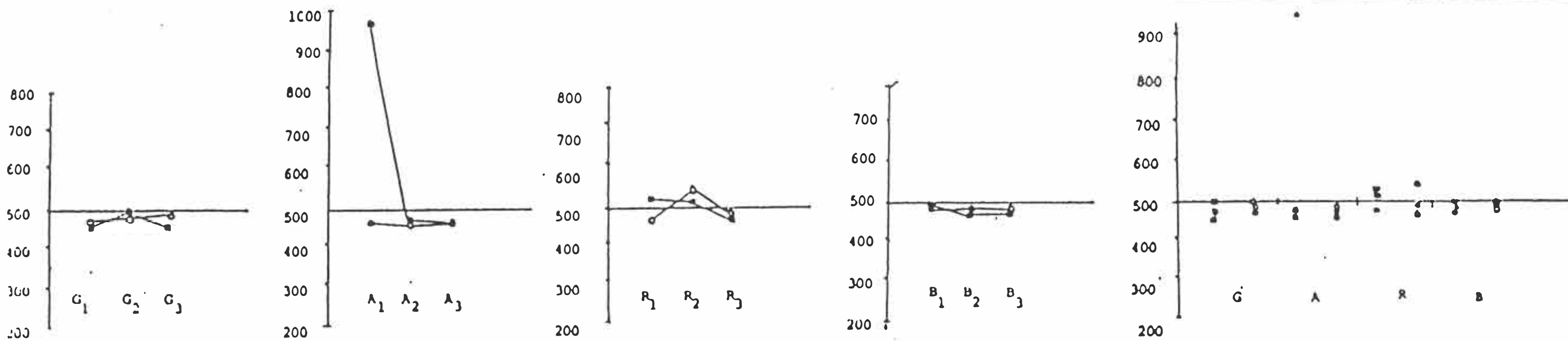
The dominant characteristics of the sixth structural dimension were phases of the lesson as orientation and work typified by verbal/nonverbal interaction. The positive pole mainly concerned the teacher's presentation of information, pupils following instructions, organizing themselves and assisting in organization. The negative pole was associated with activity situations in which the class was divided, tasks distributed among groups and within groups, the range of ideas controlled and silent guidance predominated. The female teacher's apparatus and rhythmic movement expression lessons had high loadings on this factor (Figure 16F). The structural dimension can be named "teacher-dominant verbal subject centricity -- non-verbal groupwork centricity."

Factor VI:	Cluster/ Category	Positive pole	Cluster/ Category	Negative pole
	I/05	+.79	I/10	-.62
	II/5	+.71	III/4	-.49
	II/6	+.47	II/2	-.47
	III/3			

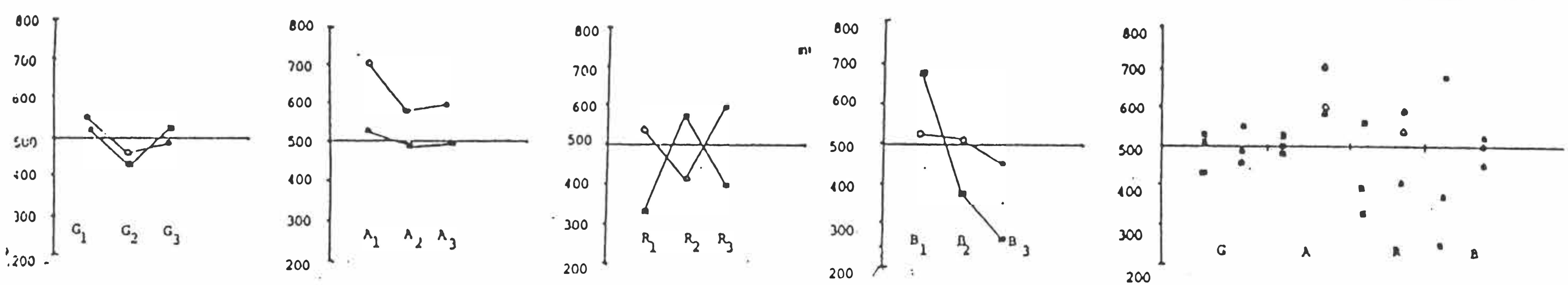
Factor VII was typified by the teacher's verbal response behavior. The positive pole was related to situations in which pupils worked individually, tasks were differentiated and the teacher stimulated the pupils, activity and thinking by acceptance of their movement ideas.

Figure 16 (cont.).

E. Factor V. SPONTANEOUS PUPIL ACTIVITY (+)/STRUCTURED ACTIVITY (-)



F. FactorVI. SUBJECT CENTRICITY - PROCESS CENTRICITY: Teacher-dominant subject centricity (+)/Group activity centricity (-)



G = Gymnastics
A = Apparatur
R = Rhythmic movement-expression
B = Ball games

1 = lower level
2 = middle level
3 = upper level

• = Teacher 1 (man)
o = Teacher 2 (woman)

The negative pole was related to situations in which the class was divided, tasks were distributed among groups and within groups, and the teacher encouraged different groups by acceptance and praise. The rhythmic movement expression lessons of both teachers had a high loading for this factor (Figure 16G). This factor was labelled "Attributing teacher's response behavior to individuals/groups."

Factor VII:	Cluster/ Category	Positive pole	Cluster/ Category	Negative pole
	III/6	+.71	III/4	-.33
	I/03	+.65	I/01	-.32

The Factor Structure by Frame Factors

The behavior of the resultant factors was considered in combination with certain variables, and frame factors, as classified according to the sex of the teacher, grade level and physical education subject area. The results are illustrated in Figure 17.

For the factor scores reported for the two teachers in Table 40, a high Factor I score indicates a predominance of behaviors extending the pupil's freedom of action, whereas a high Factor II score indicates an

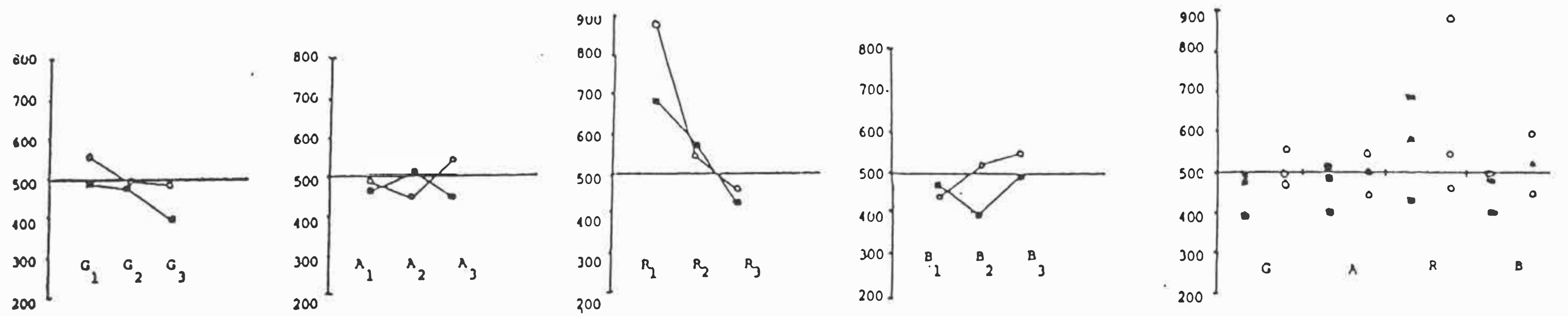
Table 40. Significance of Differences between Factor Scores Estimated for the Two Teachers (Man-Woman) (24 lessons, N=12) (ANOVA)

	MAN TEACHER		WOMAN TEACHER		df.=22
FACTOR	X	SD	X	SD	t
1	500	104	501	102	.00
2	465	48	536	428	-1.82
3	524	83	476	116	1.19
4	532	121	468	72	-1.57
5	519	141	481	22	-.94
6	467	114	533	80	1.65
7	477	82	523	118	1.11

accentuated part played by teacher initiation and direct communication, reducing the pupil's freedom of action. Factor III indicates a uniformity of teacher guidance and specificity of silent guidance in situations such as ball games and apparatus work.

Figure 16 (cont.).

G. Factor VII. INDIVIDUALITY-GROUP CENTRICITY: Attributing teacher's response behavior to individuals (+)/Groups (-)

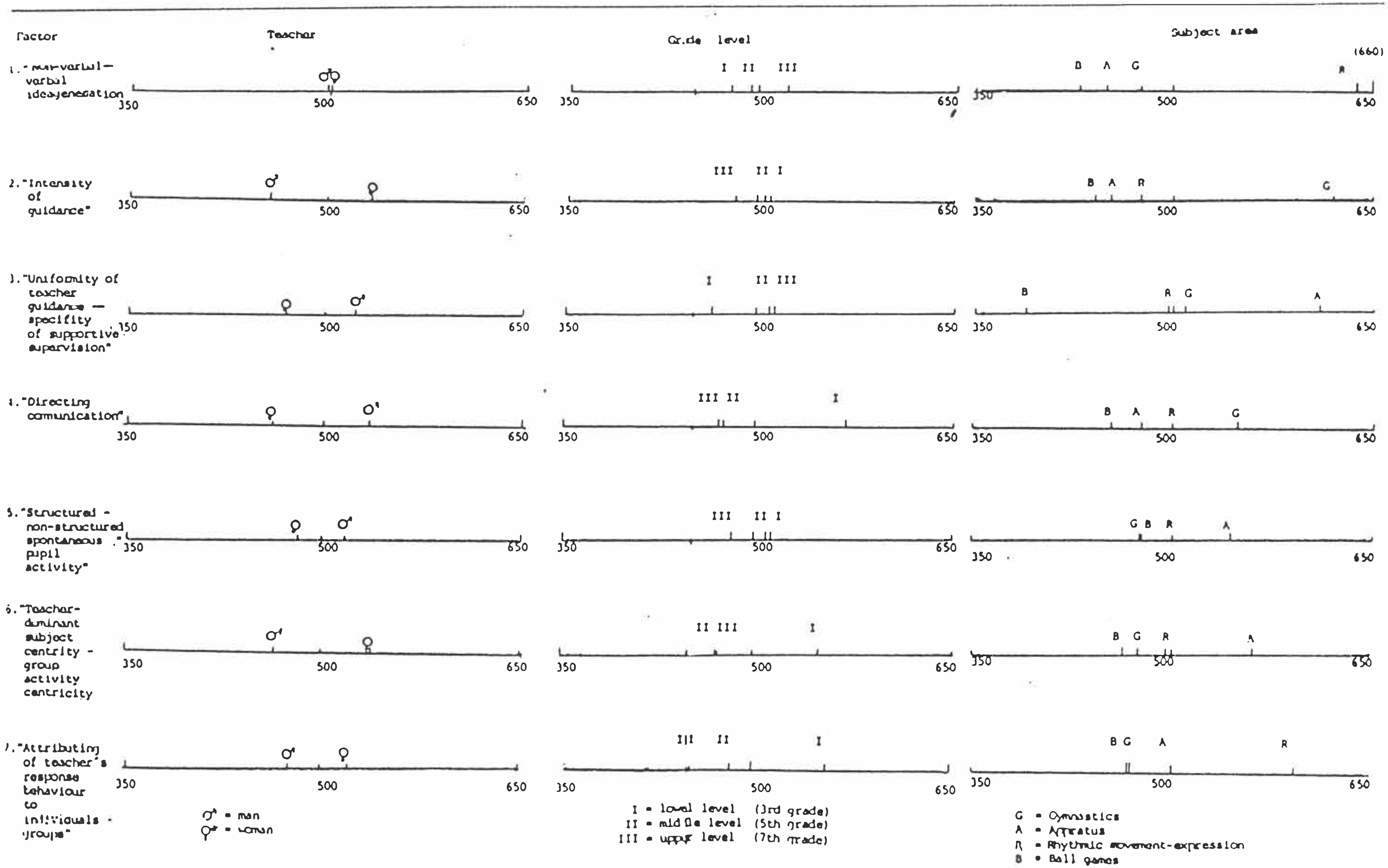


G = Gymnastics
A = Apparatur
R = Rhythmic movement-expression
B = Ball games

1 = lower level
2 = middle level
3 = upper level

• = Teacher 1 (man)
o = Teacher 2 (woman)

Figure 17. Average locations of different frame groups (teacher, grade level, subject area) in factor structure dimensions of physical education interaction process (7 factors, Varimax solution)



The differences between teachers of different sex were clearest in the case of Factor II. The female teacher's direct guidance was more intensive than that of the male teacher, whereas in Factor IV the male teacher's behavior clearly differed and was typified by directing communication.

The differences in teaching in relation to the three grade levels (Table 41) were clearest in Factor IV. The amount of directing communication varied according to the age of the pupils.

Table 41. Significance of the Difference between Factor Scores Estimated for the Lessons of Three Grade Levels (24 lessons, N=8)

	LOW GRADE		MIDDLE GRADE		UPPER GRADE		LOW-MIDDLE df=14	LOW-UPPER df=14	MIDDLE-UPPER df=14	(ANOVA) df=2
FACTOR	X	SD	X	SD	X	SD	t	t	t	F
1	482	21	498	101	521	148	-.43	-.73	-.38	.29
2	509	90	507	123	485	101	.03	.49	.38	.12
3	473	86	512	111	515	115	-.77	.83	-.07	.40
4	570	139	479	70	451	39	-1.66	-2.32	-.96	3.59
5	539	172	490	28	471	9	-.79	-1.11	-1.84	.96
6	54	114	476	71	475	111	-1.53	-1.30	.01	1.40
7	554	150	486	56	460	55	-1.20	-1.67	-.97	1.99

The differences between subject areas in relation to the factor structures (Table 42) were great and clearest in the first three factors. Rhythmic movement expression differed from the others in the first dimension, gymnastics in the second dimension, and apparatus and ball games differed greatly from each other in the third structural dimension. In this context, gymnastics and apparatus were similar to each other and differed from both ball games and rhythmic movement expression. On the other hand, in the lessons of ball games and rhythmic movement expression, the interaction was uniquely almost silent, differing from the communication of the other subject areas.

Grouping Analysis Based on Factor Scores

Procedures Used in Grouping Analysis

In the preceding section, the factor scores estimated for the lessons were considered by interpreting the content of the structural

Table 42. Significance of Differences between Factor Scores Estimated for the Four Subject Areas. (ANOVA)

Subject area / Factor No	Gymnastics N=6 m	d	Apparatur N=6 m	d	Rhytmic N=6 m	d	Ball games N=6 m	d	1-2 df=10 t	1-3 df=10 t	1-4 df=10 t	2-3 df=10 t	2-4 df=10 t	3-4 df=10 t	df=3 df=20 F
1	475	14	450	13	660	115	435	31	3.21	-3.49	2.91	-4.33	1.09	4.22	14.88
2	631	118	451	26	476	48	442	45	3.65	2.98	3.67	-1.14	.44	1.29	9.97
3	506	66	611	63	499	48	385	79	-2.82	.21	2.88	3.47	5.49	3.01	12.12
4	553	191	481	37	492	37	475	68	-.91	.77	-.95	.53	-.19	-.54	.71
5	478	17	547	203	497	32	478	9	.83	1.31	.16	-.59	-.82	-1.34	.59
6	497	45	563	82	474	108	466	143	1.71	-.50	-.51	1.61	1.43	-.10	1.12
7	469	54	477	37	585	170	469	54	.29	1.60	.01	1.53	-.27	-1.59	2.15

$p < 0.01$

N = 24 lessons
the mean = 500
standard deviation = 100

6 observers
4800 six sec. time units
tot. 28800 time units

dimensions by comparing the factor scores and location of the lessons in the different dimensions when classified according to the sex of the teacher, grade level, and the physical education subject area. In this section, the significance of these frame factors will be determined by considering the results of grouping analysis based on the factor analysis.

In grouping analysis, the goal is to form groups for each of which the sum of distances from the group mean of observation will be minumum. The number of groups must be decided in advance. For this purpose 4 to 9 groups were formed because the factor analysis had yielded seven factors. All the HYLPGA groupings were repeated with three different initial values. The emerging groupings varied to some extent, depending on the initial values.

Results of Grouping Analysis and Frame Factor Specificity

The results of the grouping analysis are presented in Table 43 and Figure 18, which illustrate the average location of the six lesson groups (1-6) on the seven varimax factor dimensions on the basis of their means and standard deviations. The principal lessons of the factors were identified by considering both the results of the grouping analysis and factor scores.

Table 43. Estimated Factor Scores of the Six Groups Formed by Means of Grouping Analysis

VARIMAX	GROUP 1		GROUP 2		GROUP 3		GROUP 4		GROUP 5		GROUP 6	
FACTOR	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD
I	441	30	465	00	454	14	707	55	477	17	505	17
II	452	46	490	00	470	45	470	46	735	37	512	35
III	390	68	531	00	611	50	487	50	491	38	528	3
IV	534	157	527	00	483	31	478	34	467	117	518	12
V	476	11	961	00	469	13	500	30	483	10	491	26
VI	475	123	529	00	544	83	492	89	499	38	437	106
VII	471	46	473	00	462	48	489	60	497	31	777	99

It was found that the lesson groups were located at the positive pole in four of the seven structural factor dimensions and at both poles in Factor III. Thus, the behavior in these lesson groups was "known", characterised by the dominating features of these poles.

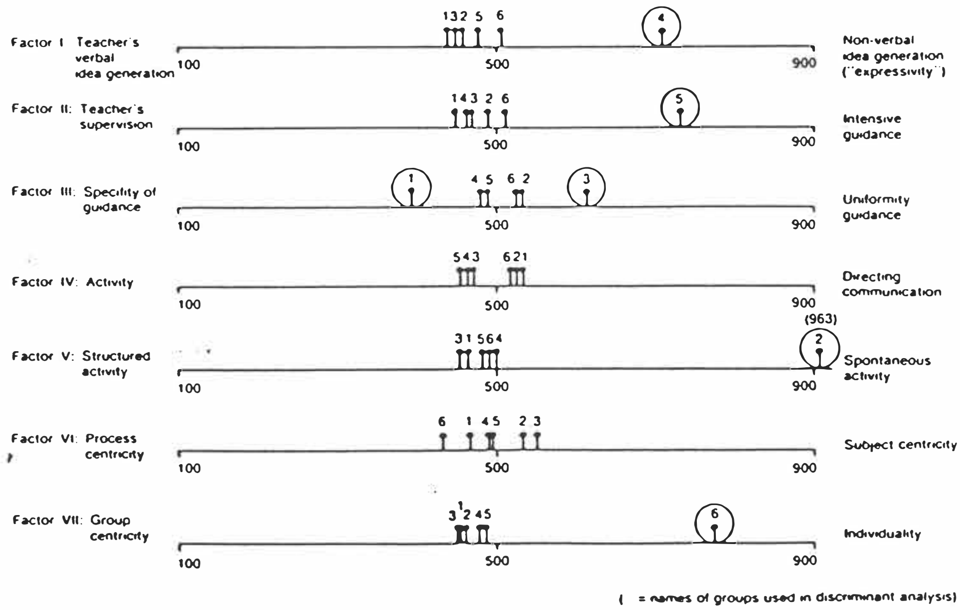


Figure 18. The average location of lesson groups 1-6 on the varimax factor dimensions based on their means and standard deviations.

By considering the behaviors of the resultant factors and lesson groups in combination with the top factors, five factors appear to be connected with the grouping of lessons, and both poles of Factor III showed the most predictive power in the grouping of lessons (Table 44).

By considering the behavior of lesson groups in combination with frame factors, as classified according to the sex of the teacher, grade level and physical education subject area, it was found that there were two principal sources of variance in the set of lessons: the P.E. subject area, and the teacher. A possible third source of variance consisted of the interaction between the first two, and a fourth, of the interaction between the first two and the grade level.

Both in the factor analysis and in the grouping analysis, the lessons had a certain tendency to cluster according to the P.E. subject area. The teachers appeared to follow the traditional ways of teaching different P.E. subject areas. Or perhaps it was the subject area itself, its structure and content, that caused the teacher to choose a certain way of teaching, using direct or indirect influence. Or maybe the measuring instrument was itself sensitive in describing this kind of

Table 44. Variation of Six Groups Through Principal Factor, Teacher, Grade Level, and Subject Area.

Group No	Lesson No	Principal factor No	Teacher Man	Teacher Women	Grade level Low	Grade level Middle	Grade level High	Subject area Gymn.	Subject area Appr.	Subject area Rhytmic,	Subject area Ball games
1	5,8,11,12 13,16,23	3(-) unif.	4	3	3	2	2	1			6
2	21	5(+) spont.	1			1			1		
3	1,2,7,18 20,22,24	3(+) specif.	4	3	1	3	3	2	5		
4	6,10,14,15	1(+) expr.	2	2		2	2			4	
5	3,4,9	2(+) intens.		3	1	1	1	3			
6	17,19	7(+) indiv.	1	1	2					2	

Principal facators in grouping analysis

I Expressivity (4)
II Intensity of guidance
III Uniformity (1) specifity of guidance (3)
IV Directing communication
V Spontaneous pupil activity (2)
VI Subject centricity - nonverbal group work centricity
VII Individuality - group centricity non directive communication (6)

behavioral differences. In any case, such grouping is regarded as too narrow (cf. Flanders, 1965). But how to describe the source of this variance, and thus the predictive power of the category system, is a problem. In the following section this variance is examined more closely using discriminant analysis techniques in an effort to describe its source.

Discriminant Analysis of Lesson Groups Formed with Factor Scores

This section reports the attempt to identify and describe the factors which predict variability and the grouping of lessons when using the categories of the PEIAC/LH-75 three-cluster category system.

Earlier, an attempt was made to explore the interaction in 24 P.E. lessons by means of the factor analytical r-technique and to form homogenous groups of lessons in a grouping analysis based on the factor scores. The behavior of the resultant varimax factors and lesson groups was considered in combination with the frame factors of the study. However, the two principal sources of variance, the P.E. subject area and the teacher, were regarded as too narrow. The principal lessons of the factors were located at the positive pole in four of the seven factoral dimensions, and at both poles in Factor III, describing "known" behavior as characterized by the dominating features of these poles (see Figure 18). The problem then was to describe the source of this variance and, thus, the predictive power of the category system used.

Applying the concepts used by Cheffers (1973) we ask now: (1) Is the instrument sensitive enough to make the discrimination required for research problems (sensitivity), and (2) does the instrument possess the ability to distinquish between groups "known" to behave differently on the construct under study (construct validity)? A useful way to explore this question further is a cumulative evaluation of the results obtained in the factor and grouping analyses.

Because there was no external criterion available to assess the construct validity of this instrument, it was decided to use multiple discriminant analysis for examining more closely the portion of variance through "criterion groups" which were predictable from or explained by the known variance on the linear combination of predictors (Cooley &

Lohnes, 1971). The design involved the assessment of two or more traits by two or more methods.

Discriminant Analysis of the Observational Data

In Part II of this report, the applicability of discriminant analysis for assessing and describing factors predicting inter-coder disagree ment was demonstrated. In the context of the present analysis, the aim is to apply this procedure to assess and describe factors predicting the variability and grouping of lessons.

Multivariate statistical correlational procedures such as factor analysis and discriminant analysis offer the investigator the opportunity to combine a relatively large number of variables into a single score. Factor analysis is commonly applied for assessing the construct validity of the measuring instrument, as in this study. The use of discriminant analysis for this purpose has been presented more comprehensively in a previous report by the present author (Heinilä, 1980) and by Cooley and Lohnes (1971), who state that "the discriminant model may be interpreted as a special type of factor analysis that extracts orthogonal factors of measurement battery for the specific task of displaying and capitalizing upon differences among criterion groups" (p. 243).

Moreover, the difference between grouping analysis and the discriminant analysis makes it possible to minimize the proportion of lessons placed in "wrong" cells. As stated earlier, the assumptions concerning the level of measurement also need to be considered, since discriminant analysis presupposes interval scales.

Results of Discriminant Analysis and their Interpretation

The data used for the discriminant analysis were the score distributions of categories from the 24-lesson data (T_2) as coded by six trained observers, and the 27 categories of the three-cluster classification system. The six lesson groups formed by using grouping analysis based on factor scores were structurally homogeneous and there were differences in the mean distributions of variables (Table 45). Missing observations were replaced by mean values. The data fulfilled the requirements set on the number of criterion groups and variables. As cited earlier, the limitations of the level of measurement were taken into account in the interpretation of the data.

Table 45. Means and standard deviations of six lesson groups formed by means of grouping analysis based on factor scores classified by six observers (N = 144)

Cluster	Cat.		Group 1 N=42 m	s	Group 2 N=6 m	s	Group 3 N=42 m	s	Group 4 N=24 m	s	Group 5 N=18 m	s	Group 6 N=12 m	s	Total N=144 m	s
I		Teacher's talk, movement; pupil talk; other														
	Teacher	01. Accepts, praises, encourages	4.69	4.17	7.83	4.83	9.71	4.34	3.46	3.20	5.83	4.45	3.58	2.3	6.13	4.68
		02. Gives corrective feedback, urges	7.45	5.32	21.00	8.12	18.45	11.41	5.54	5.65	9.11	5.10	8.58	4.76	1.21	9.17
	•	03. Uses, develops ideas, movement, tasks, suggested by pupils	.43	.78	1.00	1.26	.45	.86	.42	.83	.55	1.04	2.25	2.49	.63	1.19
		04. Asks, initiates and terminates activity	12.69	8.32	8.5	4.36	9.09	7.76	10.87	8.49	26.78	9.97	9.08	14.35	3.56	10.49
		05. Presents information, organizes	77.28	21.83	88.66	10.01	91.47	16.24	71.73	21.06	70.61	15.48	67.50	12.24	79.26	20.19
		06. Gives directions, commands during activity	3.76	3.62	4.83	5.63	5.24	4.83	4.79	4.57	28.89	8.56	5.41	4.01	7.69	9.48
		07. Criticizes pupils behaviour	1.71	2.76	1.5	1.76	1.38	2.27	1.66	3.39	.72	1.02	2.67	1.56	1.56	2.48
	Pupil	08. Answers question/clarifies, demonstrates	1.76	2.37	.83	.75	.47	1.02	.33	.82	1.72	2.53	2.58	2.11	1.17	1.93
		09. Pupil speaks spontaneously, initiates	3.29	4.24	10.33	3.61	3.33	3.78	1.75	2.15	2.61	4.31	4.17	4.60	3.33	4.11
	Teacher	10. Teacher follows pupils' activity, silent guidance	81.62	33.10	42.50	14.59	52.05	18.42	58.12	17.84	40.72	17.60	75.92	22.66	61.86	27.33
		11. Silent participation in movement activity	3.12	4.78	10.83	2.79	6.31	8.06	39.66	24.99	10.33	6.54	6.25	6.77	11.63	17.28
	Other	12. Confused situation	2.19	.89	2.17	.98	2.02	.15	2.04	.20	2.11	.47	2.00	.00	2.09	.55
II		Pupils' collective movement activity/passivity and social access														
	Activity	1. Inter-pupils contacts and movement, space, time, energy restricted; range of ideas controlled	28.81	25.29	.00	.00	3.71	9.62	10.62	15.78	92.11	24.40	.25	.62	22.79	33.65
		2. Inter-pupil contacts and/or movement free; range of ideas controlled	86.98	47.30	107.66	5.95	111.62	32.19	40.96	26.82	21.78	14.37	13.50	38.50	81.33	48.00
		3. Inter-pupil contacts free; range of ideas open	.38	2.01	.00	.00	3.10	7.01	81.62	29.20	5.33	9.16	10.67	13.33	16.17	32.35
		4. Pupils' spontaneous activity	.12	.45	15.16	2.04	.16	.69	1.12	2.40	.00	.00	.67	2.31	.96	2.17
	•Passivity	5. Pupils follow instruction, demonstrations	54.69	25.97	44.67	5.89	57.83	24.90	43.87	15.38	63.28	8.19	53.83	27.32	54.39	22.62
		6. Pupils organize themselves, assist in organization	25.83	8.35	30.10	5.58	21.10	13.76	16.12	10.86	13.94	5.60	18.38	7.12	20.94	11.10
		7. Pupils wait for turn	.92	1.60	.17	.41	.38	1.06	3.58	2.82	1.56	1.95	1.25	2.14	1.28	2.09
	Other	8. Confused situation	2.26	.94	2.33	1.03	2.09	.37	2.08	.41	2.00	.00	2.00	.00	2.13	.61
III		Social form														
	Situation	1. Complete class, uniform task	11.21	19.22	84.67	3.56	69.86	46.71	61.83	25.86	150.27	46.58	80.17	41.36	69.94	55.08
		2. Divided class, uniform task	136.71	66.60	1.16	2.40	21.36	29.67	34.91	35.84	19.00	27.28	19.92	21.86	56.01	67.86
		3. Divided class, differentiated tasks	14.07	27.11	85.83	5.45	96.43	63.67	17.37	18.95	17.39	25.33	52.30	54.95	45.25	55.65
		4. Divided class, differentiated tasks distributed amongst groups + within group	35.66	43.53	10.83	2.64	10.33	20.07	16.04	28.55	11.33	16.52	.00	.00	17.96	31.10
		5. Individual work, uniform task	5.33	2.16	2.66	6.05	.00	.00	67.70	10.80	.00	.00	39.42	41.20	14.78	28.85
		6. Individual work, differentiated tasks	.00	.00	.00	.00	.00	.00	.40	.20	.00	.00	6.00	8.21	.51	2.84
		7. Other situation, confused situation	2.00	.00	14.83	6.37	2.02	.15	2.08	.41	2.00	.00	2.00	.00	2.56	2.81

6 observers
24 lessons
4800 6 second time units, tot. 28800 time units

The five resulting discriminant functions are presented in Table 46. The table displays the structure from Wilks's lambda, indicating that the five discriminant functions separating the lesson groups can be considered highly significant statistically, which was expected in this context.

It was established that the power of the discriminant functions to separate lesson groups was great, since their canonical correlations were relatively high. The first discriminant function proved clearly more powerful than the other four. Its share of the total discrimination of the model was 47%, that of the second being 19%, and the third 18%. From the point of view of interpretation, all discriminant functions were clear and important in view of the theory.

The program selected 16 of the 27 classification categories and set them in sequence according to how much they increased the model's discriminating power. It is possible even on the basis of these categories to get an idea of the nature of the discrimination. The discrimination model included the seven categories of Cluster I (Verbal) four of Cluster II (Movement and Social Access), and four of Cluster III (Social Form). The categories of Cluster II, representing pupils' collective activity with the range of ideas closed and with open ideas, and the categories of Cluster III showed the most predictive power. Both categories which occurred rarely and those occurring most frequently were represented in the model.

Content and Interpretation of Discriminant Functions

The following principles and sequence were used in the interpretation of the contents of the discrimination dimensions: First, note was made of the variables that had obtained high weights on scaled eigenvector(s) and of their relative discriminating power. Second, it was ascertained how highly discriminant functions correlated (r) with variables selected for the model. Third, it was established how known groups of lessons were placed on the discriminant dimension on the basis of their means and standard deviations on these dimensions. Finally, their mutual placement in the discrimination plane, formed by two discrimination dimensions at a time, was studied.

Table 46. Discriminant analysis on lesson groups and categories of three cluster category system (PEIAC/LH-75)

CLUSTER	CATEGORIES		GROUPS 1. unif.	2. spont.	3. spesif.	4. expr.	5. intens	6. indiv.	Power of discrimination in categories JN	F	n_1	F_{lls}	I Discriminative functio s	r	II Discriminative function s	r	III Discriminative function s	r	IV Discriminative function s	r	V Discriminative function s	r
I	01 Accepts, praises, encourages		4.69	7.83	9.71	3.46	5.83	3.58	14	79.14	70	1.71	-.39	-.24	.27	.27	.85	-.21	.48	-.36	-.16	-.12
	03 Uses, developes ideas, movement, tasks suggested by pupils		.43	1.00	.45	.42	.56	2.25	16	72.49	80	.81	-.15	.09	-.23	.00	.14	.03	-.82	-.03	.33	.49
	04 Asks, initiates and terminates activity		12.69	8.50	9.10	10.88	26.78	19.08	15	76.17	75	1.15	.52	.04	-.16	-.19	-.19	-.38	.16	.32	.15	26
	06 Gives directions, commands during activity		3.76	4.83	5.24	4.80	28.89	5.42	3	127.70	15	322.90	.32	.05	-.41	-.05	-.92	-.80	.41	.41	.60	06
	08 Answers question/clarifies, demonstrates		1.76	.83	.48	.33	1.72	5.58	13	83.55	65	3.30	-.14	-.13	-.38	-.17	.18	.01	.43	.21	.22	.32
	09 Pupil speaks spontaneously, initiates		3.29	10.33	3.33	1.75	2.61	4.16	11	92.39	55	4.41	-.33	-.17	-.16	.33	.57	.09	.27	.07	.94	.20
	10 Teacher follows pupils' activity, silent		81.62	42.50	52.05	58.13	40.72	75.92	10	95.39	50	6.50	-.46	-.17	-.57	-.31	.49	.43	.79	.08	.14	.13
II	1 Inter-pupil contacts and movement restricted: range of ideas controlled		28.81	0.00	3.71	10.62	92.11	.25	8	107.22	40	12.94	-.22	-.12	-.29	-.24	-.19	-.59	.80	.62	.54	.09
	3 Inter-pupil contacts free: range of ideas open		.38	0.00	3.10	81.63	5.33	10.66	2	148.28	10	1859.86	.72	.85	.11	-.09	.19	.20	.24	.04	-.66	-.35
	4 Pupils' spontaneous activity		.12	15.17	.17	1.13	0.00	.66	1	147.60	5	xxxxx	.45	.09	.73	.86	.14	.23	.42	.31	-.17	.18
	7 Pupils wait for turn		.93	16.7	.38	3.58	1.56	1.25	12	87.74	60	3.19	.50	-.47	.60	-.16	.14	[illegible]	-.29	.16	-.21	-.16
III	2 Divided class, uniform task		136.7	1.17	21.36	34.92	19.00	19.92	4	112.15	20	128.90	-.75	-.41	-.43	-.39	1.14	.43	.49	.33	-.19	-.18
	4 Divided class, differentiated tasks distr. amongst groups & within group		35.67	10.83	10.33	16.04	11.33	0.00	7	107.53	35	21.18	-.40	-.19	-.41	-.14	.35	.21	.76	.19	.38	-.19
	5 Individual work, uniform task		.33	2.67	.00	67.71	0.00	39.42	5	107.39	25	64.90	1.23	.87	-.43	-.13	.25	.28	.95	.00	.60	.01
	6 Individual work, differentiated tasks		.00	0.00	.00	.04	0.00	6.00	6	112.54	30	33.67	.43	.14	-.23	-.12	.38	.07	.12	-.10	.98	.65
	7 Other situation, confused situation		2.00	14.83	2.02	2.08	2.00	2.00	9	101.15	45	9.03	.83	-.02	.47	.86	.21	.19	.22	.32	.28	.17
	Number of observations (144)		(42)	(6)	(42)	(24)	(18)	(12)	n_2 = 138				47% xxx		19,7%		17,0%		10,7%		4,32%	
	DF I: Range of ideas:	m	-4.25	-.21	-1.62	10.45	.82	4.43	4 ≠ others 1 ≠ others 2=3<5				x^2_{20}=442.3						of total discrimination			
	closed-open	d	.77	.33	.67	1.43	.82	1.80					Rc = .98									
	DF II The level of structuration: High-low	m	-3.90	12.42	-.59	-2.40	-2.60	-3.99	2 ≠ others 1 = 4 = 5 = 6 ≈ 3						xxx x^2_{16}= 332.1		xxx x^2_{16}=319.9					
		d	.61	3.25	.68	1.27	.44	.97							Rc = .96		Rc = .95					
	DF III: Intensity of teacher guidance High-low	m	5.61	5.88	1.76	4.72	-4.09	4.10	5 ≠ others 5 = 3; 1 = 2 ≈ 4 = 6 ≈ 1 ≠ 3										xxx x^2_{14}=260.3			
		d	.97	1.14	.92	.99	1,18	1.06											Rc = .93			
	DF IV: The level of spesifity of guidaance High-low	m	7.42	9.61	2.59	6.19	9.02	4.56	3 ≠ others 3 ≈ 6, 2 = 5 ≈ 1 ≈ 4												xxx x^2_{12}=164.4	
		d	1.02	1.69	1.01	.92	.89	.72													Rc = .84	
	DF V: The media of non-directive communication Verbal - non-verbal	m	.76	2.30	.57	-.29	1.22	5.80	6 ≠ others 2 ≈ 5; 1 = 3 ≈ 4													
		d	.48	1.12	.58	1.50	.49	2.17														

s = variables scaled in W-matrics, m = means on discriminative function, d = deviation on discriminative function, r = correlations between discriminant funktions and variable

From the structure coefficients of Table 46 and Figure 19, and the nature of the factors, the five functions extracted appear to measure:

DF I: Range of ideas for pupils; closed - open

DF II: The level of structuration: high - low

DF III: The level of intensity of guidance: high - low

DF IV: The level of specificity of nondirective guidance: high - low

DF V: The media of nondirective communication (attributing teachers response behavior to individuals/groups): non-verbal - verbal

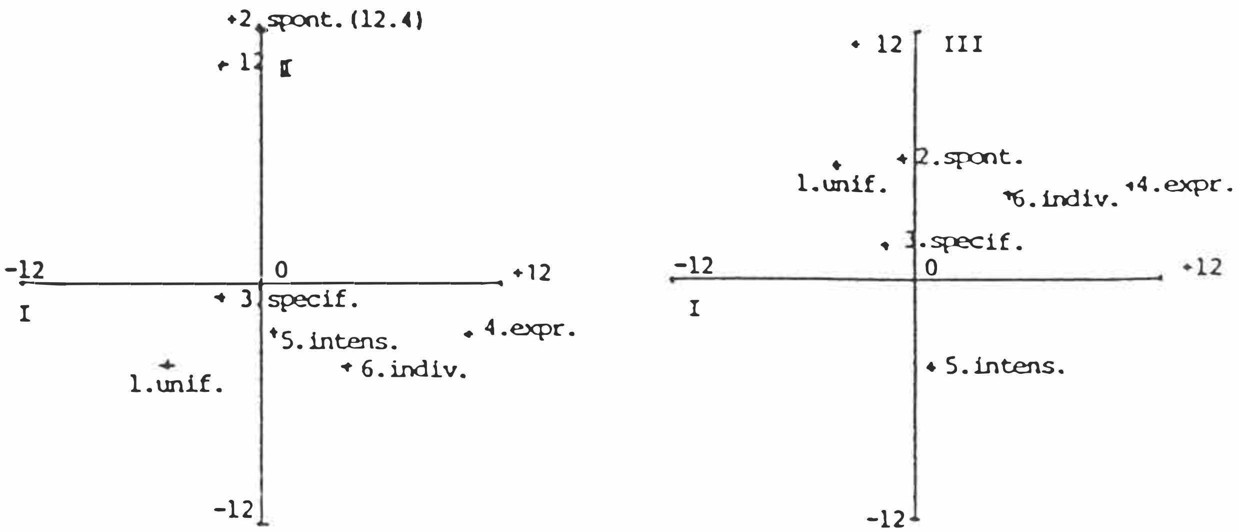
Thus, these discriminative dimensions describe different aspects and levels of "teacher's control of students' freedom of action," which is the feature that Flanders (1965) gives as the main purpose of interaction analysis.

Structure of Criterion Groups and Degree of Discriminant Validity

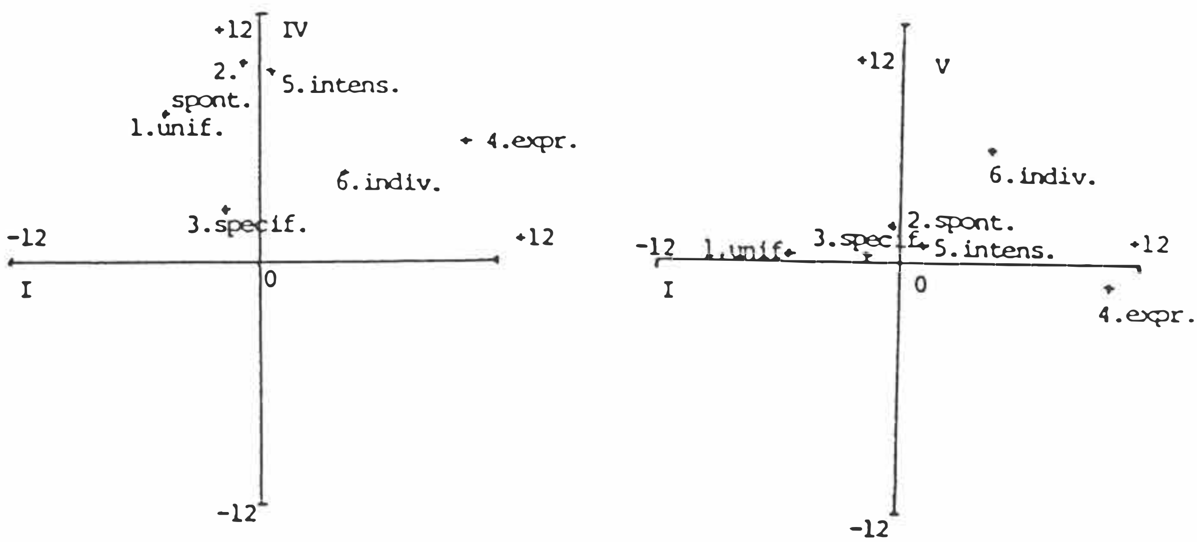
On the basis of the nature of structural differences, it is possible to describe the problems and level of the discriminant validity and sensitivity of the "testing of the instrument". The six "criterion groups" were quite heterogeneous when examined in the light of differences revealed in the variablity of scores by using this classification system on the construct under study. Group differences stood out clearly in five linear factor groups of different composition (Figure 20).

As is usual in discriminant analysis, the first linear function predictor of difference separated one group from the rest (in this case it separated two groups, 4 and 1, from the rest), then the next one (group 2) forms the rest and so on. The groups' variability was large, especially in the first three dimensions, and in the discrimination space defined by the first and the other discriminant dimensions (see Figure 19).

The structure of the discriminative model was related to the structure of the measuring instrument and produced the following sequence predicting the grouping of lessons:



A. Discriminant functions I and II   B. Discriminant functions I and III



C. Discriminant functions I and IV   D. Discriminant functions I and V

Figure 19. Placement of lesson groups 1-6 centroids on the discrimination plane on the basis of the means and standard deviations of the discriminant functions

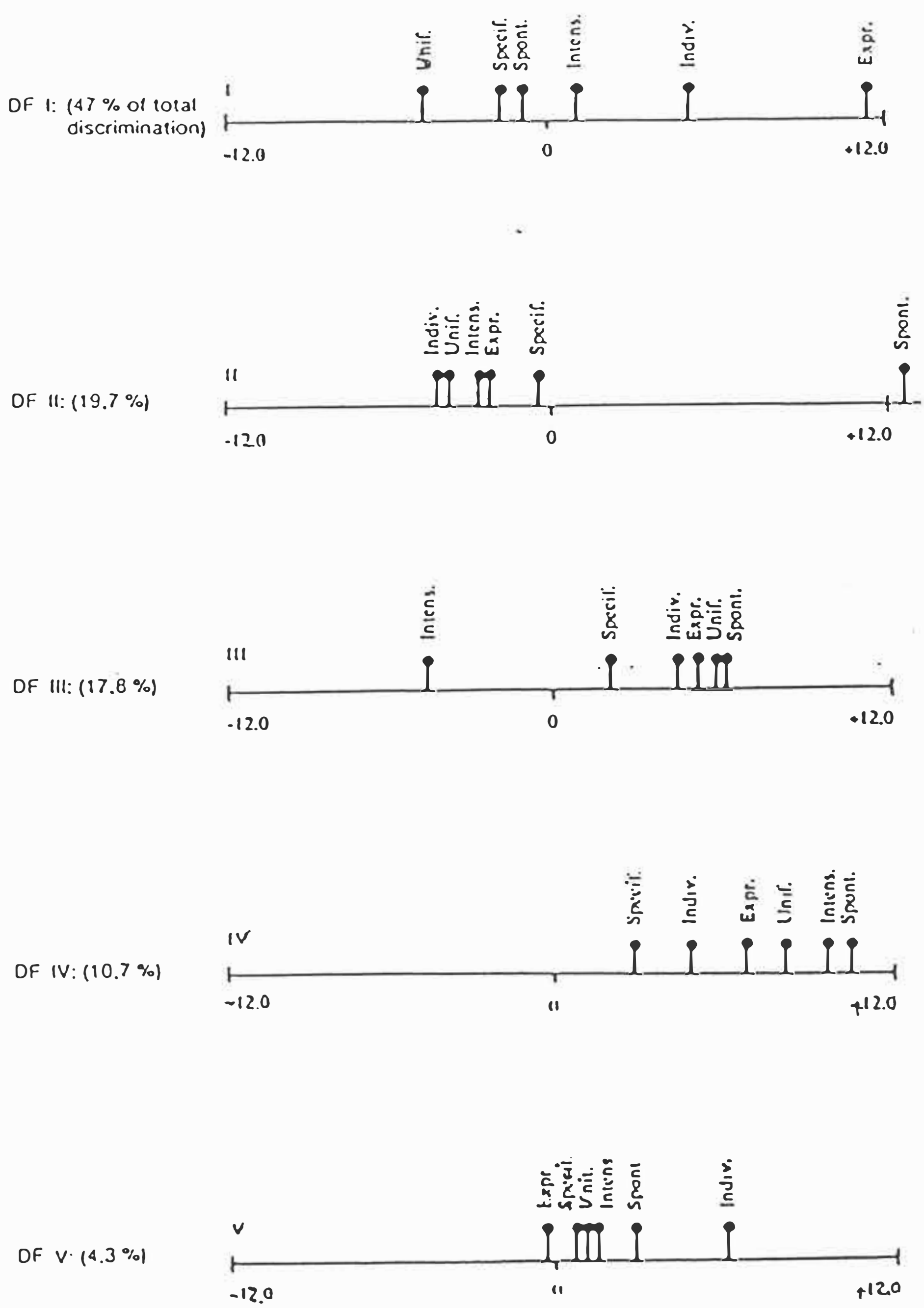
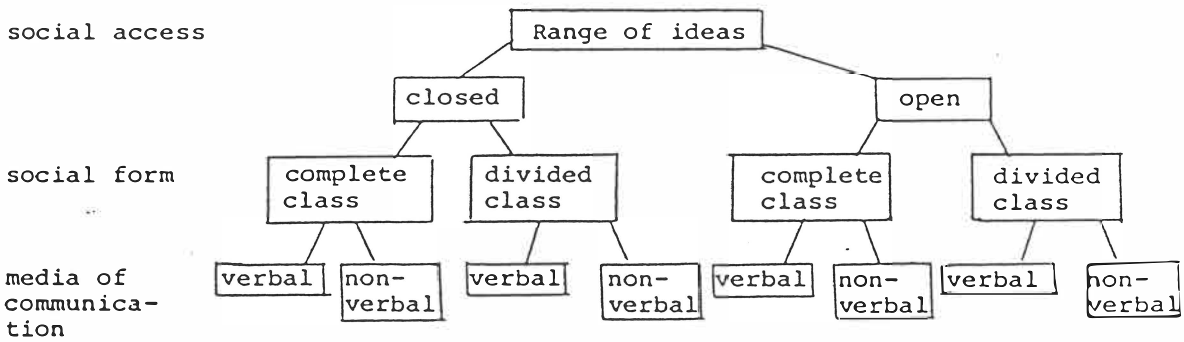


Figure 20. Placement of groups 1-6 centroids on the discrimination plane formed by discriminant functions I-V

predicting the grouping of lessons:



Summary

Part III of the results has examined the construct validity and sensitivity of PEIAC/LH-75, which is an adaptation of the Flanders Interaction Analysis System (FIAC) developed for the specific purpose of describing the instructional process in physical education classes.

The first stage explored interaction in 24 P.E. lessons by means of the factor analytical r-technique from the point of view of construct validity of Flanders' theory. The second stage examined the formation of homogenous groups of lessons in a grouping analysis based on the factor scores. The nature of the factors and lesson groups were considered in combination with the frame factors of the study.

It was found that there seemed to be two principal sources of variance in this set of lessons, (1) the P.E. subject area and (2) the teacher, and perhaps two others consisting of (3) the interaction between the first two, and (4) the interaction between the first two and the grade level. The principal lessons of the factors were identified by considering both the results of the grouping analysis and the factor scores. It was found that the lesson groups were located at the positive pole of four of the seven factoral dimensions and at both poles of Factor III. Thus, the behavior in these lesson groups was "known", characterized by the dominating features of these poles.

In the third stage, an attempt was made to describe the source of the variance, and thus to determine more closely the predictive power of the category system, by using a multiple discriminant analysis technique tically. The share of total discrimination for each discriminant function was: I,47%; II,19%; III,18%; IV,11%; V,5%.

The program selected 16 of the 27 classification categories and set them in sequence according to how much they increased the model's discriminating power. The categories of Cluster II (Movement and Social Access) and the categories of Cluster III (Social Form) showed the most predictive power. Both the categories which occurred rarely and those which occurred most frequently were presented in the model:

DF I: Range of ideas for pupils: open - closed

DF II: The level of structuration: high - low

DF III: The level of intensity of guidance: high - low

DF IV: The level of specifity of nondirective guidance; nonverbal - verbal

DF V: The media of nondirective communication (attributing teacher response behavior to individuals/groups): non-verbal - verbal.

Therefore, the lessons could be placed on the discrimination plane formed by the two discriminant functions, reflecting their aspects of direct/nondirect teaching.

Although the results of the discriminant analysis can only be regarded as tentative on account of the nature of the level of the measurement scale, it yielded quite useful information for the development of the instrument, thus, refining the discriminative model. The discriminant functions that describe factors predicting the grouping of lessons among criterion groups were interesting from the point of view of theory (Flanders 1965, 18, 1970). Lessons could thus be placed in a certain group which reflected their aspects of direct - nondirect teaching in a non-verbal and verbal context. The quality of teachers' verbal behavior had more predictione power in the grouping of lessons than the quantity of it. The quantitity and quality of teachers' nonverbal behavior posed a high predictiove power in the grouping of lessons.

In the present study, the inverse character of reliability and validity was highlighted, which had already been pointed out by Flanders (1970) in his analysis concerning the training of observers and reliability problems.

The principal sources of variance in the classes observed appear to be the subject area and the teacher, and to a lesser degree the interactions among the three frame factors. In the discriminant analysis of lesson groups, the clusters identified as movement and social access (II) and social form (III) showed the most preedictive power of the category system. These results seemed to verify the construct validity ans sensitivity of the instrument.

According to Locke (1977), "possession of reliable instruments for observation and knowing how best to use them, do not in themselves guarantee either sound research or fruitful results, but in the area of teaching they are essential first steps. And as we move to evaluative studies, we will have to confront the problem of multiple criterion measures and we will need product batteries which permit multivariate designs." This study has been an attempt to proceed in the direction recommended by Locke.

CHAPTER VII
THE APPLICATION OF INTERACTION ANALYSIS TO TEACHER TRAINING IN PHYSICAL EDUCATION

In Chapter I it was stated that the central task of the university is the planning and realization of educational programs with the ultimate aim being the quantitative and qualitative development of education. The development of educational programs should be based on scientific research. The preceding chapters have reported the results of a scientific study of physical education classroom interaction and the development of an observation instrument which will permit a detailed description and careful analysis of this interaction.

In this chapter it will be reported how this research has resulted in a program of teacher education which makes use of the observation instrument as a part of the training of future teachers of physical education. Two versions of a microteaching course are compared in order to assess the effectiveness of their components. For the purposes of this comparison, microteaching is described and its components are analyzed, particularly those on which this project focused. In the empirical part of the report, a short description is given of the teaching program, design, hypotheses, and methods of measurement and analysis. Preliminary results are then presented and discussed. In the final section some recommendations are given for courses of microteaching in P.E. teacher education, as well as for related follow-up and research activities.

Microteaching in Teacher Education

In January 1974 the Department of Physical Education of the University of Jyväskylä introduced, on an experimental basis, a new type of practice teaching in the form of a course on microteaching. It formed part of the degree requirements and was given during the last term of the third year as an obligatory course (45 hours). The experiment was started as a result of the positive reports on the use of microteaching and interaction analyzing systems as a tool of teacher education (cf.

Flanders, 1970, Ch. 11). It was considered to have a potentially beneficial effect on the attainment of the objectives of teacher education in physical education as well as on bridging the gap between the theory and practice of teaching. It was for the implementation of this course that the interaction model and observation instrument, PEIAC/LH-75, was constructed. The measuring device had been pilot-tested at the beginning of the course and its use, in modified form, proved feasible.

When the earlier forms of practice teaching, so-called order-calling exercises, were given up as not being consistent with the principles of the new type of P.E. teacher education, there was a decrease in the amount of practice teaching. The student teachers felt that this was a disadvantage, leading to a feeling of uncertainty when they started their one-year practice teaching at the "normal school." The need for new opportunities for practicing was clearly indicated. The present project was instituted in order to develop new forms and contents of practice teaching so that they satisfied the demands of changing physical culture on teacher education.

Evaluation of Curricula

This study is concerned with evaluating teacher education programs in terms of process criteria (changes in teacher's verbal and nonverbal behavior). This evaluation was undertaken as a comparison between two microteaching settings which differed with regard to (1) modeling, (2) sequencing of teaching, (3) timing, (4) number of pupils, and (5) number of reteachings. The evaluation is primarily descriptive and judgemental and its purpose is to indicate the degree of congruence between what is intended and what actually occurs.

The main activities of descriptive evaluation are (1) the study of the contingencies of antecedents, transactions and outcomes, and (2) the study of the congruence between the level of objectives and the level of observations. Congruence indicates to what extent the plan is being carried out (Stake, 1967). This report describes mainly educational intentions at the curriculum level and their degree of realization at the observation level by means of research. Stake's model (Figures 21 and 22) will be used to provide a program evaluation frame of reference for the project.

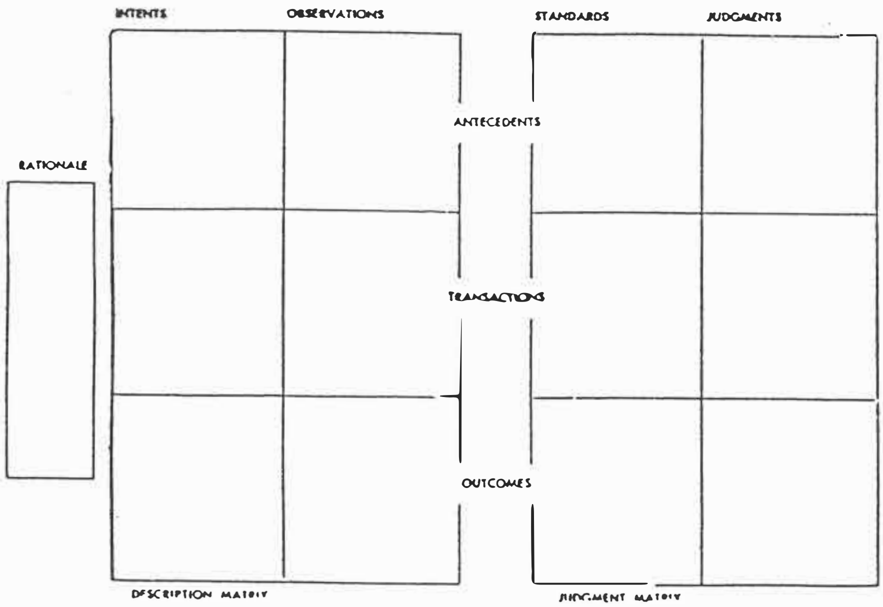


Figure 21. A layout of statements and data to be collected by the evaluator of an educational program (Stake, 1967).

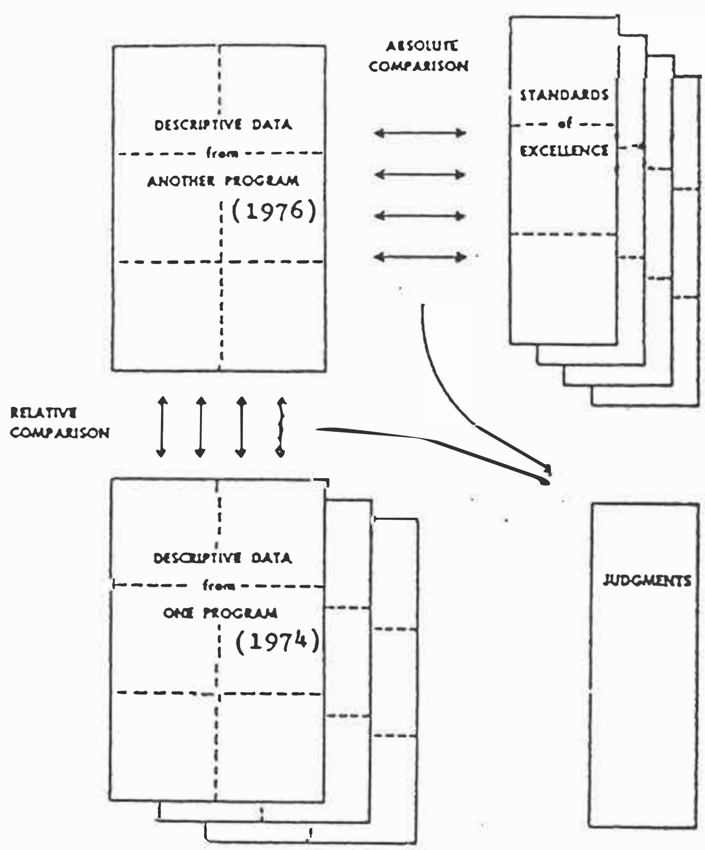
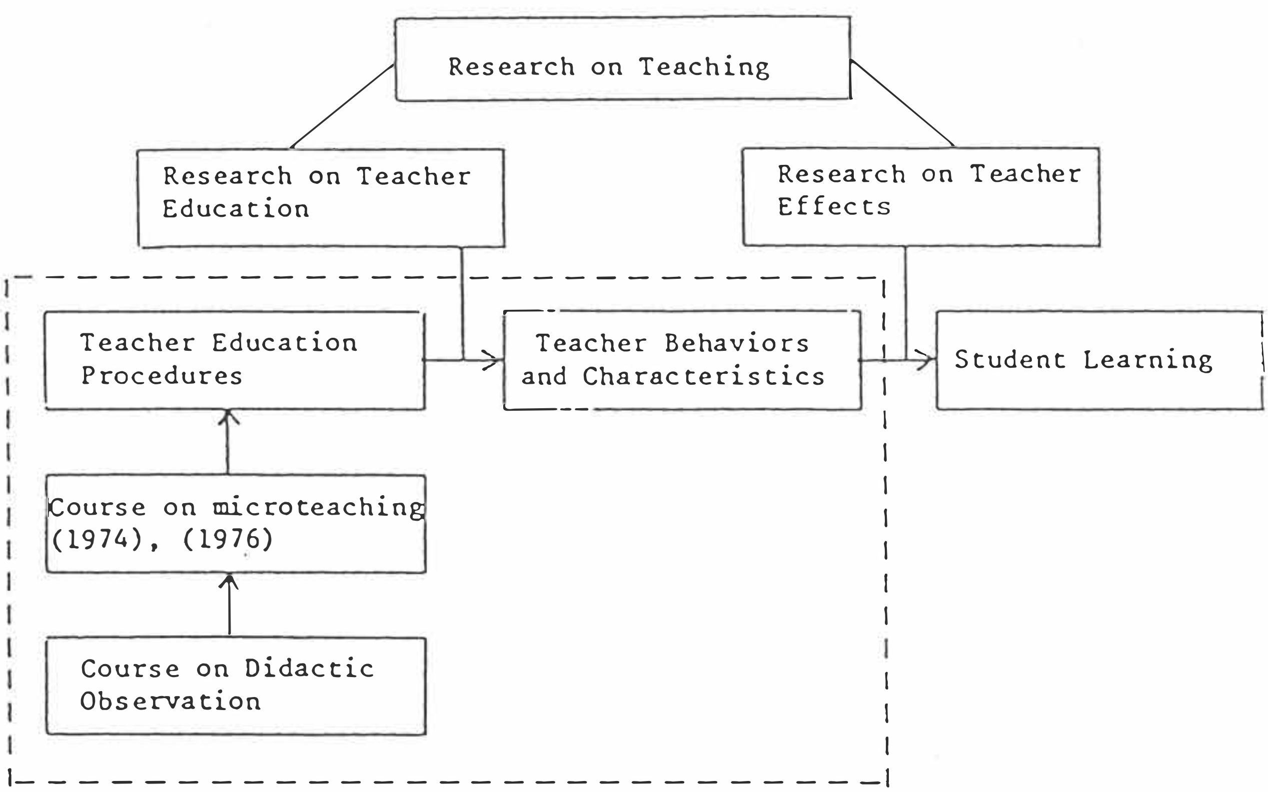


Figure 22. A representation of the process of judging the merit of an educational program (Stake, 1967).

Review of Research

Forms of Practice Teaching

The development of research on teaching and the methods of interaction analysis have been discussed in earlier chapters of this report with special attention to Gage's (1972) model of the field of research on teaching. The adapted version of this model, which was introduced in Chapter 1, is repeated here for the convenience of the reader. It illustrates the place of the present discussion within the research area and describes the starting point for the study of teaching programs. In this section we will review the concepts and purposes of microteaching and minicourses.



(Figure 2.) Adapted version of Gage's model of field of research on teaching.

When microteaching was planned at Stanford University in 1963, the aim was to develop more effective forms of practice teaching. The following criteria for the organization of initial training for pre-intern teachers were set (Allen & Clark, 1967):

First: A real teaching situation was needed so that candidates could be actively engaged in practicing and refining teaching skills and experimenting with their own and their supervisors' ideas.

Second: The teaching situation must keep the risk low both for the teacher and the students.

Third: The pre-service teaching context should take into account some well-established facts within learning theory. For example, numerous distributed practice sessions; immediate supervisory feedback; immediate opportunity to rectify errors and weaknesses; low anxiety, etc.

Fourth: The pre-service context should provide a setting in which the trainee can have experience with a wide range of student abilities and age levels and develop competence with a broad spectrum of teaching skills.

Fifth: Economy in terms of time and resources should be maximized.

Microteaching was conceived to meet these criteria.

The spread of microteaching into colleges of education was very rapid. In 1972, 50% of the colleges of education in the U.S. used various adaptations of microteaching.

Allen and Eve (1968) define microteaching as "a system of controlled practice that makes it possible to focus on specific teaching behaviors" (p. 181). The term "system" here, as well as in discussions of systematic observation, refers to the rigorous plan of choosing and controlling the components of the system beforehand for a certain specific purpose.

In microteaching, the teaching situation is usually scaled down in terms of time and number of students. The "session" lasts four to twenty minutes and the number of students varies from three to ten. Microteaching can be used for a number of purposes. Some of the variables which can be adjusted include lesson length, number of students, type of students, number of "reteachings," the amount and kind of supervision, and the use of videotape (Allen & Clark, 1967).

Microteaching's component-skill approach is used primarily to give the trainee a clear idea of the skill to be learned. The trainee has to know what he should do before he tries to do it. Instruction in a particular skill can be given by oral instructions, written directions, demonstrations or combinations of these. In the usual Stanford microteaching sessions, the procedure is to teach 5 minutes, critique 10 minutes, replay 15 minutes, reteach 5 minutes (Allen & Ryan, 1969).

Minicourses combine some of the features of microteaching such as practicing model learning and the use of feedback derived from the observation of the videotapings. Furthermore, some characteristics of programmed instruction are evident, for instance, in independent learning. In order to make minicourses as effective as possible it is of particular interest to study the effects of its various component factors. The problems are partly identical to those encountered in the development of adaptations of microteaching, the best known of which is the minicourse developed in the Far West Laboratory of Education by Flanders. It is a teaching package consisting of sound films and printed materials, which present the model and instructions.

Contents of Practice Teaching

The skills chosen as targets of practice in the new type of practice teaching programs have varied with regard to their degree of specificity and concreteness, cognitive level, the theory on which the choice has been based, etc., in accordance with the set objectives, forms of teaching and resources.

Criteria in the selection of patterns are, for instance, their relationships with student learning: knowledge, skills and attitudes. Which of them we choose to stress in P.E. teaching is a question that is related to our conception of physical education in general. It should be noted that effectiveness thinking is not the same as process-centered thinking. Often expressiveness is a condition for attaining instrumental objectives (see, e.g., Bookhout, 1967), at least in physical education in which social nonverbal communication and the affective element are also emphasized (Heinilä, 1977).

In 1965 isolated technical teaching skills were practiced in the Stanford Laboratory of Microteaching, including intitiation, presentation (communication), consolidation (of the lesson), monitoring and evaluation (Brusling, 1974; Allen, Fortune & Cooper, 1967). They are similar to the basic characteristics of the phase movement of the social interaction process (see, e.g., Bales & Strodbeck, 1967). Allen and Ryan (1969) also give a list of general skills amenable to practice whose application to the teaching of different subjects and different levels of pupils is possible: (1) stimulus variation, (2) set induction, (3) closure, (4) silence and nonverbal cues, (5) reinforcement of student participation, (6) fluency of asking questions, (7) probing questions, (8) higher-order questions, (9) divergent questions, (19) recognizing attending behavior, (11) illustrating and using examples, (12) lecturing, (13) planned repetition, (14) completeness of communication.

In connection with the use of interaction analysis these component-skills refer to the sequence of teacher-pupil interaction and are called "patterns of teaching." A pattern is a short chain of events that can be identified, occurs frequently enough to be of interest, and can be given a label (or name) since this often facilitates thinking (Flanders, 1970).

Some Research Results

The contents and forms of the practice teaching programs of physical education have been studied relatively little. The need to develop new types of practice teaching along the performance-based teacher education lines has been recognized (e.g., Lundgren, 1972; Feingold, 1972; Siedentop, 1972; Jawett & Müller, 1972; Pieron, 1975; Hanke, 1976). Exploratory studies of teaching behaviors in physical education (e.g., Heinilä, 1971, 1974; Varstala, 1973; Pieron, 1975; Hanke, 1976), which used observation instruments derived from Flanders' FIAC system, found that the behavior of teachers and student teachers in physical education was direct (teacher-centered). Pieron (1975) stated this to be the case even when student teachers were familiar with the principles of pupil-centered teaching. Typical of P.E. teacher's speech behavior was also the lack of variation in terms of the features of social interaction and the dominance of teacher talk (e.g., Heinilä, 1974; Reponen, 1979). Flanders (1970) reports 18 research projects which investigate at different levels of education the effectiveness of using interaction analysis as a means to facilitate learning. A general objective of such programs was an awareness of teaching behavior and the development of flexible teaching behavior. Research findings summarized in Flanders (1970) give rise to some generalizations:

1. An individual becomes more responsive to pupil ideas, the amount of open and higher-order questions increases, statement of reasons increases in connection with praise and criticism.
2. Teaching behavior becomes more flexible or variable and more guided by situational factors.
3. The attitudes of student teachers toward the new type of practice teaching become more positive.

Flanders states that "interaction analysis can help to develop value systems about teaching which we call convictions, by contributing information which is primarily objective" (Flanders, 1970, p. 19).

Definition of Problems and Hypotheses

The main elements of the curriculum can be briefly described as follows:

Objectives	Knowledge and mastery as well as cognitive understanding of characteristics of indirect verbal and nonverbal teaching behavior in P.E.
Contents	Teaching models 1-6 (Appendix D.1) Lectures 15 hours (theoretical background of selected models, instrument of observation PEIAC/LH-75 and model demonstrations).
Form of teaching and organization	Practice 30 hours: information, teach one (control) planning of microlesson one, teach one, videotape replay, self-observation, analysis, evaluation and discussion; replanning, reteaching, videotape replay self-observation, analysis, evaluation and comparison of microlessons one and two, summative evaluation.

Material Handout
Task plan, timing, frame factors (teaching model, subject area, pupils age level, competence), lesson plan form (Appendix D1)
Instrument of observation, coding sheet (Appendix D3)
Timeline display (Appendix D.4 and D5)
Model demonstration videotapes.

During the microlessons, the members of the course group (N=5-10) served as pupils for their classmates, then observed the lessons given by all other students on videotape, and took part in the analysis and discussions.

The questions to be answered by the study concerned the form, contents and timing of the course in microteaching. They included:

1. How should students be informed of target behavior?
2. In what way should theory be incorporated into the teaching program?
3. How many microlessons and reteachings are needed?
4. How long should microlessons last?
5. How much time is needed for the analysis of feedback after self-observation?
6. What is the optimum number of pupils in microlessons?
7. Does the constructed observation instrument, PEIAC/LH-75 facilitate model learning and are students able to observe and evaluate their own and others' teaching behavior by means of it?
8. How should the course be placed in the total educational program of P.E. teacher candidates?

The research design was the comparison of two versions of teaching programs and the evaluation of the effect of revisions. By way of hypotheses, it is assumed that at the level of program realization and learning outcomes, the revised program (1976) is more effective than the earlier version (1974).

	1974	1976
(1) information about target behavior	written	written and audio-visual
(2) timing of theory instruction	during the course	during the first third of the course
(3) number of "pupils" in microlessons	4	9
(4) length of microlessons	5 min.	10 min.
(5) number of microlessons	2	3 (of which 1 was used for information and control measurements)

It would have been possible to derive an experimental design on the basis of the above hypotheses for studying the effects of different components. In this exploratory study it was decided to aim at obtaining more global descriptive data. The following null hypotheses were, however, set for testing the differences in the effectiveness of the two programs:

H 1: At the level of program realization there are no statistically significant differences between the teaching behaviors during microlessons 1 and 2 in the two groups (1974 and 1976) in terms of proportional distribution of time in different categories of the PEIAC/LH-75 instrument nor in the selected indices formed on the basis of them (1, 2, 3, 5, and 7).

H 2: At the level of program realization and learning outcomes there are no statistically significant differences between the 1974 and 1976 groups in ratings that concern (1) information about target behavior, (2) timing of theory instruction, (3) number of "pupils" in microlessons, (4) length of microlessons, and (5) number of microlessons.

Research Data and Data Collection

The project focused on the study of the congruence between the objectives of the two microteaching courses held in 1974 and 1976 and the actual outcomes. The subjects were the female and male third year students (1974, N=58; 1976, N=74) at the Faculty of Health and Physical

Table 47. Categories and Main Parameters of Modified PEIAC/LH-75[1] and their calculation

CLUSTER I - TEACHER TALK - PUPIL TALK - SILENT TEACHER ACTIVITY				
		1.		Praises, encourages, accepts the feeling tone of a pupil
		2.		Gives corrective feedback, directs, clarifies, answers pupil´s questions
		3.		Makes use of the ideas and movement patterns suggested by a pupil or group of pupils
	Response		3.1.	Clarifies, expands, builds questions and movement initiations on the ideas expressed by a pupil
			3.2.	Summarizes pupil´s ideas or movement patterns, asks a pupil to demonstrate
			3.3.	Compares the ideas or movement patterns expressed by one pupil to those of another or to those given, repeats pupil´s ideas, asks a pupil to demonstrate
		4.		Asks questions, initiates, terminates activity:
TEACHER TALK			4.1.	Asks questions requiring narrow answers, initiates short-term activity, terminates activity
			4.2.	Broad, open questions which clearly permit choice in ways of answering and moving
		5.		Content emphasis:
			5.1.	Presents information, opinions, demonstrates movement patterns, makes a pupil demonstrate
	Initiation		5.2.	Organizes pupils, material, division of labour and responsibility
		6.		Gives directions, commands during activity (pupils expected to comply)
		7.		Criticizes pupil behaviour, rejects movement pattern, justifies authority
PUPIL TALK		8.		Pupil answers question made by the teacher
		9.		Pupil initiates speech, asks for instructions, expresses own ideas or movement patterns
SILENT TEACHER ACTIVITY AND OTHER		10–12.		(10) Teacher follows pupil´s activity, silent guidance (11), Teacher´s silent participation in movement activity, (12) Confused situation, uproar
CLUSTER II		1.		Pupils collectively passive
PUPILS COLLECTIVE MOVEMENT BEHAVIOR		2.		Pupils collectively active

DEFINITION OF SELECTED INDICES APPEARING IN CONNECTION WITH PEIAC/LH-75

INDICES	1.	Percent teacher talk (TT) =	$\frac{\text{categories } 1,2,31,32,33,41,42,51,52,6,7}{\text{row totals cluster I}} \cdot 100$
	2.	Percent pupil talk (PT) =	$\frac{\text{categories } 8,9}{\text{row totals cluster I}} \cdot 100$
	4.	Teacher´s silent guidance and silent participation in movement activity ratio (TSGPR) =	$\frac{\text{categories } 10,11,12}{\text{categories } 1,2,31,32,33,41,42,51,52,6,7,10,11,12} \cdot 100$
	5.	Teacher response ratio (TRR) =	$\frac{\text{categories } 1,2,31,32,33}{\text{categories } 1,2,31,32,33,6,7} \cdot 100$
	7.	Content emphasis ratio (CCR) =	$\frac{\text{categories } 41,42,51,52}{\text{row totals cluster I}} \cdot 100$

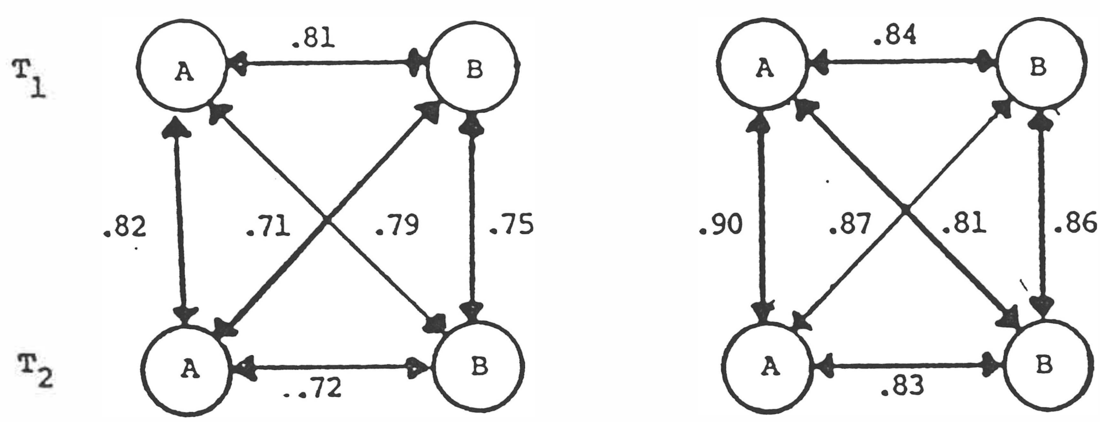
1) PEIAC LH/75, Heinilä 1977, Bulletin, vol 46,1,13-25
Flanders 22-category system, Flanders 1970,140-141

Education of the University of Jyväskylä. During the 1974 course both microlessons of 27 students were videotaped, 54 lessons and in 1976 the microlessons of all students were videotaped 221 microlessons. The teacher was the senior lecturer responsible for the planning of the course.

Two measuring instruments were used in the study. The PEIAC/LH-75 (Heinilä, 1977) observation instrument was used in a modified form for measuring the verbal and nonverbal behavior of the teacher and students (Table 47). The microlessons were recorded by means of the CCTV system of the University. There was a manually controlled camera on site and a camera manipulated from the control room (see Appendix B). The wireless throat microphone used by the teacher recorded the teacher's voice and partly the voices of "pupils." A trained observer observed and coded the videotaped microlessons on computer sheets using the constructed coding instructions (Appendix D.2). Scott's coefficient (Scott, 1955) was used to estimate reliability (Table 48). Towards the end of the 40 hours of

Table 48. Means of Scott's Coefficients for Inter-coder Agreement, Within-coder Constancy, Between-coder Constancy by Cluster (I,II) and by Occasion (T_1,T_2) in Microteaching Observations (N=11 microlessons)

CLUSTER I	CLUSTER II
Teacher Talk Pupil Talk Silent Teacher Activity	Pupils' Collective Movement Activity/Passivity



Coding situations and order of coding:
A=observer 1
B=observer 2
T_1=1st observation
T_2=2nd observation (2 months after T_1)

↔ =inter-coder agreement
↕ =within-coder constancy
⤡ =between-coder constancy

training, program reliability was estimated by means of the degree of agreement between the codings by the researcher and the observer. When the reliability index of a sample of two 60 minute videotaped coding sessions reached the level of .78, the observation of the research data was started.

The questionnaire used classified the student teachers' reactions toward the course program and the form of its realization. The questionnaire consisted of 58 items presented as positive and negative statements to which the experimental population were to react by choosing one of five steps on a scale ranking from "very divergent opinion" (step 1) to "uncertain" (step 3) to "complete agreement" (step 5). Distribution of frequencies were tested by a modified three-step scale with a chi-square test.

Results and Conclusions

Teaching Behavior

The data presented in Table 49 show the results of a one-way analysis of variance for the percentages of category distributions in two clusters and some selected indices based on them indicating the behavior used by students (N=27 and N=74) participating in two different versions of the studied practice teaching program. The data were based on the marks of a reliable observer who coded the events of the videotaped microlessons One and Two, given by students at one week intervals. The coding was done at six-second intervals to both clusters.

In the testing of the hypotheses, statistically significant F-values were obtained in 10 out of 16 analyses of category distributions and in 4 out of 5 analyses of indices. This would indicate that the null hypotheses can be rejected with regard to these dependent variables. These results indicate that the revised course program differed clearly from the first version on the level of realization. On the basis of the differences between indices, the changes can be described in the following fashion: (a) the percentage of teacher talk (TT) decreased (from 76%, 1974, to 68%, 1976), (b) the silent teachers' didactic activities (TSGPR) increased (19%to 29%), (c) the amount of teacher response behavior (TRR) (response ID ratio, by Flanders, 1970) increased (59% to

Table 49. Comparison of the Curriculum Groups 1974 and 1976 on the Percentages of Behavior Used in Microlessons 1 and 2; ANOVA and t-test Computed by Categories of Clusters I and II and by Selected Indices Based on Row Totals

Categories and Indices	Curriculum 1974								Curriculum 1976								Diff. 74-76				F	
	1st Lesson (N=27)		2nd Lesson (N=27)		Total (N=54)		Diff. 1-2 df=52		1st Lesson (N=74)		2nd Lesson (N=74)		Total (N=148)		Diff. 1-2 df=146		Diff. 1-1 df=99		Diff. 2-2 df=99		df=1 df=199	
	x̄	S.D.	x̄	S.D.	x̄	S.D	t	p<	x̄	S.D.	x̄	S.D.	x̄	S.D.	t	p<	t	p<	t	p<		p<
I Cluster																						
1.	5.8	3.0	4.8	3.6	5.3	3.3	-1.07		5.1	3.9	4.6	3.4	4.9	3.7	-.75		-.83		-.22		0.57	
2.	7.2	6.1	7.0	6.[illegible]	7.1	6.2	-.13		5.1	4.7	5.1	4.8	5.1	4.8	-.02		-1.80		-1.56		5.63	.05
3.1.	2.1	3.5	3.7	7.1	2.9	5.7	1.01		12.7	6.7	13.6	6.1	13.2	6.4	.97		7.72	0.001	6.92	.001	107.31	.001
3.2.	0.1	0.5	0.1	0.5	0.2	0.5	0.0		1.4	1.6	2.0	2.5	1.7	2.1	1.58		4.07	0.001	3.80	.001	29.35	.001
3.3	0.8	2.4	0.4	1.2	0.6	1.9	-.71		1.4	2.3	1.9	2.8	1.6	2.6	1.27		1.06		2.61	.05	6.88	.01
4.1.	13.0	7.0	13.1	5.0	13.0	6.0	.09		8.6	4.3	8.5	3.8	8.5	4.1	-.20		-3.76	0.001	-4.96	.001	36.73	.001
4.2.	0.4	1.0	1.0	2.8	0.7	2.1	1.02		1.9	1.4	1.9	1.3	1.9	1.4	-.18		5.25	0.001	2.23	.05	23.66	.001
5.1.	29.6	11.7	27.6	12.2	28.6	11.9	-.64		19.4	8.2	17.1	7.2	18.2	7.8	-1.79		-4.94	0.001	-5.29	.001	52.52	.001
5.2.	5.6	4.4	4.6	3.4	5.1	3.9	-.90		5.1	3.4	4.4	3.0	4.7	3.2	-1.30		-.58		-.28		.39	
6.	9.9	9.7	10.3	12.3	10.1	11.0	.12		2.9	3.4	3.2	3.4	3.0	3.4	.54		-5.45	0.001	-4.54	.001	48.82	.001
7.	2.7	3.1	2.4	3.6	2.6	2.4	-.40		0.9	1.1	1.0	1.4	0.9	1.3	.33		-4.36	0.001	-2.82	.01	24.85	.001
8.	3.4	3.8	2.5	2.4	3.0	3.2	-1.04		2.8	2.3	3.4	2.7	3.1	2.5	1.60		-1.05		1.52		0.08	
9.	2.4	3.2	2.7	2.7	2.6	2.9	.46		1.4	1.6	1.6	1.7	1.6	1.7	0.0		-1.64		-2.57	.05	8.74	.01
10 - 12.	16.9	14.4	19.8	15.0	18.3	14.6	.72		31.2	12.1	31.7	12.6	31.5	12.3	.23		5.00	0.001	4.00	.001	40.47	.001
II Cluster																						
1.	50.2	14.6	47.5	13.0	48.9	13.7	-.73		48.9	12.5	48.2	12.2	48.6	12.3	-.34		-.45		.26		.48	
2.	49.8	14.6	52.5	13.0	51.2	13.7	.73		51.1	12.5	51.8	12.2	51.5	12.3	.34		.45		-.26		.48	
Indices																						
1. (TT)	77.3	13.9	75.0	14.6	76.1	14.2	-.59		64.4	11.0	63.3	11.1	63.9	11.0	-.61		-4.81	0.001	-4.28	.001	41.51	.001
2. (PT)	5.8	5.2	5.3	3.8	5.5	4.5	-.42		4.3	2.8	5.0	3.2	4.7	3.0	1.32		-1.79		-.36		2.42	
4. (TSCPR)	17.7	15.0	20.8	15.5	19.3	15.2	.72		32.5	12.2	33.1	12.7	32.8	12.4	.33		5.04	0.001	4.08	.001	41.53	.001
5. (TRR)	54.4	21.8	60.2	30.8	59.3	26.4	.25		87.4	12.2	86.7	11.6	87.0	11.9	-.33		8.39	0.001	6.30	.001	104.28	.001
7. (CCR)	48.5	12.9	46.2	11.5	47.4	12.2	-.69		35.0	10.7	31.9	10.1	33.4	10.5	-1.82		-5.31	0.001	-6.09	.001	64.81	.001

74%), and (d) the proportion of the content emphasis (CCR) decreased (47% to 42%).

Furthermore, an examination of the F-values and t-values of statistically significant category distribution differences shows that the behavior of the student teachers of the revised course differed in the following ways: (a) the teacher gave less corrective feedback and answered pupils' questions, (b) made much more use of pupil's ideas and movement themes by extending (cat. 3.1.), summarizing (cat. 3.2.) and comparing them (cat. 3.3.), (c) the teacher asked fewer questions which pupils were expected to answer in a given way or initiated and terminated movement activity (cat. 4.1.), (d) the teacher asked more broad and open questions demanding a higher level of thinking which clearly permitted choices in ways of answering and moving (cat. 4.2.), (e) the teacher presented and demonstrated information and his/her own opinions less (cat. 5.1.), (f) the amount of teacher ordering and direction during movements (cat. 6) decreased as well as (g) the amount of criticism and rejection of pupil behavior or movement pattern (cat. 7), and (h) pupil-initiated talk decreased (cat. 9), whereas (i) the amount of teachers' silent didactic activities (cat. 10-12) increased. It is worth mentioning that these changes were not observed to have influenced pupils' collective movement activity, which was just over 50% of the time in the microlessons of both groups. (See Appendix E for a complete comparison by category.)

The two courses differed quite clearly with regard to the above-mentioned respects in terms of both the first and second microlessons, whereas differences between the two lessons within courses were small.

In summary, it may be stated that at the level of realization of the course program, the group whose program had been revised with regard to (1) information about target behavior, (2) timing of theory instruction, (3) number of pupils in the microlessons, (4) length of microlessons, and (5) number of microlessons, displayed more indirect behavior which had been set as a goal. The teacher offered the pupils more opportunities to create ideas and solve problems, was more inclined to observe pupil responses, and took advantage of these responses in the progress of the topic treatment. Pupil-initiated talk did not increase. However, this may be due to the type of pupils who may have been less inclined to "role playing" or the teacher may have directed his main attention to movement ideas and activity.

Student Ratings of the Microteaching Course

From questionnaires filled out by the students (N=121), (Appendix F) data were obtained on student reactions to revisions made. The significance between the percentage distributions of statements was tested by the chi-square test. There were statistically significant differences between the answers given by the students of the two courses. Contrary to the students of the first course, the students attending the revised course were of the opinion that the course could well be placed in the third year program, not before (from 15%, 1974, to 60%, 1976, agree, $X^2 = 25.7$, $p < .001$). The course did not, in the opinion of the students of the revised course, overlap with other teaching (69% to 89% agree, $X^2 = 7.75$, $p < .05$) and they were interested in the theory lessons (20% to 42% agree, $X^2 = 7.95$, $p < .05$). The students of the revised course were more satisfied with the amount of use of audiovisual material (21% to 47% agree, $X^2 = 15.47$, $p < .001$) but they still wanted more. The students of the revised course thought that the time available for exercises was not sufficient, however, they were more satisfied with the time arrangement than the students of the first course (73% to 60% disagree, $X^2 = 6.15$, $p < .05$).

The students of the revised course were less satisfied with the selection of exercises (from 18%, 1974, to 41%, 1976, $X^2 = 9.25$, $p < .01$) but they thought that the exercises were sufficiently varied (27% to 53% agree, $X^2 = 12.9$, $p < .01$). The students attending the revised course were more often of the opinion that the course had opened a new outlook (50% to 60% agree, $X^2 = 6.63$, $p < .001$) and the organization of the course was judged to be better (33% to 69% agree, $X^2 = 4.72$, $p < .001$). The students of the revised course considered themselves to have learned better than the students of the first course to discriminate between teaching patterns in observing and coding feedback (45% to 82% agree, $X^2 = 19.00$, $p < .001$).

In addition, the students of both courses were very satisfied with lecture handouts (in both 1974 and 1976, 92% agree), thought that demonstration tasks had been well selected (70% to 62% agree) and lectures and demonstrations were well coordinated (50% to 69% agree). The students reported that the course had been useful (77% to 87% agree) and that they intended to use in their future practical teaching the teaching patterns they had learned (79% to 82% agree). They also thought that their views of teaching behavior had broadened (77% to 88% agree) and that during the course they had become aware of errors and weaknesses in their teaching behavior (68% to 78% agree).

Summary

Two versions of a practice teaching program have been described and compared. The congruence between the intended and actual outcomes was examined in order to draw conclusions about the rationale of component revision and to provide some basis for the placement of the different modifications of the course in the P.E. teacher education program.

The congruence between objectives, which were identical in both programs, and the degree of their realization was improved in the revised program judging from observation of the students' teaching behavior and their ratings of the courses. The revised program, which included written and videotaped materials, instruction of theory during the early part of the course, and microlessons with nine students and lasting 10 minutes, proved more effective than the original. The students applied better patterns of indirect teaching and were aware of and understood better their theoretical background. The differences between the first and second microlessons of both courses were not significant in terms of any variables. It follows that the number of reteachings should be carefully considered as well as developing their contents and the gradual increasing of level of difficulty.

The instrument of interaction analysis (PEIAC/LH-75 modification) used in the courses was based on an empirical study of physical education teaching (Heinilä, 1974, 1977) and on the theory of Flanders (1965, 1970) and his FIAC system. It proved feasible both from the point of view of research and of teaching. It appeared to facilitate the operationalization, information, evaluation and measurement of intended behavior code patterns. In addition, it helped to teach discrimination and to create teaching patterns stated as objectives.

Recommendations

1. The new form of practice teaching supplies a framework in which research evidence can quite readily feed back into training practice, and in which training practice can open up doors for research in P.E. teaching. Full advantage should be taken of these opportunities.
2. The creation of the new type of teaching practice programs presupposes the construction and testing of different theory based measurement instruments, in other words, improved methods of observation. This should be one of the central tasks of research on P.E. teacher education.
3. The selection of teaching contents and forms should be based on research-based information, i.e., the theory of P.E. Linking these two components together helps bring theory and practice closer to one another.
4. In informing students about target behavior, videotaped materials should be used to supplement written materials. Teaching component-skills should be demonstrated both in natural teaching situations and in microteaching contexts. This would facilitate their recognition and understanding.
5. It is advisable to put the theoretical section at the beginning of the course. Its extent and contents should be carefully considered. As much as possible of the theoretical component should be linked with practical situations by increasing, for example, the time used on the analysis of results of observations and discussions.
6. It would appear desirable that the number of "pupils" in microlessons should be at least eight. The social form of the activity situation is of great importance in physical education. It may be a group work situation, which presupposes a larger number of students. At the same time the teacher's field of observation widens and he is offered opportunities to compare and summarize students' ideas, performances, functions and roles.
7. The length of microlessons should be at least 10 minutes in the P.E. practice teaching, otherwise the situation microlesson may become "truncated" as a process of social interaction. It may lack, for instance, the phase of orientation, activity or evaluation, as

regrettably is often the case in physical education lessons. The stages of orientation and evaluation should be duly attended to.

8. Microteaching is a form of student teaching which aims at improving its effectiveness. Therefore the number of reteachings should be carefully considered. Students who have received training in observation methods are able to adopt and carry out target behaviors more easily than those who have received less training. The contents and form of microlessons might be gradually modified and made more demanding. Microteaching should be closely integrated with other forms of practice teaching and the trying out of different patterns should take place in "natural" teaching conditions and within the framework of longer periods.

Finally, the value of the new forms of practice teaching in P.E. depends on the validity of the chosen forms and contents. Do they effect student teacher behavior as predicted? This is a great challenge for us in our efforts to work for the development of teacher training in physical education.

CHAPTER VIII
SUMMARY AND CONCLUSIONS

Overview

In this chapter the main results of the study are summarized and some conclusions are drawn. The summary first recapitulates the main findings on the primary research problems. Then some of the strengths and limitations of the study are critically discussed. This is followed by an outline of areas suggested for further investigation. Finally, some possible implications of the study for research on teacher education and on the teaching of physical education are discussed.

The main purpose of this study was to develop and test a system for describing instructional procedures in physical education. Its aim was to construct a method for providing good descriptions of teacher-student interactions in P.E. classes, rather than to test theoretical hypotheses or evaluate the effects of such interactions.

Thus the study has a clear methodological orientation. Drawing mainly on interactionist theories of the teaching-learning process and on available research, it sought to develop a theoretically justifiable system for describing and analyzing what happens in the physical education classroom. The second research task was to critically test the reliability and validity of the constructed system. Finally, the third research task was the application of this system to teacher education through curriculum evaluation in microteaching.

The approach used in this study is primarily based on the theoretical and practical work done by Flanders, with reference to his paradigm and the research literature related to the original FIAC system and its several adaptations. The impetus for the present study came from the DPA Helsinki project. Professor Matti Koskenniemi encouraged the author to start an enquiry into interaction in the gymnasium. Assosiate Professor Erkki Komulainen's exhaustive and perceptive methodological studies on classroom observation have served as a model whose sophistication is worthy of emulation but not easy to achieve.

A study of related research literature and consideration of the specific character of physical education indicated a clear need to adapt the FIAC system. Since movement is an intergral part of the instructional processes in P.E. classes, it was obviously necessary to be able to take into account how movement communicates and influences. Consequently, three clusters were included in the developed PEIAC/LH-75 system. The first cluster describes teacher and student talk and teacher's silent activity. The second cluster deals with students' collective movement activity/passivity and social access. The third cluster records the social form of the class. These three clusters contain 12, 8, and 7 categories respectively, altogether 27 categories. Since this cluster arrangement required triple coding, a six-second interval was used instead of the three-second interval employed in the FIAC system. The decision was based on the consideration that three seconds was too short a time for the complex coding required of coders.

The data was collected in such a way that the developed system could be tested in a number of ways. The data used to evaluate the descriptive adequacy of the developed observation schedule and observation training consisted of 24 P.E. lessons, altogether 28,000 six-second time units. The objectivity of coding was assessed by studying the level of agreement between six observers. The sensitivity of the system to faithfully reflect similarities and differences in P.E. classes was studied by including in the 24 lessons four different areas of subject matter (gymnastics, apparatus, rhythmic movement expression, and ball games). For the same reason, lessons from three different grade levels (lower grades: 1-3; middle grades: 4-6; and upper grades: 7-9) were sampled. The construct validity of the system was studied by examining the patterns of data obtained through primary and secondary analyses in the light of the posited model.

The Reliability of PEIAC/LH-75

The first aspect of the reliability of the developed system deals with the objectivity of coding. It was studied in both live and videotaped situations. The results indicated that the intercoder agreement was somewhat higher with the videotaped material than in the live situation. This might be explained by the fact that the situational complexity is reduced in a videotape recording.

The second aspect of reliability dealt with the objectivity of coding in terms of inter-coder agreement, within-coder constancy and between-coder constancy. The method used was Scott's pi coefficient. Summarizing the main results, the average level of mean coefficient values was rather low and varied according to cluster: Cluster I, .61; Cluster II, .65; and Cluster III, .69. The inter-coder agreement was .65, within-coder constancy .69, and between-coder constancy .60 when the two observations of the videotape recordings (T2 and T3) were compared.

The third aspect of reliability focused on reliabilities of the various individual categories, operationalized as inter-coder agreement, and assessed by means of Kendall's coefficient of concordance (W). This analysis indicated that agreement was fairly high, with 23 out of 27 categories yielding a value of W significant at the .01 level (chi-square test). In all coding situations, however, the coefficients of four categories of infrequent occurrence (I/03), and confused situation (I/12, II/8, and III/7) were not statistically significant.

As a fourth aspect, the construct validity of coding was studied using discriminant analysis. The first two of five discriminant functions were statistically highly significant and a third one nearly significant (58%, 21%, and 11% of total discrimination, respectively). The first discriminant function distinguished those observers who made a wide use of the categories of verbal communication from those who used only some of these categories. The second function separated coders by their coding choice in a situation which might be variably interpreted as either confused or as displaying spontaneous student activity. The third discriminant function distinguished coders who described a sequence of verbal and nonverbal communication by using also infrequently occurring categories from those who employed only frequently occurring categories.

The results indicate that there may be an inverse relation between reliability and validity in the case of observation research. Crude coding may be advantageous in terms of reliability, but be detrimental to construct validity. It was concluded that the three-dimensional measuring instrument (PEIAC/LH-75 was reliable when estimated by using a nonparametric coefficient of concordance, W.

The Validity of PEIAC/LH-75

The first aspect of the validity of PEIAC/LH-75 addressed the question of construct validity. To enhance this crucial aspect of all research, a model was developed to define the overall research strategy for the project (see Figure 3). This model served as a guide (1) in specifying the entry situation by defining a valid theoretical and conceptual framework, (2) in constructing a set of exhaustive and mutually exclusive observable behavior categories on the basis of the conceptual framework, (3) in selecting the unit of observation and in developing an adequate coding procedure for accurate use of the system, and (4) in selecting the unit of analysis. The instrument was developed on the basis of a detailed review and analysis of available liturature on research on classroom interaction. This critical survey showed that the Flanders one-dimensional, verbally oriented system needed to be complemented. The feasibility of a multi-dimensional coding system was affirmed in pilot work (Heinilä, 1971).

Construct validity is often determined in an indirect way. The researcher uses a theory to establish a set of hypotheses about how the data should behave. For instance, the researcher predicts certain internal relationships between measured variables: high, intermediate or low correlations. A construct-valid instrument will produce scores that correlate only with those variables with which, on the basis of theory, it should correlate, and the scores of those variables to which it should not be related will not correlate with it (convergent vs. discriminant validity). Similarly, a construct-valid instrument should distinguish between groups that are "known" to behave differently on the construct under study.

In the primary analyses, it was noted that all of the PEIAC/LH-75 categories were used in coding. Thus, the instrument does not appear to contain superfluous categories. Second, 22 statistically significant differences out of the total of 27 categories were found as functions of frame factors: 4 between the two teachers of the sample, 5 between grade levels, and 13 between the various subject areas of physical education classes. Third, matrix analysis showed the interaction sequences to be different in the three clusters, as expected, providing a good description and yielding more information concerning critical teaching behavior. In the first cluster, more than half of all sequence pairs were in the steady state cells while the corresponding figures were more than 80% and more than 90% for Clusters II and III, respectively. This indicates that decisions concerning social form, division of labor and responsibility as well as the forms of students' collective activity/passivity were the general dominating features of teacher behavior.

As another indirect indicator of construct validity, teacher directiveness decreased as a function of grade level while teacher's silent guidance, participation, use of student ideas, and pupil responsibility increased. Also, the variety of critical sequence patterns increased and was stongly related to the content area of physical education.

In the secondary analyses, 18 indices were computed to reduce the primary descriptive analyses. These indices were based on unit coding and the statistical procedures were based on category frequencies, percentages, and ratios. They were computed separately from the matrices of the three clusters. The results indicated that in all 18 parameters of PEIAC/LH-75, statistically significant differences (Mann-Whitney U-test) were found as a function of the key frame factors: teacher (5 statistically significant differences), grade level (6), and subject area of physical education (14).

Two studies conducted by Akkanen (1976, 1979) and Reponen (1979) respectively used the PEIAC/LH-75 instrument. Akkanen's study verified the teaching model. Reponen established that (1) the order of indices revealed differences between two experienced teachers with regard to the rank order of behaviors, (2) the order of indices revealed differences between two groups of student teachers, and (3) the order of indices distinguished between the two experienced teachers and the student teachers.

Factor analysis yielded seven factors which accounted for 68.6% of the total variance. The variables in the factors were concerned with: Factor I, indirect nonverbal integrative idea generation; Factor II, intensity of the teacher's verbal direct guidance; Factor III, uniformity of the teacher's nonverbal guidance as opposed to the specificity of verbal supportive supervision; Factor IV, direction of teacher-pupil communication; Factor V, spontaneous student activity; Factor VI, subject-centricity vs. process centricity; and Factor VII, teacher's response behavior focused on individuals vs. groups.

Grouping analysis was used to relate lessons to the extracted factor dimensions. This made it possible to establish the type of lesson that was most characteristic of each factorial dimension. Through this procedure, we have empirical knowledge of what the lessons were like. The six structurally homogeneous lesson groups, formed on the basis of grouping analysis, were used as the starting point for a further exploration using discriminant analysis.

Five discriminant functions were extracted: DF I, range of ideas for students (closed vs. open); DF II, level of structuring (high vs. low); DF III, level of intensity of guidance (high vs. low); DF IV, level of specificity of nondirective guidance (high vs. low); and DF V, media of nondirective communication (nonverbal vs. verbal and attributing of teacher's response to individuals as opposed to groups).

The results showed that the discriminant functions clearly distinguished between lesson groups. The first discriminant function proved much more powerful than the other four. Its share of the total discrimination of the model was 47%, the four other shares being 19%, 18%, 11%, and 5%, respectively.

The analysis selected 16 out of the total of 27 categories and set them in sequence according to how much they increased the model's discrimination power. The categories of the second cluster (students' collective activity/passivity and social access) and the categories of the third cluster (social form) proved to possess the highest discrimination power.

Through the extensive set of explorations summarized briefly in the above, it was concluded that (1) the instrument possesses a definite degree of construct validity, and that (2) it is sufficiently sensitive to discriminate aspects of direct-nondirect teaching behavior.

The Application of PEIAC/LH-75 to Teacher Education

The final stage of this research project was to apply the instrument that had been developed to the task of training the future teachers of physical education. This was carried out through a microteaching program. A scientific management of the teaching process is set as a goal of the new system of the training of P. E. teachers. Research has indicated that the systems of interaction analysis as tools in teacher education offer better opportunities of achieving this goal, the interaction of theory and practice.

The development of new programmes of practice teaching presupposes the controlling and evaluation of their basic elements. The purpose of this study was to evaluate and compare two curricula, whose purpose was to develop the verbal indirect teaching behaviour of student teachers. The congruence between intended and actually occurring outcomes was studied. The curricula of courses differed in terms of the following elements: (I) information about (models of) target behaviour (written, audiovisual), (II) timing of instruction of theoretical considerations (before/during the course), (III) size of training groups (5-10), (IV) length of microlessons (5-10 min.), and (V) number of microlessons (2-3)

The data cover the courses of microteaching arranged by the faculty in 1974 and 1976 and the subjects were male and female students who started their studies in 1971 (N = 48) and in 1974 (N = 74), 275 microlessons.

The measurement instrument (PEIAC/LH-75, Heinilä 1977) had been constructed for teaching and testing purposes and it was used in a somewhat modified form. It has been derived from Flanders' FIAC-system and contains two clusters, speech and movement, and altogether 18 categories. It made it possible to give information about target behaviour, to operationalize model behaviour and to analyze TV-feedback using a systematic observation method. Reliability (.78) was estimated by means of Scott's pi-coefficient. The category frequencies, indices and student evaluations of courses were compared using analysis of variance and t-test, and chi-square test. (ANOVA)

Statistical comparisons of the outcomes of each course showed clearly that the revised course program differed from the first version on the level of realization. The success of the program was reflected in (a) a decrease of teacher talk, (b) and increase of teachers' silent didactic activities, (c) an increase in teacher response behavior, and (d) a decrease in content emphasis. The increase of indirect behavior was evident in the second session, in which the teachers offered the pupils more opportunities to create ideas and solve problems, observed pupil responses, and took advantage of these responses in the progress of the topic treatment.

The students of both sessions were asked to evaluate the course. A comparison of the responses indicated that, although the students in both sessions were generally pleased with the content, timing and organization of the course, the second group clearly benefited from the revisions that had been made. They felt that the course had opened a new outlook and that they had learned to discriminate between teaching patterns in observing and coding feedback. They thought the course had been useful, making them aware of errors and weaknesses in their teaching behavior, and broadening their views of teaching behavior. Most importantly, they reported that they intended to use the teaching patterns they had learned in their future practical teaching.

Thus, the instrument of interaction analysis (PEIAC/LH-75 modification) used in the courses proved feasible both for research and for teaching. It facilitated the operationalization, information, evaluation and measurement of intended behavior code patterns, and helped to teach discrimination and to create desirable teaching patterns.

Strengths and Weaknesses of the Study

In spite of the many successful aspects of the study, it has several limitations. The most obvious is the limited scope of the empirical data. The data consisted of boys' and girls' P.E. classes at three different grade levels taught by one male and one female teacher and covering four different areas of subject matter, a total of 24 lessons. This would have been a severe limitation if the purpose had been to make a generalizable description of what is happening in P.E. classroom interaction in Finnish schools. Such a description was not, however, the purpose of the study. For the purposes of initial testing of the developed instrument, the data was sufficient.

The major methodological problem of the study was the selection of the length of the time unit. Pilot studies had indicated that the three coded aspects (Clusters) had different natural rhythms. It was not possible to employ the much-used three-second time unit due to the complexity of triple coding. The decision to use a six-second arbitrary time unit to code all three clusters was a compromise made to allow the use of the same time unit in the simultaneous analysis of the whole process. Naturally it was assumed that the aspects with slower tempo, such as the social form and the students' collective movement activity/ passivity, would be reflected in various analyses as dominating features. This assumption was to be explored through many-sided analyses.

Within these limitations, the study has contributed to the area of the study of instructional processes in P.E. classes. An observation instrument and a coding procedure were developed which went beyond the verbal orientation of most classroom interaction studies, and which incorporated features that reflected better the special characteristics of physical education classes. Thus, on the basis of the work done, the instrument can be used to carry out more extensive and representative studies on the nature of interaction in P.E. classes. Since it was clearly demonstrated that the subject matter area was closely related to variation in the kind of classroom interaction, it would be useful to replicate the study with more subject matter areas and with more representative student samples.

Further, the study has highlighted the importance of the quality of teacher-student interaction and student-student interaction in physical education. Attention to this aspect is important if P.E. classes are to have the kind of impact on students' continued interest in physical activity that is a major goal of P.E. teaching in our syllabuses.

Implications for Classroom Teaching

This study was carried out by a former P.E. teacher who has also worked long in teacher education and who has a lifelong commitment to the improvement of teaching. The ultimate motivation for this study was thus to help develop P.E. teaching. Some recommendations can be made on the basis of the work done during the many years of the project.

Teacher education programs in physical education cannot afford to focus too closely on one facet of personality, the psychomotor domain. The cognitive and affective aspects of physical education need to be fully appreciated by future and practicing P.E. teachers. The emphasis on the affective domain, which features prominently in PEIAC/LH-75, seems warranted on the basis of the extensive research on the Flanders system, but this should be ascertained specifically for physical education classes.

Through in-service education, teachers should become familiar with the concept of indirect teacher behavior and its effects on classroom climate and interaction. This should be followed by a demonstration of how classroom interaction can be observed and analyzed. Microteaching in the pre-service training of future P.E. teachers has clearly indicated that this is possible and that it opens a new perspective for students (Heinilä, 1979).

Recommendations for Further Study

During the more than ten-year period of the present study, a serious effort was made to explore a variety of issues and problems related to the empirical study of interaction in P.E. classrooms. However, several technical and methodological problems remain to be explored.

The results suggest that the following questions need to be addressed:

1. The development of rules for coding the verbal and nonverbal communication of teacher and students and of their sequences with a higher degree of specificity is desirable.
2. The optimum length of the coding interval needs careful consideration. A three-second interval might be appropriate in coding the first talk cluster, but a one-minute unit might be more reasonable in the other two clusters.
3. Rules for more decisive coding of students' collective movement activity and the forms of social access (categories II/3 and II/4) need to be developed.
4. The rules for coding students' collective passivity (II/7), waiting for turn, should be refined.
5. The rules for videotape recording need to be determined more exactly, specifying how the total situation is to be filmed.
6. Techniques for voice-recording need to be refined (e.g., using wireless throat microphones), with special attention to the problems of recording student talk.
7. The training of coders needs careful attention, with special

emphasis on the content of training material so that sufficiently varied situations are presented to coder trainees.

8. Agreement controls carried out only at the beginning of coding are not enough to avoid systematic errors in coding. Recurring constancy control needs to be instituted.
9. The criteria for the selection of coders should take into account not only the cognitive but also the affective characteristics of rater candidates.

The empirical findings reported in this study concerning validity and sensitivity established clearly:

1) that in research work in connection with physical education several dimensions describing the influence patterns of the teacher are needed (see Cheffers 1973, 1977; Komulainen 1973),
2) that high frequencies of occurrence are not necessary prerequisites for the discriminant validity and sensitivity of the instrument. Nor should we be deterred from attempting to measure particular behaviors of interest from the point of view of theory on the ground that their occurrence is relatively infrequent,
3) that the aspect represented in the categories of the third and second clusters was found to be the dominating characteristic of the discrimination on the construct under study. It was related to the subject area of p.e., by analysing the formation of homogeneous groups based on factor-scores. But whether it must be so, is another question.

The feasibility of the instrument for different purposes need to be considered more closely. It may be subscripted and postcripted so as to describe different patterns of students and teachers. The cluster can be used singly and/or as entirely as has been done in connection with teacher education programmes, e.g. in microteaching (see Heinilä 1979).

The technique of multiple discriminant analysis outlined in this study was found to be applicable, e.g. for:

1) refining the classification system to that its ability to make the discrimination required for the research problem will be broadened,
2) implementation of observer training programmes so that aspects important from the theoretical point of view can be emphasized,
3) selecting different "criterion groups" and examining factors causing variation among them,

4) studying and assessing "inter-investigation construct validity" by having the same classification system used on different constructs under study and by different investigators.

Thus, the observation instrument and the coding procedures still need refinement. This is to be expected. Flanders (1967) pointed out that "the fact that teaching is a complex social process, hard to define and evaluate, does not mean that all evidence is useless simply because it is incomplete. The tools and techniques to establishing criteria of teaching effectiveness are crude, but they can be improved only by further experimentation and development" (p. 242).

The in-depth study of classroom interaction in P.E. classes in Finland has only begun. (see Laakso, 1984, pp 131-134). This project was mounted in order to get a better grasp of the conceptual and methodological issues and problems. It is to be hoped that this beginning will be greatly extended and intensified by future studies.

REFERENCES

Allen, D. W., & Clark, R. J., Jr. (1967). Microteaching: Its rationale. High School Journal, 51(2), 75-79.

Allen, D. W., & Eve, A. W. (1968). Microteaching. Theory into practice, 7(5), 181-185.

Allen, D. W., Fortune, J. C., & Cooper, J. M. (1974). The Stanford summer microteaching clinic, 1965. In C. Brusling, Microteaching: A concept in development. Stockholm: Almqvist & Wicksel International.

Allen, D. W., & Ryan, L. (1969). Microteaching. Palo Alto, CA: Addison-Wesley.

Akkanen, O. (1976). Use of different teaching patterns based on the analysis of verbal behavior and collective movement activity/passivity of pupils in P.E. classes of junior comprehensive school teachers. In T. Tammivuori (Ed.), Evaluation: International Congress of Physical Education (Report No. 64, pp. 89-95). Helsinki: Finnish Society for Research in Sport and Physical Education.

Amidon, E., & Flanders, N. A. (1967a). The effects of direct and indirect teacher influence on dependent-prone students learning geometry. In E. J. Amidon & J. Hough (Eds.), Interaction analysis: Theory, research and application (pp. 210-216). Reading, Mass: Addison-Wesley.

Amidon, E., & Flanders, N. A. (1967b). The role of the teacher in the classroom. Interaction analysis as a feedback system. In E. J. Amidon & J. B. Hough (Eds.), Interaction analysis: Theory, research and application (pp. 121-149). Reading, Mass: Addison-Wesley.

Amidon, E. J., & Hough, J. B. (Eds.). (1967). Interaction analysis: Theory, research and application. Reading, Mass.: Addison-Wesley.

Amidon, E. J., & Simon, A. (1965, February). Implications for teacher education in interaction research in student teaching. Paper presented at the American Education Research Association, Chicago, IL. (Eric Document Reproduction Service No. Ed 012 695)

Anderson, H. H. (1939). The measurement of domination and of socially integrative behavior in teachers' contacts with children. Child Development, 10, 73-89.

Borgatta, E. F., & Bales, R. F. (1953). Notes on research and teaching: The consistency of subject behaviour and the reliability of scoring in interaction process analysis. American Sociological Review, 18, 566-569.

Brusling, C. (1974). Microteaching: A concept in development. Stockholm: Almqvist & Wicksel International.

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

Cheffers, J. T. F. (1973). The validation of an instrument designed to expand the Flanders System of Interaction Analysis to describe non-verbal interaction, different varieties of teacher behaviour and pupil responses (Doctoral dissertation, Temple University). (University Microfilms No. 73-23, 327)

Cheffers, J. T. F. (1977). Observing teaching systematically. Quest, 28, 17-28.

Cheffers, J. T. F., & Mancini, V. 1978). Teacher-student interaction. In W. G. Anderson & G. T. Barrette (Eds.), What's going in gym: Descriptive studies of physical education classes. Monograph. No. 1, pp. 39-50.

Cohen, J. A. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

Cooley, W. C., & Lohnes, P. R. (1971). Multivariate data analysis. New York: Wiley.

Cooley, W. C., & Lohnes, P. R. (1976). Evaluation research in education: Theory, principles and practice. New York: Wiley.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), Educational Measurement (pp. 443-507). Washington, D.C.: American Council on Education.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). The dependability of behavioral measurement: Generalizability of scores and profiles. New York: Wiley.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 7, 281-302.

Darwin, J. H. (1959). Note on the comparison of several realizations of the Markoff Chain, Biometrika, 46, 412-419.

Dougherty, N. J. (1970). A comparison of the effect of command, task and individual program styles of teaching in the development of physical education fitness and motor skills (Doctoral dissertation, Temple University). (University Microfilms No. 71-10, 813).

Dougherty, N. J. (1971). A plan for the analysis of teacher-pupil interaction in physical education classes. Quest, 15, 39-49.

Dunkin, M. J., & Biddle, B. J. (1974). The study of teaching. New York: Holt, Rinehart & Winston.

Ebel, R. L. (1960) (Ed.). Encyclopedia of Educational Research. Chicago: Rand McNelly.

Emmer, E. T. (1972). Direct observation of classroom behavior. International Review of Education, 18(4), 473-490.

Evertson, C. M., & Green, J. L. (1986). Observation as inquiry and method. In M. C. Wittrock, (Ed.), Handbook of research on teaching (Third Edition) (pp. 162-213). New York: Macmillan Publishing Company.

Feingold, R. S. (1972). The evaluation of the teacher education programs in physical education. Quest, 18(June), 33-40.

Fishman, S., & Anderson, W. (1971). Developing a system for describing teaching: Educational change in the teaching of physical education. Quest, 15, 9-16.

Flanders, N. A. (1965). Teacher influence, pupil attitudes and achievement (HEW, Office of Education Cooperative Research Monograph No. 12). Washington, D.C.: U.S. Government Printing Office.

Flanders, N. A. (1966). Subscripting interaction analysis categories: A 22 category system. Ann Arbor: University of Michigan.

Flanders, N. A. (1967a). The problems of observer training and reliability. In E. J. Amidon & J. Hough (Eds.), Interaction analysis: Theory, research and application (pp. 158-166). Reading, Mass.: Addison-Wesley.

Flanders, N. A. (1967b). Teacher influence in the classroom. In E. J. Amidon & J. Hough (Eds.), Interaction analysis: Theory, research and application (pp. 103-116). Reading, Mass.: Addison-Wesley.

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

Flanders, N. A., & Nuthall, G. (Eds.) 1972). International review of education, 18.4, Special number: The classroom behavior of teachers.

Hamburg: UNESCO Institute of Education.

Flanders, N. A., & Simon, A. (1970). Teaching effectiveness: A review of research 1960-'66. In R. L. Ebel (Ed.), Encyclopedia of Educational Research. Chicago: Rand McNally.

Foa, U. G. (1965). New development in facet design and analysis. Psychological Review, 72, 262-274.

Gage, N. L., (Ed.). (1963a). Handbook of research on teaching. Chicago: Rand McNally.

Gage, N. L. (1963b). Paradigms for research on teaching. In N. L. Gage (Ed.), Handbook of Research on Teaching (pp. 94-141). Chicago: Rand McNally.

Gage, N. L. (1972). Teacher effectiveness and teacher education. Palo Alto, Cal.: Pacific Books.

Galloway, C. M. (1962). An exploratory study of observational procedures for determining teacher nonverbal communication (Doctoral dissertation, University of Florida). (University Microfilms No. 62, 6529).

Galloway, C. M. (1966). Teacher nonverbal communication. Educational Leadership, 24, 55-63.

Galloway, C. M. (1968). Theory into practice. Nonverbal Communication, 7(5), 172-175.

Galloway, C. M. (1970). Teaching in communication, nonverbal language in the classroom (Bulletin No. 29). Association for Student Teaching.

Galloway, C M. (1971). Teaching is more than words. Quest, 15, 67-71.

Gasson, I. S. (1972). The development of an observational instrument to record selected teacher-pupil behaviors in primary school physical education (Doctoral dissertation, The Ohio State University). (University Microfilms No. 71-27473)

Gorman, A. H. (1969). Teachers and learners: The interactive process of education. Boston: Allyn and Bacon.

Guilford, J. . (1948). Factor analysis in a test development paradigm. Psychological Review, 55, 79-94.

Guttman, L. P. (1954). A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), Mathematical thinking in social sciences. Glencoe, Ill: The Free Press.

Hanke, U. (1976). The importance of evaluation in modelling and feedback for the acquisition of teaching skills. In T. Tammivuori (Ed.), Evaluation: International Congress of Physical Education (Report No.

64, pp. 74-81). Helsinki: Finnish Society for Research in Sport and Physical Education.

Hanke, U. (1980a). Die Lehrer-Schuler-Interaction aus handlungs-theoretischer Sicht (Teacher-student interaction from the activity-theoretical point of view). In G. Schilling & W. Bauer (Eds.), Audiovisual Means in Sport (pp. 296-297). Basel: Birkhaus Verlag.

Hanke, U. (1980b). Training des Lehrerverhaltens von Sport-studenten. Ein Vergleich zweier Trainingsverfahren auf der Basis des Microteachings (Training of the teacher behavior of sport students). Inaugural Dissertation, Ruprecht-Karls-Universität zu Heidelberg.

Harris, C. W. (Ed.). (1963). Encyclopedia of educational research (3rd edition). New York: Macmillan.

Heinilä, L. (1970). Opettajan ja oppilaiden välisistä vuoro-vaikutussuhteista liikunnan opetustilanteissa (Report No. 22, pp. 80-94). Helsinki: Liikuntatieteellisen seuran julkaisuja.

Heinilä, L. (1971). Liikunnan opetustapahtuma sosiaalisena vuorovaikutusprosessina (Teaching of physical education as a process of social interaction). Unpublished master's thesis, University of Jyväskylä, Finland.

Heinilä, L. (1974). Developing a system for describing teacher-pupil interaction in physical education classes. FIEP Bulletin, 44(3), 16-20. (Also published in Education physique des enfants avant l'époque de la puberté (1976) (pp. 218-223). Warsaw: Edition Scientifiques de Pologne.)

Heinilä, L. (1976, July). Objectivity of coding in a system (PEIAC/LH-75) developed for describing teacher-pupil interaction in physical education. Paper presented at the International FIEP Congress of Physical Education, Jyväskylä, Finland.

Heinilä, L. (1977). Analysing systems in the evaluation of the teacher-pupil interaction process in physical education classes. FIEP Bulletin, 47(1), 20-34. (Also published in T. Tammivuori (Ed.), Evaluation: International Congress of Physical Education (Report No. 64, pp. 37-58). Helsinki: Finnish Society for Research in Sport and Physical Education.)

Heinilä, L. (1979). Application of interaction analysis to the teacher education in physical education (Research Bulletin No. 15).

Jyväskylä, Finland: University of Jyväskylä, Department of Physical Education.

Heinilä, L. (1980). Developing a system (PEIAC/LH-75) for describing teacher-pupil interaction in physical education classes: Objectivity and content validity of coding. In G. Schilling & W. Bauer (Eds.), Audiovisual Means in Sport (pp. 361-370). Basel: Birkhaus Verlag.

Heinilä, L. (1983). Developing a system (PEIAC/LH-75) for describing teacher-pupil interaction in physical education classes: Construct validity and sensitivity. In R. Telama, V. Varstala, J. Tiainen, L. Laakso & T. Haajanen (Eds.), Research in school physical education (Report No. 38, pp. 124-132). Jyväskylä, Finland: University of Jyväskylä, Formation for Promotion of Culture and

Honigman, F. K. (1970). Multidimensional analysis of classroom interaction (MACI): The Honigman system of interaction analysis. In A. Simon & E. G. Boyer (Eds.), Mirrors for behavior: An anthology of classroom observation instruments (Vol. II). Philadelphia: Research for Better Schools, Inc.

Hough, J., & Ober, R. (1967). The effect of training in interaction analysis on the verbal teaching behavior of pre-service teachers. In E. J. Amidon & J. Hough (Eds.), Interaction analysis: Theory, research and application (pp. 329-345). Reading, Mass.: Addison-Wesley.

Itälä, J. (1969). Koulutussuunnittelu ja koulusuunnittelu (Planning of schooling and school planning). In J. Itälä (Ed.), Koulusuunnittelu (pp. 11-22). Helsinki: Tammi.

Jawett, E., & Mullan, M. R. (1972). A conceptual model for teacher education. Quest, Spring Issue, 18, 76-87.

Kemper, H., Verschuur, R., Ras, J., Snel, J., Splinter, P., & Tavecchio, L. (1976). Development of an instrument for the analysis of the social-emotional teacher-pupil interaction in physical education. In Education physique des enfants avant l'époque de la puberté (pp. 234-239). Warsaw: Edition Scientifiques de Pologne.

Kerlinger, F. N. (1964). Foundations of behavioral research. London: Holt, Rinehart & Winston.

Komiteanmietintö. (1970a). Peruskoulun opetussuunnitelmakomitean mietintö I: Opetussuunnitelman perusteet (No. A4). Helsinki: Valtion painatuskeskus.

Komiteanmietintö. (1970b). Peruskoulun opetussuunnitelmakomitean mietintö II: Oppiaineiden opetussuunnitelmat (No. A5). Helsinki: Valtion painatuskeskus.

Komulainen, E. (1968). Opetustapahtuman tutkimuksesta observointimenetelmällä (On the study of teaching by observa-tion). Unpublished phil.lic. thesis, University of Helsinki, Finland.

Komulainen, E. (1970). Investigations into the instructional process (Vol. II): Objectivity of coding in a modified Flanders Interaction Analysis (Research Bulletin, No. 27). Helsinki, Finland: University of Helsinki, Institute of Education.

Komulainen, E. (1971a). Investigation into the instructional process (Vol. III): P-technique treatment of observational data (Research Bulletin, No. 28). Helsinki, Finland: University of Helsinki, Institute of Education.

Komulainen, E. (1971b) Investigation into the instructional process (Vol. IV): Teaching as a stochastic process (Research Bulletin, No. 29). Helsinki, Finland: University of Helsinki, Institute of Education.

Komulainen, E. (1973). Investigation into the instructional process (Vol. VIII): On the problems of variable construction from Flander's interaction matrix with special emphasis on the stochastic nature of classroom communication (Research Bulletin, No. 34). Helsinki, Finland: University of Helsinki, Institute of Education.

Komulainen, E. (1974). Sattumakorjattujen yksimielisyyskertoimien käytöstä luokitteluun perustuvan tutkimusaineiston yhteydessä (Chance-corrected agreement coefficients applied to nominal data). Tutkimuksia, Helsingin yliopiston kasvatustieteen laitos, No. 33.

Komulainen E. (1974)Opettajankoulutus ja opetusharjoittelun uudet muodot. Suomen kasvatustieteellinen aikakauskirja kasvatus 5, 4, (pp. 227-232).

Komulainen, E. (1978). Developmental change in interaction patterns of the DPA classes. In E. Komulainen & M. Koskenniemi (Eds.), DPA Helsinki Investigations II: Research on Teaching (pp. 17-28). Helsinki: Academy of Finland.

Komulainen, E., & Koskenniemi, M. (Eds.). (1978, October). DPA Helsinki investigations II: Research on teaching. Symposium conducted at the Academy of Finland, Helsinki.

Koskenniemi, M., & Hälinen, K. (1970). Didaktiikka. Helsinki: Otava.

Koskenniemi, M., & Komulainen, E. (1969). Investigation into the instructional process (Vol.I): Some Methodological Problems (Research Bulletin, No. 26). Helsinki, Finland: University of Helsinki, Institute of Education.

Kuhn, T. S. (1962). The structure of scientific revolutions. International Encyclopedia of Unified Science: Vol. 2, No. 2. Chicago: University of Chicago Press.

Laakso, L. (1984) The Norwegian, Swedish and Finnish Student in Physical Education Teacher Training I. Theoretical background, methods and preliminary results. Research Bulletin No. 17. University of Jyväskylä, Department of Physical Education, Finland.

Lamarre, G., & Nygaard, G. (1977, June). A comparison of the effects of command and guided discovery styles of teaching on college students' cognitive achievement and their attitudes toward a winter camping class. Paper presented at the meeting of the International Congress of AIESEP, Madrid.

Levin, L. (1968). Observationer av elevaktiviteter under gym-nastiklektioner (Observations of student activities during physical education lessons) (Report No. 27). Göteborg: Göteborgs Universitet, Pedagogiska Institutionen.

Lewin, K. (1935). Dynamic theory of personality. New York: H. W. Wildon.

Lewin, K., Lippitt, R., & White, R. (1939). Patterns of aggressive behavior in experimentally created 'social climates'. Journal of Social Psychology, 10, 271-299.

Lippitt, R., & White, R. K. (1943). The 'social climate' of children's groups. In R. G. Barker, J. S. Kounin & H. F. Wright (Eds.), Child Behavior and Development. New York: McGraw-Hill.

Locke, L. F. (1977). Research on teaching physical education: New hope for dismal science. Quest, 28, 2-16.

Love, A., & Roderick, J. (1971). Teacher nonverbal communication: The development and field testing of an awareness unit. Theory Into Practice: The Challenge of Nonverbal Awareness, 4, 295-299.

Lundgren, U. P. (1972). Frame factors and teaching process. Stockholm: Almqvist & Wiksell.

Mancuso, J. T. (1973). The verbal and nonverbal interaction between secondary school physical education student teachers and their pupils (Doctoral dissertation, University of Illinois).

McGaw, B., Wardrop, J. L., & Bunda, M. A. (1972). Classroom observation schemes: Where are the errors? American Educational Research Journal, 9(1), 13-27.

Medley, D. M. (1971). The language of teacher behavior: Communication of the results of structures observations to teacher. The Journal of Teacher Education, 22(2), 157-165.

Medley, D. M. (1982). Systematic observation. In H. E. Mitzel (Ed.), Encyclopedia of Educational Research (5th edition) (pp. 1841-1851). New York: The Free Press.

Medley, D. M., & Mitzel, H. E. (1958). A technique for measuring classroom behavior. Journal of Educational Psychology, 49(2), 87-92.

Medley, D. M., & Mitzel, H. E. (1963). Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), Handbook of Research on Teaching (pp. 298-321). Chicago: Rand McNally.

Mitzel, H. E. (1963). Teacher effectiveness. In C. W. Harris (Ed.), Encyclopedia of Educational Research (3rd edition). New York: Macmillan.

Mitzel, H. E. (Ed.). (1982). Encyclopedia of educational research (5th edition). New York: Free Press.

Mosston, M. (1966). Teaching physical education. Columbus, Ohio: Merrill Publishing Company.

Nixon, J. E., & Locke, L. F. (1973). Research on teaching physical education. In R. M. W. Travers (Ed.), Second Handbook of Research on Teaching (pp. 1210-1242). Chicago: Rand McNally.

Nygaard, G. (1978). Three research papers on the analysis of teaching in physical education. In M. Pieron (Ed.), Toward a science of physical education: Teaching analysis (pp. 53-58). Liege: AIESEP.

Pankratz, R. (1967). Verbal interaction patterns in the classrooms of selected physics teachers. In E. J. Amidon & J. B. Hough (Eds.), Interaction analysis: Theory, Research and Application (pp. 189-209). Palo Alto, CA: Addison-Wesley.

Parsons, T. (1968). Social interaction. In D. L. Sills (Ed.), International Encyclopedia of The Social Sciences (vol. 7., pp. 429-441). New York: Macmillan.

Pieron, M. (1975). Etude de la relation professeur-élève en éducation physique par le system de la Hough (Study of teacher-student relationship in physical education using the Hough system). Paper presented to the International Conference on the Pedagogy of Physical Education, Karlsruhe, West Germany.

Pieron, M. (Ed.). (1978). Toward a science of physical education: Teaching analysis. Liege: AIESEP.

Pieron, M. (1983). Teacher and pupil behavior and the interaction process in P.E. classes. In R. Telama (Ed.), Proceedings of the International Symposium on Research in School Physical Education: Research in School Physical Education (Report No. 38, pp. 13-30). Jyväskylä, Finland: University of Jyväskylä, Research Institute for Physical Culture and Health.

Quarterman, J. (1978). A descriptive analysis of teaching physical education in the elementary schools. Doctoral dissertation, Ohio State University, Columbus. (University Microfilms No. 78 12376)

Reponen, P. (1979). Personality and teaching behavior in physical education students. In T. Tammivuori (Ed.), Evaluation: International Congress of Physical Education (Report No. 64, pp. 96-115). Helsinki: The Finnish Society for Research in Sport and Physical Education.

Rosenshine, B. (1971). Teaching behaviours and student achievement. Slough, England: National Foundation for Education in England and Wales.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R. M. V. Travers (Ed.), Second Handbook of Research on Teaching (pp. 122-183). Chicago: Rand McNally.

Rowley, G. I. (1976). The reliability of observational measures. American Educational Research Journal, 13(1), 51-59.

Safrit, M. J. (1973). Evaluation in physical education. Englewood Cliffs, NJ: Prentice-Hall.

Schilling, G., & Bauer, W. (Eds.). (1980). Audiovisual means in sport. Basel: Birkhaus Verlag.

Scott, W. A. (1955). Reliability of content analysis: The case of nominal scale coding. Public Opinion Quarterly, 19, 321-325.

Siedentop, D. (1972). Behavior analysis and teacher training. Quest, Spring Issue, 18, 26-32.

Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. Tokyo: McGraw-Hill.

Sills, D. L. (Ed.) (1968). Interaction. International Encyclopedia of The Social Sciences (vol. 7., pp. 429-470). New York: Macmillan.

Simon, A., & Boyer, E. (Eds.). (1970). Mirrors for behavior: An anthology of observation instruments (Vol. II). Philadelphia: Research for Better Schools, Inc.

Smith, B. O., & Meux, M. O. (1970). A study of the logic of teaching. Chicago: University of Illinois Press.

Soar, R. S. (1968). Optimum teacher-pupil interaction for pupil growth. Educational Leadership Research Supplement, 1, 275-280.

Soar, R. S. (1972). The teacher behaviour related to pupil growth. International Review of Education, 18(4), 508-528.

Splinter, P. G. (1980). Observation of teaching behavior in physical education with the Physical Education Interaction Analysis System (PEIAS). Doctoral dissertation, University of Amsterdam, The Netherlands.

Stake, R. E. (1967). The countenance of educational evaluation. Teacher College Record, 68(7), 523-540.

Stufflebeam, D. L. (1968). Toward a science of educational evaluation. Educational Technology, 30, 5-13.

Tammivuori, T. (Ed.). (1976). Evaluation: International Congress of Physical Education (Report No. 64). Helsinki: The Finnish Society for Research in Sport and Physical Education.

Tavecchio, L. W. C. (1977). Quantification of teaching behavior in physical education. Unpublished doctoral dissertation, University of Amsterdam, Groningon, Holland.

Tavecchio, L. W. C., Splinter, P. G., Kemper, H. C. G., Koos, G. A., Ras, J. S., & Vershuur, R. (1977). Development and application of physical education interaction analysis system. International Journal of Physical Education, 14(1), 12-19.

Telama, R., Varstala, V., Tiainen, J., Laakso, L., & Haajanen, T. (Eds.). (1983). Research in school physical education (Report No. 38). Jyväskylä, Finland: University of Jyväskylä, Formation for Promotion of Culture and Health.

Thorndike, R. L. (Ed.). (1971). Educational measurement. Washington, D.C.: American Council on Education.

Travers, R. M. W. (Ed.). (1973). Second handbook of research on teaching. Chicago: Rand McNally.

Underwood, G. L. (1977). The use of interaction analysis videotape recording in studying teaching behavior in physical education. In T. Tammivuori (Ed.), Evaluation: International Congress of Physical Education (Report No. 64, pp. 59-68). Helsinki: The Finnish Society for Research in Sport and Physical Education.

Valkonen, T. (1971). Haastattelu- ja kyselyaineiston analyysi sosiaalitutkimuksessa II. Helsinki: Korjattu Painos.

Varstala, V. (1973). Opettajan ja oppilaiden toiminta koulun liikuntatunneilla. Unpublished pro gradu of University of Jyväskylä, Finland.

Vuoden 1973 opettajankoulutustoimikunnan mietintö (Report of the 1973 Commission on Teacher Education). Komiteanmietintö 1975:75. Helsinki: Government Printing Office.

Westbury, I., & Bellack, A. (Eds.). (1971). Research into classroom processes: Recent developments and next steps. New York: Teachers College Press.

Withall, J. (1949). The development of a technique for the measurement of social-emotional climate in classroom. Journal of Experimental Education, 17(7), 347-361.

Wittrock, M. C. (Ed.). (1986). Handbook of research on teaching (Third Edition). New York: Macmillan Publishing Company.

TABLE 2 Physical Education Interaction Analysis Category System (PEIAC/LH-75)

		I CLUSTER - TEACHER TALK - PUPIL TALK - SILENT TEACHER ACTIVITY category			II CLUSTER - SOCIAL ACCESS (PUPILS' COLLECTIVE MOVEMENT ACTIVITY/PASSIVITY) category	III CLUSTER - SOCIAL FORM (DIVISION OF LABOUR AND RESPONSIBILITY) category
TEACHER TALK	RESPONSE	01. Accepts, praises, encourages 02. Gives corrective feedback, directs, urges 03. Uses pupils' ideas, accepts, clarifies, develops ideas, movement, tasks suggested by pupils	PUPILS' COLLECTIVE MOVEMENT	ACTIVITY	1. Inter-pupil contacts and movement (space, time, energy) restricted; range of ideas controlled	1. Complete class, uniform task 2. Divided class, uniform task
	INITIATION	04. Asks, initiates and terminates activity 05. Presents information, uses demonstration, describes, organizes pupils/material 06. Gives directions, commands during activity (pupil expected to comply) 07. Criticizes pupil behaviour, rejects movement pattern			2. Inter-pupil contacts and/or movement free; range of ideas controlled 3. Inter-pupil contacts and/or movement free; range of ideas open 4. Pupils' spontaneous activity	3. Divided class, differentiated task 4. Divided class, differentiated tasks distributed amongst groups & within group
PUPIL TALK	INIT./RESP.	08. Answers question/clarifies, demonstrates 09. Initiates speech (asking for instructions expressing own ideas, movements)		PASSIVITY	5. Pupils follow instruction, demonstration	
TEACHER SILENT ACTIVITY		10. Follows pupils' activity, silent guidance 11. Silent participation in movement activity			6. Pupils organize themselves, assist in organization 7. Pupils wait for turn	5. Individual work, uniform task 6. Individual work, differentiated task
OTHER		12. Confused situation, uproar			8. Confused situation, uproar	7. Other situation,confusion
		The decision on classification is made on the basis of the didactic function of the activity.				

L. Heinilä 1975

TABLE 2, Definitions of clusters and instructions for classification

I CLUSTER - TEACHER TALK - PUPIL TALK - SILENT TEACHER ACTIVITY	II CLUSTER - SOCIAL ACCESS (PUPILS' COLLECTIVE MOVEMENT ACTIVITY/PASSIVITY)	III CLUSTER - SOCIAL FORM (DIVISION OF LABOUR AND RESPONSIBILITY
When analysing teacher's authority in use the observation is focused on teacher's and pupil's speech behaviour and the other didactic teacher activity. The decision on classification is made on the basis of the above mentioned didactic function of the teacher activity. Sequence of the actions should be retained. Categories 1-9 The major feature of this category system lies in the analysis of initiative and response which is a characteristic of interaction between two or more individuals. To initiate, in this context, means to make the first move, to lead, to begin, to introduce an idea or concept for the first time, to express one's own will. To respond means to take action after an initiation, to counter, to amplify or react to ideas which have already been expressed, to conform or even to comply to the will expressed by others. Teacher's and pupil's initiative-response behaviour can be directed toward individuals (teacher and/or pupil), group of pupils or the entire class. The behaviour may refer either to the situation, activity or behaviour in the past, in the present or in the future. Categories 10-11 Teacher's silent, purposeful activity is classified into categories 10 and 11. In 10 his role is that of a "teacher's"; in 11 his actions are characterized by an affective identification with the pupils' actions.	Categ. 1-4 Pupils' movement responses By collective activity is meant the movement-activity which has a learning tion. The decision on classification is made through observation of the activity in the entire classand the degree of pupils' freedom in movement, social contacts and range of ideas. Categ. 5-7 Other purposeful activity Collective movement-passivity means that pupils are not moving but are involved in other activity which has a learning function. Categories 1-3 Movement response means the movement-activity which is initiated by teacher's direct or indirect actions based on his own and/or collective decisions. Category 4 Activity is classified as pupils' spontaneous activity when pupils are allowed to move in a certain situation under teacher's supervision and given facilities, teacher assisting and guiding if needed. The problems are set by the pupils.	The observation is aimed at the instructional situation as a whole - at its social form which is considered to appear in division of labour and responsibility. To classify the division of labour and responsibility those behaviours, functions and roles which the group members have during the instructional situation are observed. Behaviours are actions of individual group members expressed in verbal or symbolic terms (eg movement expression). Functions are behaviours directed purposefully toward building the group and toward helping it accomplish its task. Labour: behaviours and functions, which occur in the instructional situation of P.E., may be uniform to all the pupils. Roles mean characteristic playing of certain sets of functions by group members. These functions may be permanent or occasional, more or less conscious. If the tasks are distributed within the group it is the role functions which are often in question. The decision on classification is not only determined by the teacher's but also by the pupils' verbal expressions as a result of which a certain social form is created in the instructional situation.

L.Heinilä 1976

Appendix A3

The procedure of observation (PEIAC/LH-75)

The observer places himself where he can hear and see both the teacher and the pupils, or the video recording on the TV monitor. He observes the first five minutes from the beginning of the lesson without marking the card. The observation period is started and terminated by marking 1287 in the first and last row of the appropriate column. Then every six seconds, either on hearing the signal of by following the hands of the large clock placed on top of the TV receiver, the observer decides which of the three clusters in the classification system the events of the previous six seconds best belong to. The observer writed down the numbers selected while following the events of the next period. Thus he continues for twenty minutes making four digit codings in the appropriate row of the answer card in the six second columns, ten codings per minute. The chronology of the events is retained. A louder signal marks the end of a five minute period, whereupon the observer must continue marking in the first column of the row reserved for the next five minutes.

Where certain events in the observation period have been unclear, this is indicated in the rows (2 vertical lines) at the beginning or end of the said period and a more precise explanation is given at the right-hand edge of the card or on the back. Other features which are necessary for the later interpretation of results are indicated, for example, whether the class was divided, the size of the group observed that was moving etc.

L. Heinilä 1976

APPENDIX A 3

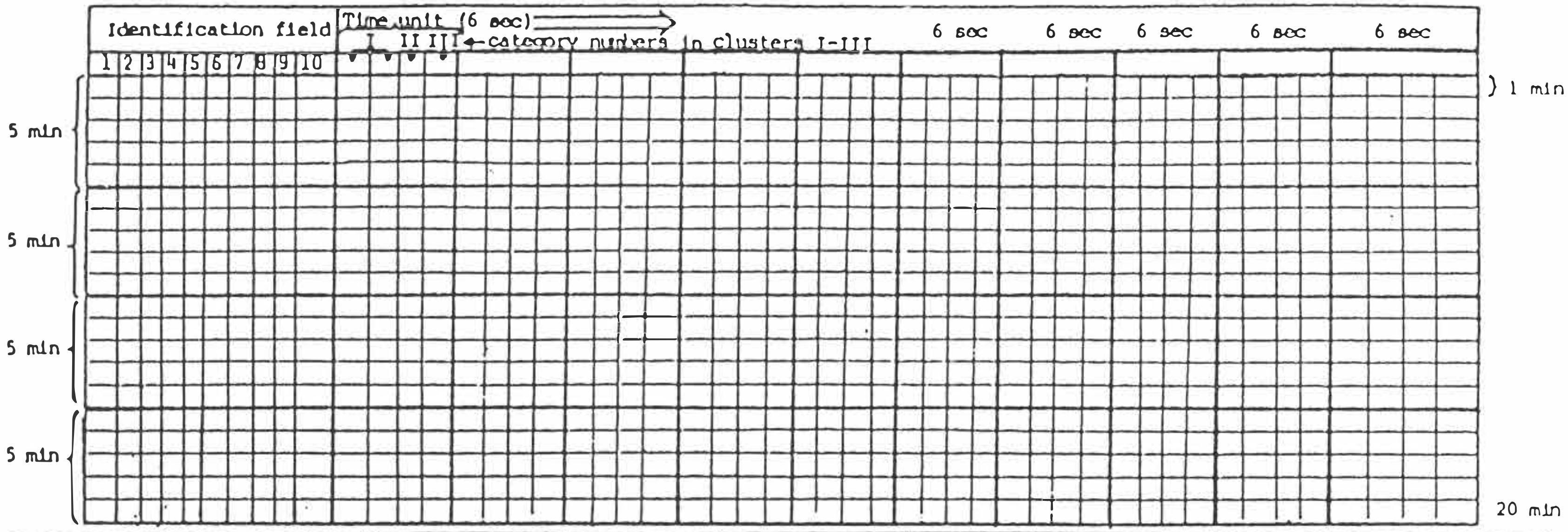
The classification time sheet (see appendix) is the same as an ADP coding sheet where information on the variables connected with lesson material is located in columns 1 - 8, the sequence number of the card in columns 9 - 10, and the observations on the teaching process within the time units in columns 11 - 78.

Before the commencement of the observation period the observer fills in information on the factors below in the first ten columns of the time sheet.

Column:	1.	Observer number (1-6)
	2. & 3.	Situation 01-24
	4.	Classification time: 1. natural situation, 2. video-tape, 3. video-tape, 4. sound tape
	5.	measure 1-9
	6.	Class level: 1. preschool, 2. junior comprehensive, 3. intermediate comprehensive, 4. senior comprehensive 5. sixth form comprehensive, 6. other
	7.	Teacher: 1. man 2. woman
	8.	Subject matter: 1. free gymnastics, 2. apparatus, 3. rhythmic movement expression 4. ball games, 5. basic sport
	9. & 10.	Sequence number of card
	11.-80.	Variables

APPENDIX A 4

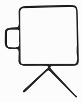
THE CODING SHEET EMPLOYED IN RECORDING (1974)

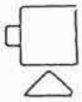


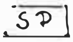
1. Column	2-3. Columns	4. Column	5. Column	6. Column	7. Column	8. Column	9-10. Columns
Coder	Coding situation	Coding occasion	Instrument	Grade level	Teacher	Subject area	Sequence number of punch-card
1 (A) 2 (B) 3 (C) 4 (D) 5 (E) 6 (F)	01 - 24	1 Live situation 2 Video-tape I 3 Video-tape II 4 Sound tape	1 LI/3 clusters 6 sec 2 LI/1.cluster 6 sec 3 LI/2.cluster 6 sec 4 LI/1 & 2 6 sec 5 LI/3 & 2 6 sec 6 LI/2 & 1 6 sec 7 Other	1 Preschool 2 Junior comprehensive 3 Intermediate comprehensive 4 Senior comprehensive 5 Sixth form comprehensive 6 Other	1 Man 2 Woman	1 Gymnastics 2 Apparatus 3 Rhytmic movement expression 4 Ball games 5 Basic sports L.Heinilä	01 - 99 1976

APPENDIX B
AUDIO-VISUAL EQUIPMENT AND ARRANGEMENT
(1974 AND 1976)

1. Symbols used for technical equipment

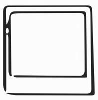

- movable, manual controlled video camera

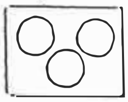

- video camera with remote controlled pan and tilt head and remote controlled lens unit

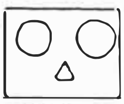


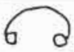
- remote control unit for camera (S) and pulse generator (P)


- remote control box for pan and tilt head and lens unit


- video monitor

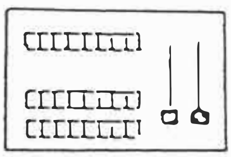

- videotape recorder

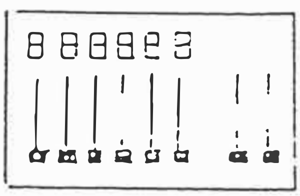

- tape recorder


- headphone


- microphone

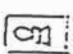

- loudspeaker


- video mixer


- audio mixer

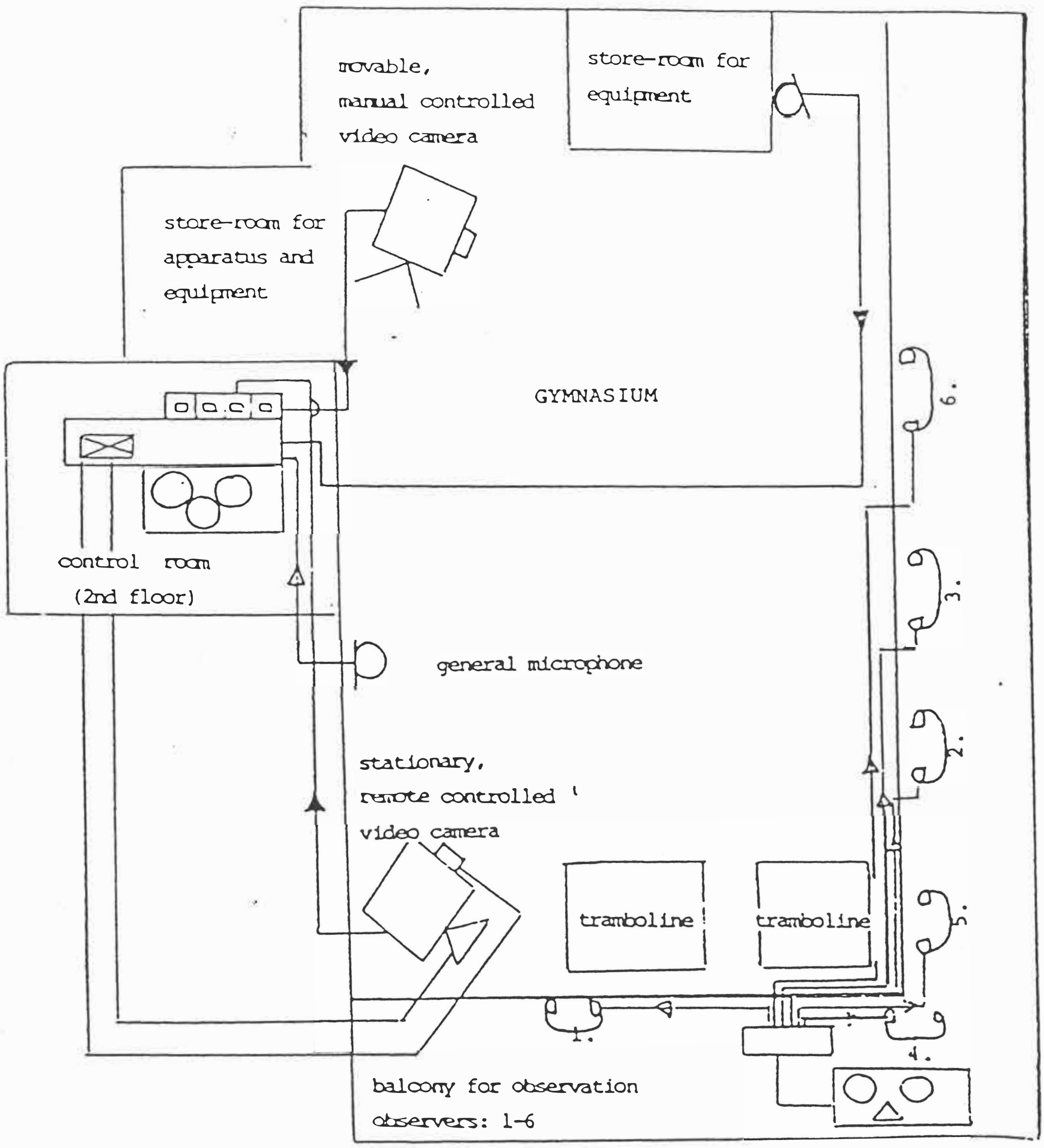

- video signal


- audio signal


- intercom

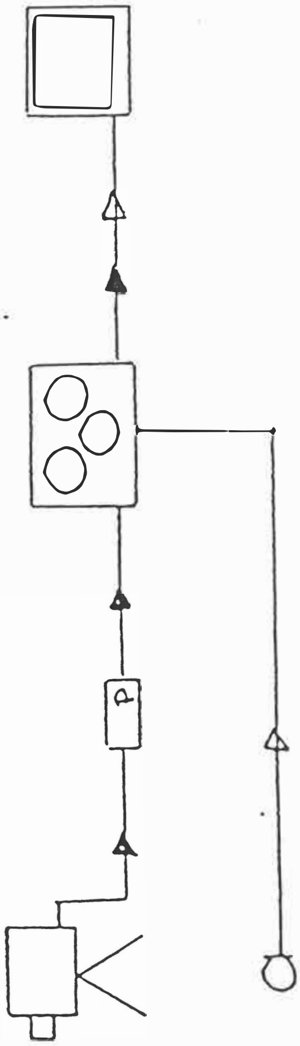
L. Heinilä 1976

2. Placement of video cameras, microphones and observers in the gymnasium



L. Heinilä 1976

3. Scheme of SHIBADEN video equipment (in recording)



L. Heinilä 1976

Categories of the Three Clusters on Correlation Matrix for Observation T_2. The Highest Correlation Coefficient of Each Variable is Placed on the Diagonal.

VARIABLES CLUS. No.	CAT. No.																											
I	01	55																										
	02	55	66																									
	03	-02	05	51																								
	04	-17	-35	-14	54																							
	05	10	32	-02	-19	61																						
	06	02	-10	09	54	-02	72																					
	07	-00	-02	32	06	29	03	52																				
	08	-04	-18	51	45	20	31	52	58																			
	09	-26	18	50	-23	48	05	43	49	58																		
	10	-24	-19	-09	-31	-60	-47	-16	-36	-35	-60																	
	11	-15	-37	-15	01	-31	-08	-34	-26	-30	-21	77																
	12	00	-08	-01	09	-02	12	04	06	-05	-04	-00	39															
II	1	-27	-25	-21	54	-11	72	-21	09	-22	-24	03	30	22														
	2	46	33	13	-44	-17	-51	07	-23	14	54	-38	-19	-65	-65													
	3	-21	-26	-06	-12	-25	-10	-04	-17	-28	-11	77	-08	-18	-41	77												
	4	06	22	17	-10	07	-12	04	-06	54	-14	06	05	-19	13	-03	98											
	5	-20	05	09	41	61	25	03	46	08	-57	-18	01	26	-55	-20	-16	61										
	6	-13	-02	04	-19	58	-14	38	31	58	-08	-45	08	-17	-02	-31	19	25	58									
	7	-44	-39	-08	14	-07	17	08	09	-19	-25	56	21	19	-57	59	-10	07	-06	59								
	8	21	-05	15	13	17	-14	50	58	47	-19	-12	05	-13	01	-15	19	24	26	-02	58							
III	1	07	07	15	36	-14	47	-12	-24	03	-32	18	-18	41	-16	-03	15	-08	-34	-08	-23	-59						
	2	-39	-42	-18	18	21	-12	13	32	08	11	-20	39	18	-20	-19	-21	34	27	07	22	-59	-59					
	3	41	66	06	-36	22	-10	04	02	20	-07	-27	-19	-41	34	-18	10	06	18	-17	-11	-19	-47	66				
	4	30	-06	-23	-23	-35	-21	-10	05	-10	38	03	-08	-12	25	08	-10	-38	-16	-15	25	-26	-06	-18	-38			
	5	-33	-38	19	-20	-29	-22	-09	-24	-10	11	62	-12	-25	-15	76	06	-35	-20	48	-13	13	-24	-31	-08	76		
	6	-17	03	51	40	-07	01	19	42	-06	04	-15	-08	-15	-02	04	-07	22	-00	16	-11	-08	-05	22	-13	-10	51	
	7	10	28	12	-11	10	-07	-00	-05	52	-16	00	07	-15	12	-10	98	-10	18	-15	17	09	-18	15	-05	-03	-05	98
		01	02	03	04	05	06	07	08	09	10	11	12	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7
		I												II								III						

-257-

Categories of the Three Clusters on Correlation Matrix for Observation T_3. The Highest Correlation Coefficient of Each Variable is Placed on the Diagonal.

VARIABLES CLUS. No.	CAT. No.																											
I	01	47																										
	02	47	64																									
	03	03	-10	57																								
	04	-16	-29	-03	65																							
	05	23	42	-05	-28	-63																						
	06	-03	-02	-15	65	-01	76																					
	07	06	-12	-13	16	20	17	64																				
	08	11	-20	17	45	07	24	64	64																			
	09	16	03	11	-12	24	-03	40	49	53																		
	10	-31	-25	-21	-26	-63	-43	-15	-29	-15	-63																	
	11	-10	-33	42	-03	-26	-13	-35	-19	-26	-28	80																
	12	08	-05	-08	11	-00	22	16	14	-13	-02	-15	33															
II	1	-19	-17	-02	61	-19	76	-05	16	-10	-19	-06	32	76														
	2	38	27	-33	-54	-02	-57	-12	-26	17	50	-33	-18	-63	-63													
	3	-11	-27	57	-07	-29	-14	-08	-15	-31	-14	80	-11	-18	-42	82												
	4	06	03	-05	-11	05	-10	27	-05	39	-06	04	-03	-23	07	14	71											
	5	-30	10	-00	43	51	36	17	40	-04	-56	-14	00	25	-60	-14	-24	-60										
	6	-12	-10	-20	-19	-41	-18	50	24	45	07	-48	15	-23	09	-35	28	15	50									
	7	-34	-37	11	12	-12	-05	05	12	13	-26	64	-16	-06	-42	51	36	16	-02	64								
	8	24	20	-12	-13	07	-08	06	04	53	-09	-05	15	-07	12	-13	65	-13	16	19	97							
III	1	10	13	13	34	-05	47	-21	-20	00	-34	15	-20	41	-14	-04	16	-09	-34	-02	-04	-55						
	2	-43	-42	-11	19	02	-04	29	37	-01	16	-18	33	16	-20	-15	-15	30	22	-06	02	-55	-55					
	3	37	64	-14	-34	35	-17	-10	-08	00	-11	-24	00	-38	28	-16	-07	08	23	-07	01	-25	-50	64				
	4	21	-21	-11	-35	-25	-29	-02	-12	02	48	-09	-21	-20	38	-04	-13	-39	-05	-20	10	-35	01	-13	48			
	5	-22	-36	40	-10	-39	-21	-04	-18	-15	04	72	-12	-22	-22	82	27	-28	-25	54	-14	11	-20	-28	-12	82		
	6	-03	-05	07	46	-10	06	25	36	00	-06	-13	-08	-13	-12	10	-09	28	-01	17	-08	-15	-01	21	-12	-06	46	
	7	25	26	-12	-09	10	-05	05	04	53	-15	-03	01	-11	13	-11	71	-09	15	22	97	06	-10	08	01	-12	-05	97
		01	02	03	04	05	06	07	08	09	10	11	12	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7
		I												II								III						

APPENDIX C2

Symmetric Transformation Analysis: Varimax-rotated factors transformation matrices for three occasions T_1, T_2, T_3

	Factor No	T_1 1	2	3	4	5	6	7
T_2	1	95	-00	01	-04	08	30	01
	2	05	06	-13	99	-05	00	-08
	3	-05	98	-10	-07	04	15	06
	4	-03	-10	-27	05	26	03	92
	5	06	05	10	04	88	-42	-20
	6	11	03	91	14	08	29	23
	7	28	17	26	01	-38	-79	23

		T_2 1	2	3	4	5	6	7
T_3	1	82	01	06	-01	11	56	-03
	2	-01	-12	-01	99	-05	05	-01
	3	-15	-12	-20	03	95	06	-15
	4	-28	86	-08	08	00	41	11
	5	-19	-05	96	01	14	15	06
	6	32	20	09	06	-01	-43	82
	7	32	44	17	09	26	-56	-54

		T_3 1	2	3	4	5	6	7
T_2	1	98	-04	-01	-11	-11	10	-06
	2	04	98	03	15	-15	-02	-01
	3	08	-17	10	96	-14	07	-05
	4	-08	02	19	-07	-04	94	28
	5	01	-02	96	-10	01	-17	-10
	6	12	13	03	14	97	07	-06
	7	-10	00	-59	-06	-05	28	-95

APPENDIX C3

Regression coefficients employed in estimating factor scores (Ledermans' method)

a. Occasion T_1 (live situation)

Cluster	Cat.	1	2	3	4	5	6	7
I	01	-01	-28	-09	-08	-00	04	10
	02	02	-22	06	01	-01	05	-08
	03	04	00	01	-01	-12	-26	03
	04	-01	-04	-03	-21	-02	05	14
	05	01	03	28	12	01	-01	04
	06	01	-02	-01	-09	-01	-03	10
	07	-01	-01	-05	-02	-01	-04	25
	08	-01	-02	-02	-03	-01	02	18
	09	07	03	07	06	02	-23	14
	10	16	17	-19	12	-01	-02	-20
	11	-31	-02	06	-01	04	08	-25
	12	-17	01	03	08	03	13	-04
II	1	03	06	-04	-27	04	05	-04
	2	16	-06	-04	19	-04	-11	-12
	3	-27	-11	-06	01	-04	03	21
	4	-00	-03	-04	-02	32	-06	04
	5	00	19	33	07	02	11	-12
	6	06	10	02	13	02	-00	11
	7	-03	-06	-01	00	-01	-00	07
	8	-05	-01	-12	01	03	05	14
III	1	09	-03	11	-26	00	-22	-21
	2	01	14	-04	03	02	10	09
	3	-06	-29	11	15	-05	12	02
	4	-00	04	-18	06	03	08	04
	5	-18	12	01	19	-06	-46	-01
	6	-04	-05	-02	-01	-02	05	10
	7	-03	07	-02	-02	64	13	-15
Scalarproduct of factor loadings		.92	.89	.91	.89	.95	.87	.81

APPENDIX C3

c. Occasion T_3 (second videorecorded observation)

Cluster	Cat.	1	2	3	4	5	6	7
I	01	-02	-04	-03	-22	-06	13	06
	02	06	01	-01	-20	11	-10	-04
	03	-07	-02	-03	-08	-01	12	09
	04	03	-23	01	-02	-07	03	-41
	05	03	13	-02	00	32	05	26
	06	04	-24	00	-07	02	-00	-01
	07	-03	02	-01	-01	-03	23	-01
	08	-06	-03	-04	-08	-06	44	-08
	09	-03	-00	02	-07	-04	21	08
	10	10	07	02	19	-29	-13	-15
	11	-26	-02	03	00	06	-12	11
	12	01	-01	00	04	01	-00	04
II	1	07	-34	01	10	-01	-08	24
	2	07	08	-02	-18	-21	04	-01
	3	-30	05	-00	-05	02	04	-05
	4	-04	01	08	02	01	03	01
	5	05	05	-00	15	27	-06	-14
	6	03	09	01	09	05	07	01
	7	-08	06	07	12	09	-07	-20
	8	04	-02	33	12	-02	-00	16
III	1	-03	-15	-00	-16	-03	-02	07
	2	04	06	01	29	03	05	11
	3	06	11	-02	-13	12	-11	-27
	4	00	01	-01	-02	-09	05	04
	5	-24	02	02	00	-06	03	00
	6	00	01	-01	-05	-02	04	-31
	7	05	-06	58	-04	02	-13	-31
Scalarproduct of factor loadings		.94	.92	.96	.86	.89	.84	.77

APPENDIX C4

Table Correlation matrices of original groups estimated on the discrimination functions 1-5

Group 1.

	1	2	3	4	5
2	.26				
3	-.30	-.32			
4	-.14	-.59	.05		
5	-.12	-.20	-.29	-.05	

Group 2.

	1	2	3	4	5
2	.31				
3	.24	.73			
4	.22	.92	.83		
5	-.32	.77	.56	.81	

Group 3.

	1	2	3	4	5
2	.14				
3	-.40	-.06			
4	.10	-.62	-.07		
5	-.21	-.41	-.26	.10	

Group 4.

	1	2	3	4	5
2	-.04				
3	.38	-.08			
4	-.21	.32	.01		
5	-.39	.11	.07	-.46	

Group 5.

	1	2	3	4	5
2	.10				
3	-.36	.06			
4	.25	-.53	-.62		
5	-.19	-.77	-.25	.23	

Group 6.

	1	2	3	4	5
2	-.42				
3	.58	-.25			
4	.30	.28	.40		
5	.61	.30	.49	.22	

APPENDIX D1
MAIN ELEMENTS OF THE MODIFIED CURRICULUM (1976)
TEACHING MODELS IN PHYSICAL EDUCATION

1. GENERAL

1.1. The Concept of teaching model

During the teaching situation the same series of events is often repeated again and again. This kind of sequence can be called a teaching model. It can be defined as a short chain of events that can be identified, occurs frequently enough to be of interest, and can be given a label (or name) since this often facilitates thinking (Flanders 1970).

Teaching model must not be connected with subject matter but with the process behavior. By using a model features in teaching behavior, either stable or occasional, can be described. There are some teachers who are able to produce several different teaching models, and others who use only few models. The former are noticed to be flexible and their work is more productive. It is also noticed that certain models of teaching behavior are connected with certain positive or negative pupil responses and attitudes. The teaching model is also a more concrete term than, for example, teacher's role or teaching method. It can be limited to a specific area of teaching behavior, for example to speech behavior. Purposeful changing of teaching models within a longer period of time is called teaching strategy.

Teaching models can be identified with the help of the process analysis technique available. Among other things the systematic observation method enables the quantification and measurement of the features in teaching-learning process. On the basis of frequencies behavior can be placed in certain dimensions such as teacher initiation - pupil initiation, teacher-centeredness - pupil-centeredness and direct - nondirect.

The applied process analysis system (measuring instrument), and the theoretical basis of teaching models must be related to each other. The teaching models in physical education to be described, and measured, are based on the PEIAC/LH-75 system (Heinilä 1976) (appendix 1) developed and enlarged from the FIAC system (Flanders 1970). It is justified to describe teaching models, in physical education, by using a modified form of the multidimensional system. (Appendix D3, p. 268)

2. TEACHING MODELS IN PHYSICAL EDUCATION

I will describe some teaching models which are central in physical education. I will further operationalize these models into skills which can be practised and learned. The models will be focused to process readiness. The starting point for the selection of the models is the present knowledge of teaching physical education. It is defined as an interaction process within school surroundings, and is aimed at the promotion of personal development of the pupils in accordance with the set educational objectives (Koskenniemi & Hälinen 1970, 101). In interaction process a social system moves from one state to another as a function of time (Komulainen 1970, 1) (Heinilä 1970, 1976, 1977).

Teaching models are centered, in the first place, around teacher's speech behavior which is considered to be the most significant single factor affecting the teaching-learning process (Amidon & Hough 1967). The teaching-learning process in physical education is, however, exceptional. Non-verbal communication (e.g. movement) has a central position in it. Movement communicates and affects. It is a means and an end. It's movement which is often the answer of a pupil. Even movement of the teacher, and other activity, significantly affect the formulation of the social structure and the sequence of events in the process. (Heinilä 1970, 1976, 1977).

The teaching models selected can be roughly divided into the models of direct and indirect teaching. In interaction analysis this refers to the balance of teacher initiation and response. In the direct way of influence teacher, initiation is stressed. This usually restricts pupil's freedom of action, while the indirect teaching tends to increase it.

Out of the 9 models selected for description 6 are models of indirect teaching. The practice will be focused on these models. The starting point is, however, the mastery of direct models - the basic elements in teaching. Physical education has traditionally been subject matter centered and direct method - the command method - has been the main method. The school reform and the notion of the teaching-learning process, however, presuppose the mastery of a more pupil-centered and pupil-initiative way of teaching in physical education. For this reason, it is justified to concentrate in practicing indirect teaching models in physical education teacher-training.

APPENDIX D2

1. The indirect teaching models used

1. Teacher initiatives based on pupil responses
The P.E. teacher has to be able to make use of pupils' earlier performance or initiatives by making questions and suggestions related to them or by making the pupil demonstrate his performance. The teacher must then clarify essential points.

2. Summarizing model
The P.E. teacher has to be able to summarize what pupils have done or said and then proceed to the next logical stage by making use of the summary. He can also make pupils demonstrate the functional solutions of the sub-stage and describe them verbally. This is effective reinforcement of pupils' initiative.

3. Comparison model
The P.E. teacher has to be able to observe and compare pupils' movements or their previous ideas to other pupils' movements or given task requirements. In this way the teacher can help pupils to solve problems and guide them to identify key ideas while showing or giving the pupils the impression that they solved the problems on their own. This kind of teacher activity, in which pupils' performance is informed or described to other pupils serves to reinforce their initiative and independent behavior.

4. Model of guiding feedback
The P.E. teacher has to be able to give guiding feedback to the whole class, smaller groups and individual pupils. The giving of feedback presupposes exact definition of objectives and tasks. The teacher has to be able to give feedback wisely, in a variety of ways and giving reasons for his statements. The use of guiding feedback is common in physical education. For instance, in the teaching of some "closed" motor skill (in given circumstances and restricted) it has a decisive role. The role of guiding feedback is to help a pupil to become aware of his performance and to find solutions to problems concerning e.g. movement paths, timing,use of power or space. Giving guiding feedback with statement of reasons for it willl help to promote independence. The teacher has tried to see the pupil as a person with whom things can be discussed and planned before decisions are made. The pupil can thus be guided towards a goal which he understands and accepts.

5. Model of reinforcement and extinction
The P.E. teacher has to be able to observe -- to watch and listen to -- pupils' ideas and movements with a view to organizing them in terms of teaching objectives and to reinforce selectively those ideas and movements which are on-target. The teacher also has to be able to state without hesitation and clearly what is not relevant or useful from the point of view of the teaching objective. Such responses may be directed to the whole class, to smaller groups or to individual pupils. Praise and reward and criticism may concern pupil's behavior or movements. Praise can be verbal but also symbolic (e.g. smile, applause), similarly rejection. In acting on pupils' conditions the teacher's reasons must be related

either to the whole class, to groups of pupils or to individual pupils.

6. Discrimination model
   The P.E. teacher has to be able to clarify -- verbally and through demonstration -- the logic of classroom discourse and progress. For instance, he can clarify the pupils' degree of freedom of social activities by stating given or accepted directions -- customs, norms, rules of the game, etc. This includes the maintenance of a consistent meaning of words, concepts and movements. Accurate concepts aid communication and classroom discipline is improved. It is especially important to help pupils to distinguish between facts, opinions and valuations. This presupposes that the teacher monitors and evaluates the situation.

TABLE 47 b. Specified Classification System for Physical Education Interaction Process: Cluster I (PEIAC/LH-75)

Teacher talk	Response	1.	Praises, encourages, accepts the feeling tone of a pupil
Teacher talk	Response	2.	Gives corrective feedback, directs, clarifies, answers pupil´s questions
Teacher talk	Response	3.1.	Makes use of the ideas and movement patterns suggested by a pupil: clarifies, expands, builds questions and movement initiations on the ideas expressed by a pupil
Teacher talk	Response	3.2.	Summarizes pupil's ideas or movement patterns, asks a pupil to demonstrate
Teacher talk	Response	3.3.	Compares the ideas or movement patterns expressed by one pupil to those of another or to those given, repeats pupil's ideas, asks a pupil to demonstrate
Teacher talk	Question	4.1.	Asks questions requiring narrow answers, initiates short-term activity, terminates activity
Teacher talk	Question	4.2.	Makes questions requiring higher level of thinking or activity
Teacher talk	Initiation	5.1.	Presents information, opinions, demonstrates movement patterns, makes a pupil demonstrate
Teacher talk	Initiation	5.2.	Organizes pupils, material, division of labour and responsibility
Teacher talk	Initiation	6.	Gives directions, commands during activity (pupils expected to comply)
Teacher talk	Initiation	7.	Criticizes pupil behaviour, rejects movement pattern, justifies authority
Pupil talk	Resp.	8.	Pupil answers question made by the teacher
Pupil talk	Init.	9.	Pupil initiates speech, asks for instructions, expresses own ideas or movements
Other	Silence, confused situation	10-12	(10) Teacher follows pupils' activity, silent guidance (11) Teacher's silent participation in movement activity (12) Confused situation, uproar
The decision on classification is made on the basis of the didactic function of the activity.			

2. Coding instructions and coding sheet (Appendix D4)

Instructions for Classification

Before the beginning of the observation period the observer enters on the reverse side of the form data on teaching situation.

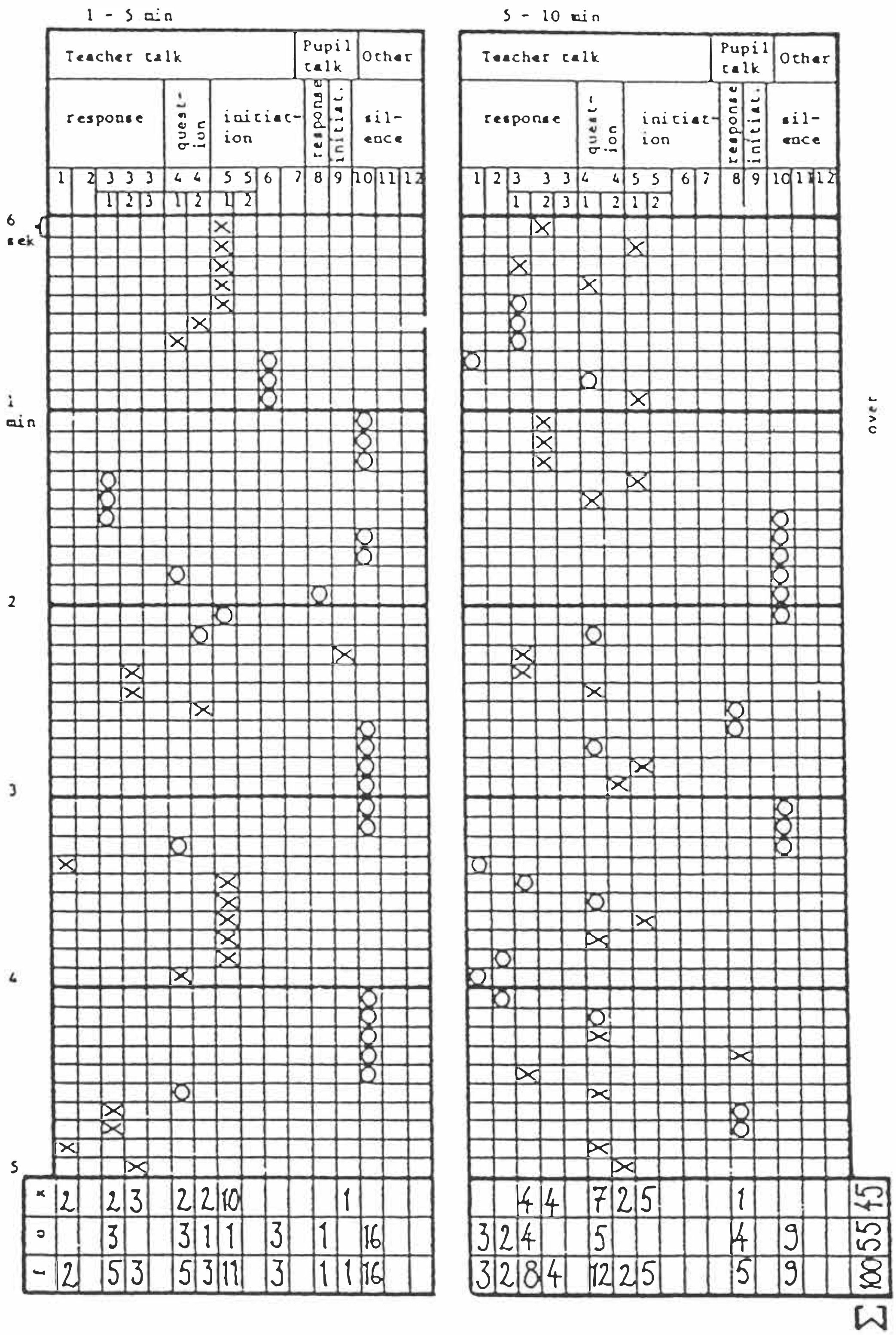
The observer places himself where he can hear and see well the TV-display. Every sixth second, either on hearing a signal or observing the clock placed on top of the TV set, he decides which of the categories of the classificaiton system best represents the events of the previous six-second period. The observer directs his attention to the speech and movement behaviors of the teacher and students. Students' movement behavior is viewed collectively. The observer marks the relevant category column entering either 0 or X depending on whether the class was active or passive in terms of movement during the six-second period. At the same time he observes what is happening during the next period. This produces 10 entries per minute and 100 entries in ten minutes. At every full minute timing should be checked. Entries on the form constitute a series going from top to bottom, which preserves the sequence of events. Categories have been placed on the form so that entries yield an immediate basis, for example, for (I) a visual evaluation of teacher's initiating and response behavior, (II) a general idea of the amount of movement, (III) a general idea of the stability and variability of the process, and (IV) a general picture of nature of points.

At the end of the observation period, column totals are computed and entered on the form, separately of 0 and X and combined. The column totals of 0, X and F are also added up. After that computations for obtaining indices (on the back page of the form) are carried out. The obtained results are used in analysing, comparing and evaluating micro lessons in relation to set objectives.

(Heinilä, 1977)

APPENDIX D4

Coding Sheet: Time Line Display (1976)



X = collective movement passivity
O = collective movement activity

(Heinilä 1977)

APPENDIX D5

Microteaching course/Heinilä, L. -75

Class n:o 1 2

Student (subject) n:o ___ Name: ________________ Model n:o _____ Date: ______

Class information: age of pupils ___y. skill level ____ Subject matter _________

Calculate the following indices:

1) Percent teacher talk $= \frac{100 - (8+9+10+11+12)}{100} \times 100$

2) Percent pupil activity $= \frac{100 - \Sigma X}{100} \times 100$

3) Teacher response ratio $= \frac{1+2+3+11}{1+2+3+6+7+11} \times 100$

4) Occurance of models (frequencies)

5) Percent model occurance $= \frac{\text{categories in the model}}{100} \times 100$

6) Intensity of teacher quidance $= \frac{4+6}{1+2+3+4+5+6+7} \times 100$

Observation instructions for teaching models:

Way of teaching	Model n:O	Category n:o
Direct teaching		
Information presentation model	1.1.	5.1., 4, 6
Organization model	1.2.	5.2., 4, 6
Initiation variation model	1.3.	4.1., 4.2., 5.1.
Indirect teaching		
Teacher initiations based on pupils responses	2.1.	3.1.
Reinforcing pupil initiations, summarizing model	2.2.	3.2.
Comparison making model	2.3.	3.3.
Accepting pupil's feeling -model	2.4.	1, 2
Corrective feedback -model	2.5.	2, 3.1.
Changing the level of abstraction	2.6.	1, 2, 3, 5.1.
Teacher's reinforcing and extinguishing reactive behavior	2.7.	1, 2, 3, 5.2.
Discrimination making model	2.8.	2, 3, 4, 5

Analyse: ______________________________

____

Suggestions for improvements: ______________________________

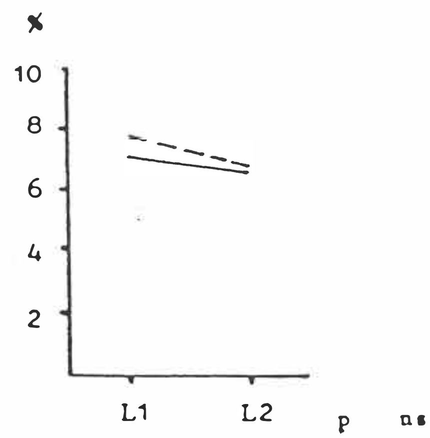
____

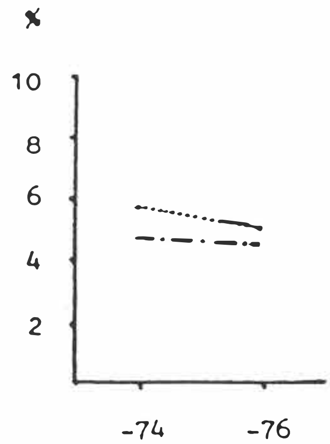
______________________ date

______________________ signature

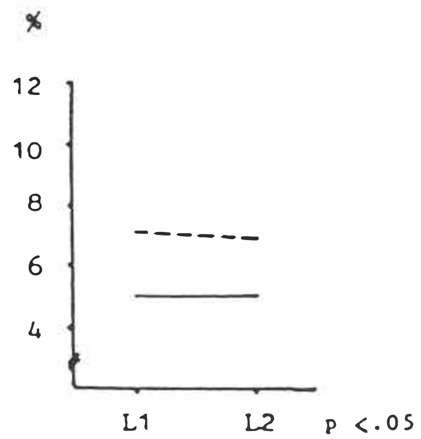
APPENDIX E
COMPARISON OF PROGRAM 1 (1974) AND PROGRAM 2 (1976)
BY CATEGORY AND INDEX

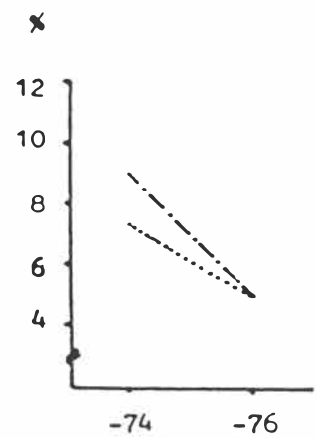
Cat. 1. Praises, encourages, accepts the feeling tone of a pupil



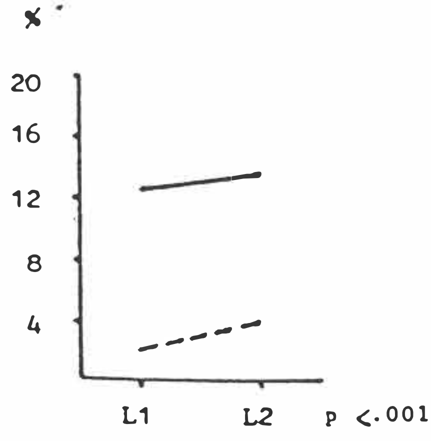


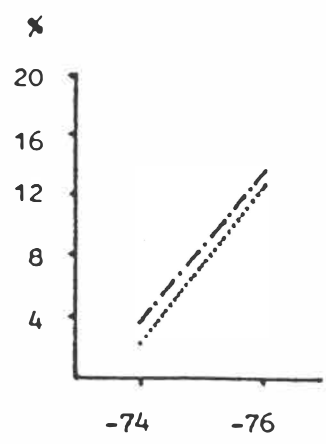
Cat. 2. Gives corrective feedback, directs, clarifies, answers pupil's questions





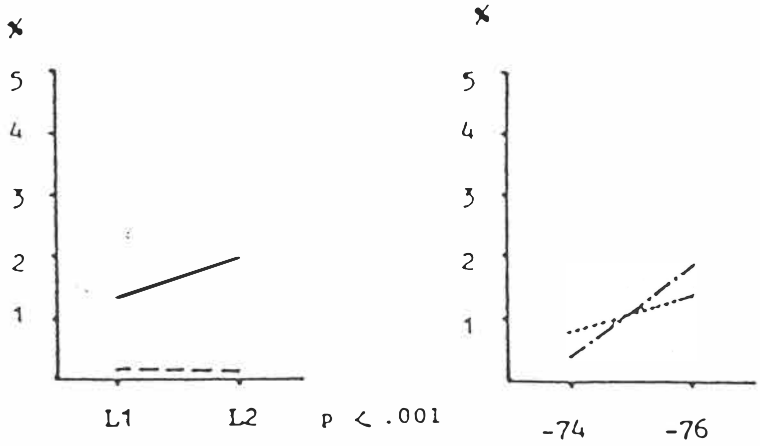
Cat. 3.1. Makes use of the ideas and movement patterns suggested by a pupil or group of pupils. Clarifies, expands, builds questions and movement initiations on the ideas expressed by a pupil.



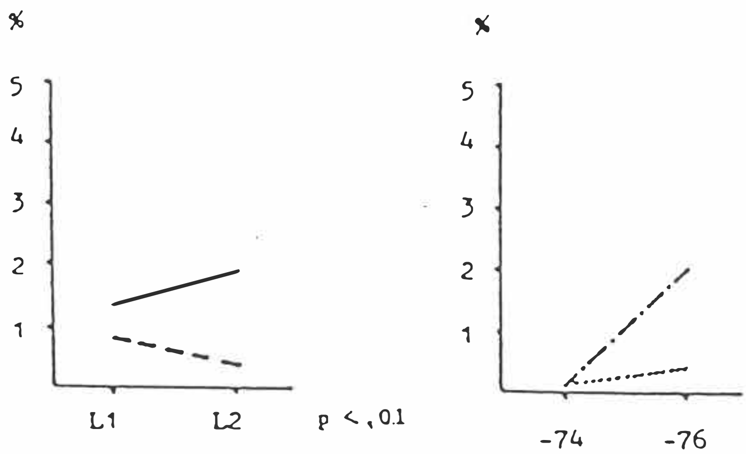


---- Program 1 (1974)
——— Program 2 (1976)
.... Microlesson 1
-·-·- Microlesson 2

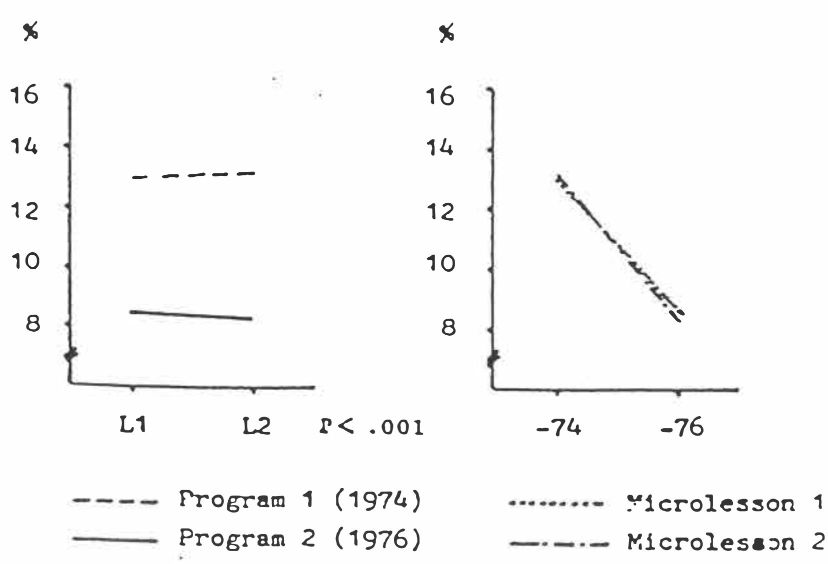
Cat. 3.2. Summarizes pupil's ideas or movement patterns, asks a pupil to demonstrate



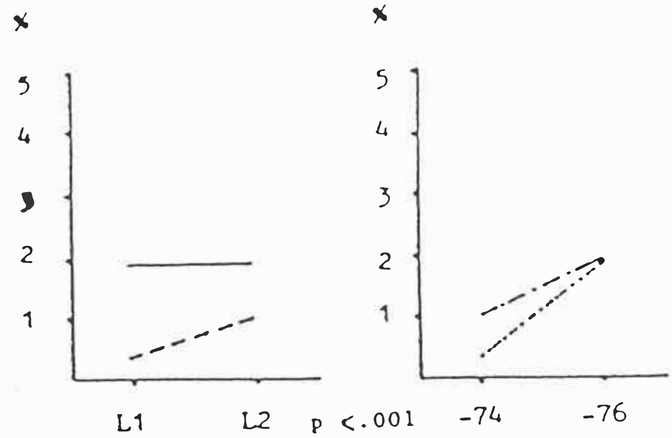
Cat. 3.3. Compares the ideas or movement patters expressed by one pupil to those of another or to those given, repeats pupil's ideas, asks a pupil to demonstrate



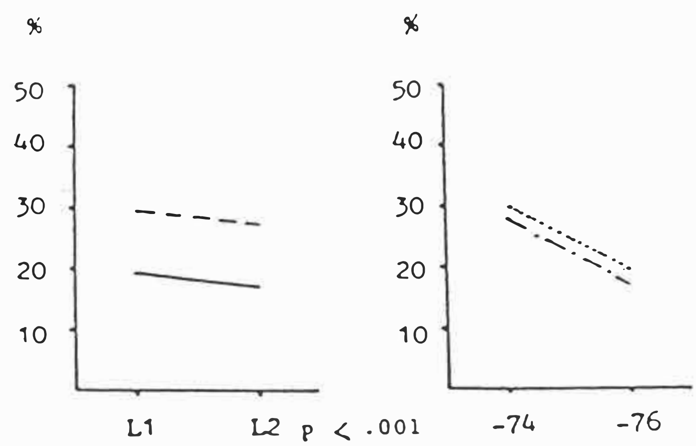
Cat. 4.1. Asks questions, initiates, terminates activity: Asks questions requiring narrow answers, initiates short-term activity, terminates activity



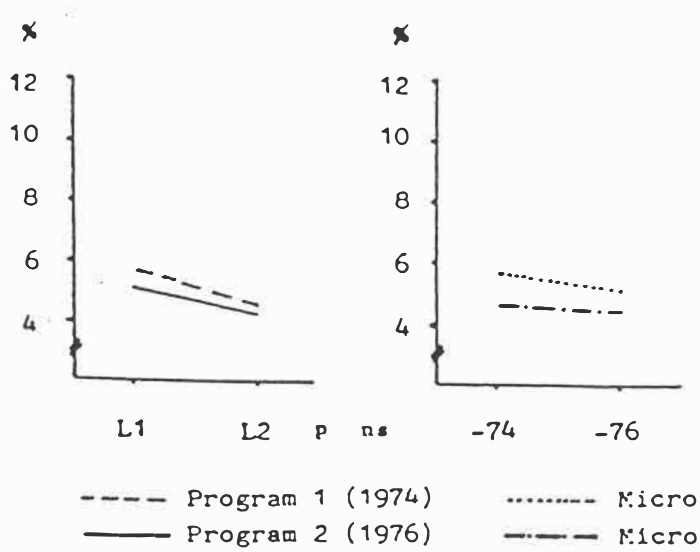
Cat. 4.2. Broad, open questions which clearly permit choice in ways of answering and moving



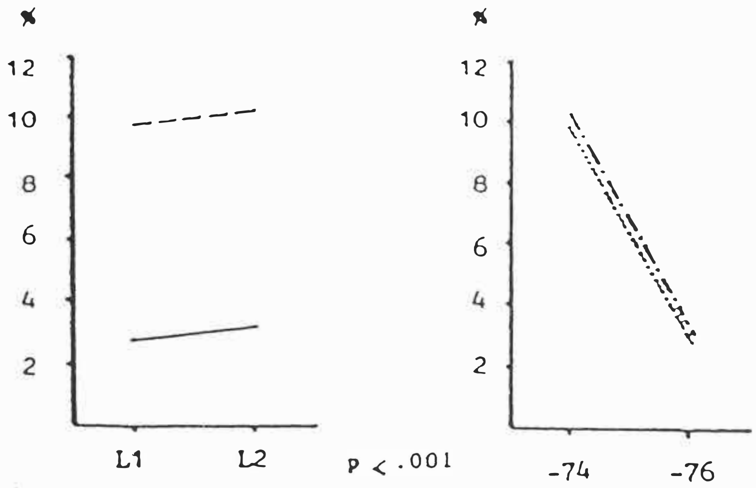
Cat. 5.1. Presents information, opinions, demonstrates movement patterns, makes a pupil demonstrate



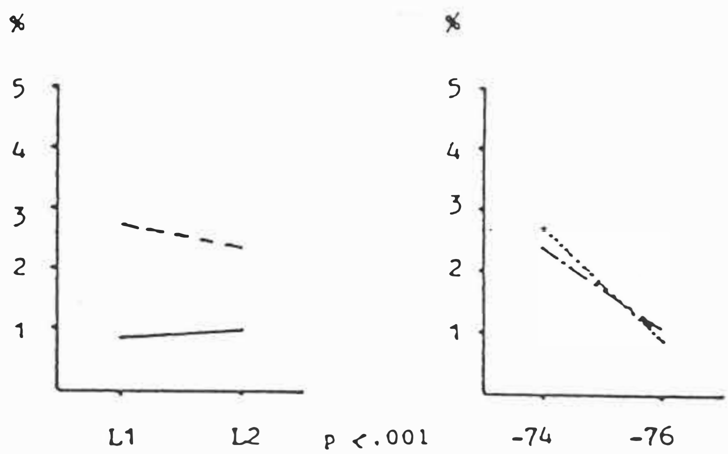
Cat. 5.2. Organizes pupils, material, division of labor and responsibility



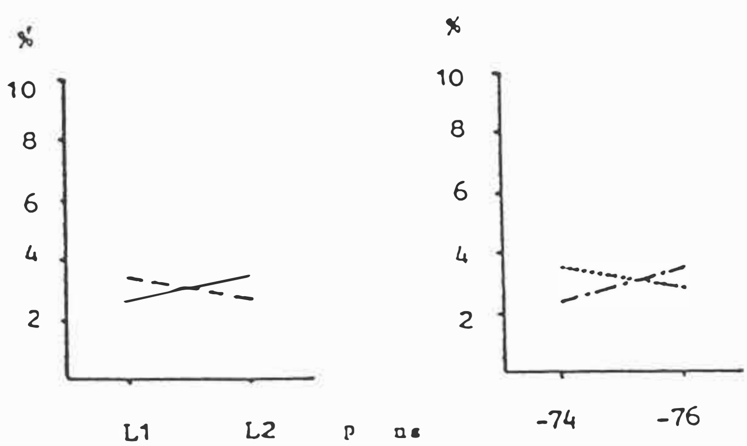
Cat. 6. Gives directions, commands during activity (pupils expected to comply)



Cat. 7. Criticizes pupil behavior, rejects movement pattern, justifies authority



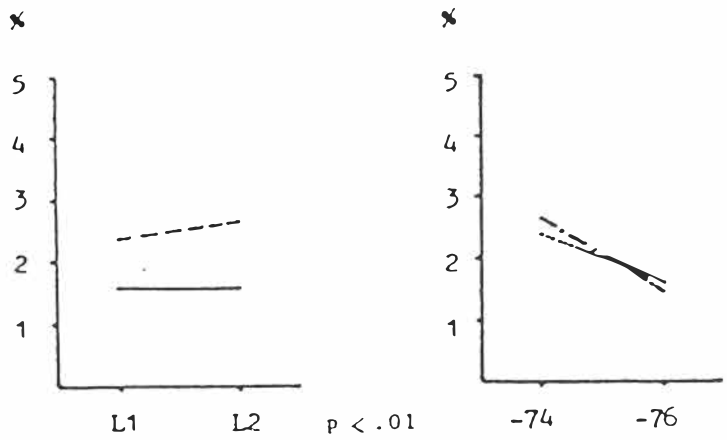
Cat. 8. Pupil answers question made by the teacher



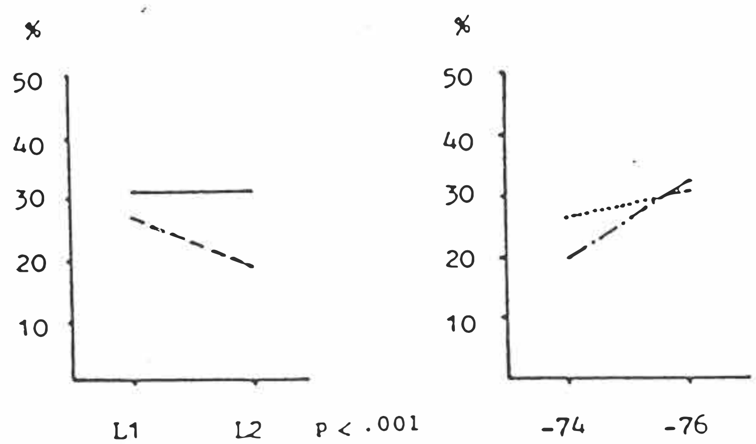
- - - - Program 1 (1974) ·········· Microlesson 1

——— Program 2 (1976) —·—·— Microlesson 2

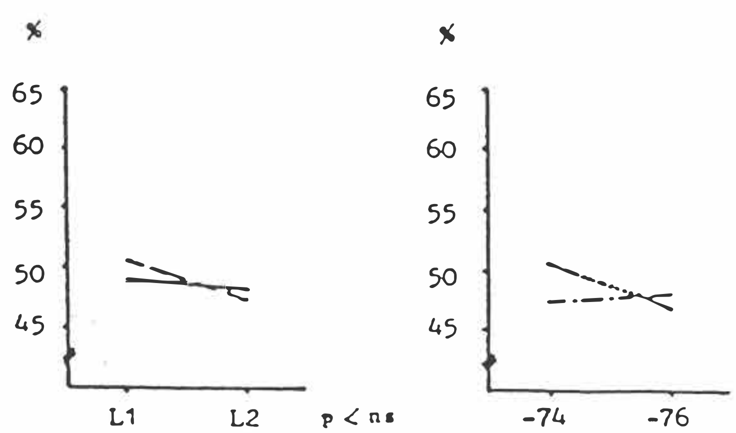
Cat. 9. Pupil initiates speech, asks for instructions, expresses own ideas or movement patterns



Cat. 10-12. (10) Teacher follows pupil's activity, silent guidance, (11) Teacher's silent participation in movement activity, (12) Confused situation, uproar



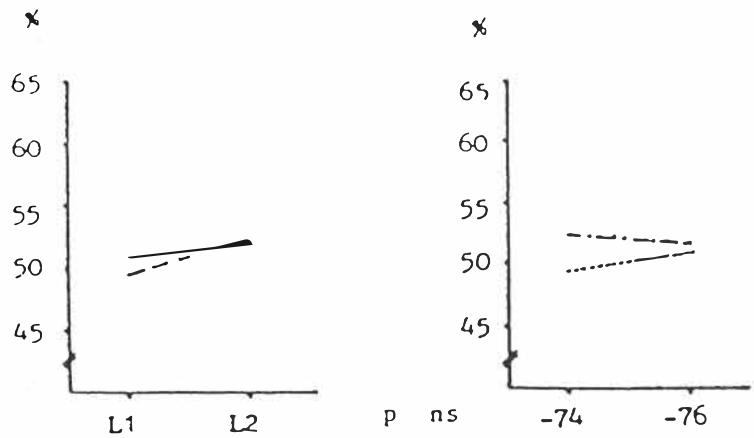
Cat. 1. Pupils collectively passive



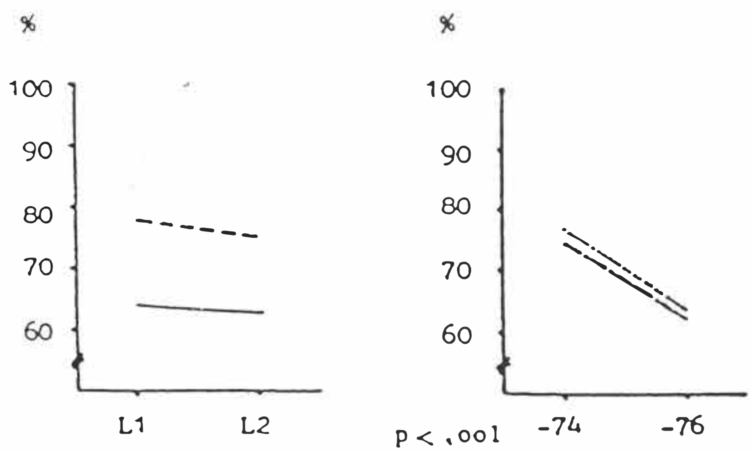
---- Program 1 (1974) .... Microlesson 1

—— Program 2 (1976) —·—·— Microlesson 2

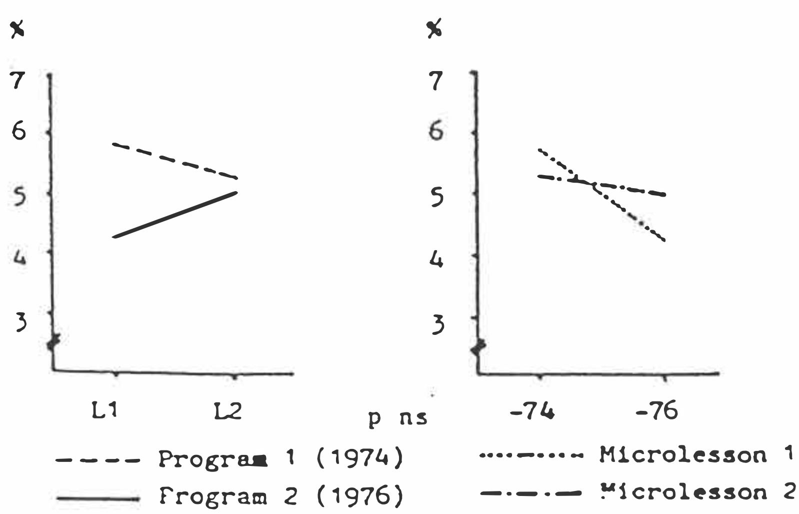
Cat. 2. Pupils collectively active



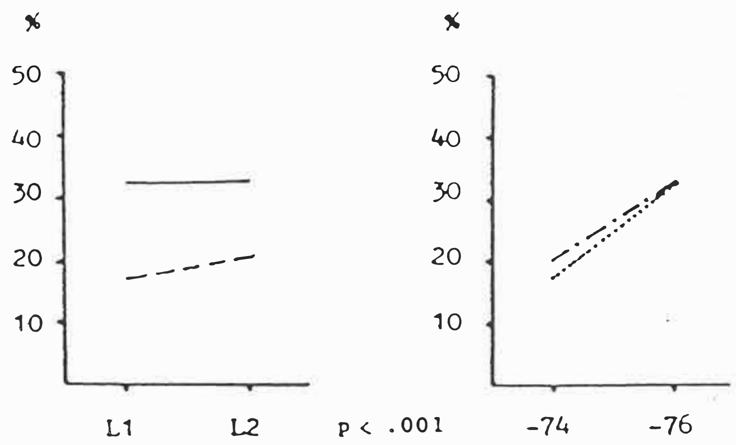
Index 1. Percent teacher talk (TT)



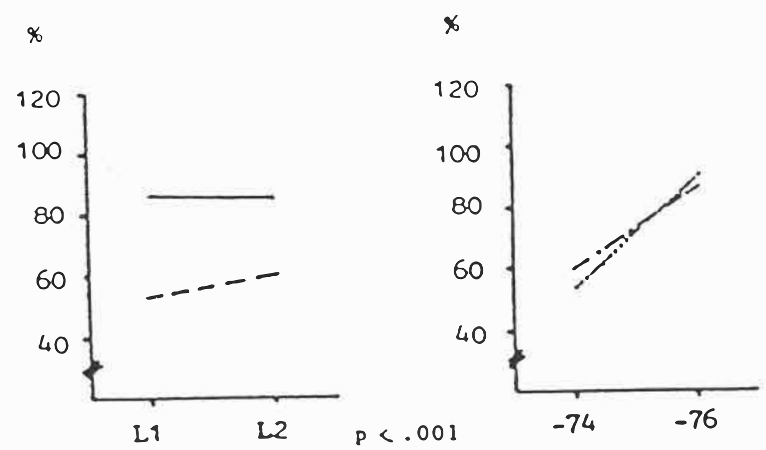
Index 2. Percent pupil talk (PT)



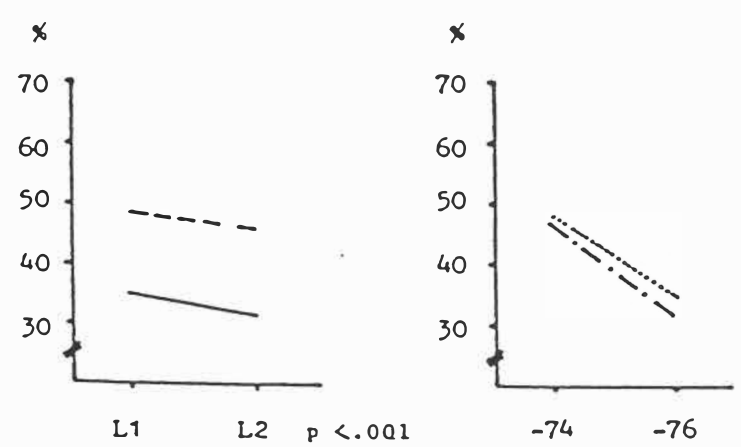
Index 4. Teacher's silent guidance and silent participation in movement activity ratio (TSGPR)



Index 5. Teacher response ratio (TRR)



Index 7. Content emphasis ratio (CCR)



---- Program 1 (1974) .... Microlesson 1

——— Program 2 (1976) —·—·— Microlesson 2

Liitetaulukko

OPISKELIJOIDEN ARVIOINNIT PIENOISOPETUKSEN KURSSISTA VUOSINA 1974 JA 1976

VÄITTÄMÄT	1974 N 48 Eri mieltä %	Ei osaa sanoa %	Samaa mieltä %	1976 N 73 Eri mieltä %	Ei osaa sanoa %	Samaa mieltä %	YHT. N 121 Eri mieltä %	Ei osaa sanoa %	Samaa mieltä %	χ^2 74-76 df 2	
31. Luennoitsija olisi saanut edetä ripeämmin	45.8	45.8	8.3	79.5	13.7	6.9	66.1	26.5	7.4	16.34	1
32. Luennoitsijalla ei ollut riittäviä tietoja aiheesta	66.7	27.1	6.3	82.2	13.7	4.1	76.0	19.0	5.0	3.92	2
33. Luennoitsijan omat käsitykset ovat värittäneet opetusta liikaa	25.0	54.2	20.8	45.2	23.3	31.5	37.2	35.5	27.3	12.16	
34. Harjoituksissa aika yleensä loppui kesken	16.7	10.4	72.9	35.6	4.1	60.3	28.1	6.6	65.3	6.15	2
35. Kurssi on sinänsä aika hyödyllinen	10.4	12.5	77.1	5.5	6.9	87.7	7.4	9.1	83.5	2.36	2
36. Kurssissa ei ole puututtu todella oleellisiin ja tärkeisiin asioihin	52.1	27.1	20.8	54.8	11.0	34.3	53.7	17.4	28.9	6.18	
37. Luennot ja demonstraatiot ovat niveltyneet hyvin yhteen	25.0	29.2	45.8	37.0	13.7	49.3	32.2	19.8	48.0	4.86	
38. Luennoilla on ollut helppo pysyä aiheesta kiinnostuneena	53.3	29.2	12.5	68.5	16.4	15.1	64.5	21.5	14.1	2.78	
39. Harjoituksissa on ollut helppo pysyä aiheesta kiinnostuneena	16.7	4.2	79.2	19.2	5.5	75.3	18.2	5.0	76.9	.26	2
40. Kurssin opetus ei ole herättänyt kiinnostusta aiheeseen	54.2	29.2	16.7	64.4	17.8	17.8	60.3	22.3	17.4	2.20	
41. Luennoitsijalla on turhia tapoja ja maneereja, jotka johtavat huomion pois opetuksesta	27.1	52.1	20.8	43.8	32.9	23.3	37.2	40.5	22.3	4.90	
42. Tämän kurssin pitäisi kuulua opinnoissa myöhempään vaiheeseen	77.1	20.8	2.1	75.3	16.4	8.2	76.0	18.2	5.8	2.20	2
43. Olen oppinut tämän kurssin luennoilla enemmän kuin luennoilla yleensä	68.8	31.3	0.0	57.5	31.5	11.0	62.0	31.4	6.6	5.85	2
44. Luennoitsija ei ottanut kuulijoitaan riittävästi huomioon	29.2	31.3	39.6	56.2	17.8	26.0	45.5	23.1	31.4	8.60	
45. Monistetut luentojäsennykset olivat hyödyttömiä	87.5	6.3	6.3	97.3	1.4	1.4	93.4	3.3	3.3	4.47	4
46. Luentosarja avasi minulle uusia näkökulmia liikunnanopetusta ajatellen	16.7	33.3	50.0	20.6	13.7	65.8	19.0	21.5	59.5	6.63	
47. Demonstraatioiden tehtävät olivat huonosti valittuja	70.8	22.9	6.3	61.6	17.8	20.6	65.3	19.8	14.9	4.74	
48. Luentosarja ei ollut kuuntelemisen arvoinen	43.8	35.4	20.8	67.1	20.6	12.3	57.9	26.5	15.7	6.49	
49. Liikunnanopetuksen koulutuksen kannalta olisi hyödyllisempää käyttää tämän kurssin aika muunlaisiin opetusharjoituksiin	58.3	16.7	25.0	71.2	9.6	19.2	66.1	12.4	21.5	2.36	
50. Pelkkä luentosarjan ja opetusmallien demonstrointi riittäisi ilman harjoituksia	89.6	10.4	0.0	91.8	1.4	6.9	90.9	5.0	4.1	8.08	4
51. Luennoilla edettiin liian nopeasti	29.2	43.8	27.1	43.8	26.0	30.1	38.0	33.1	28.9	4.48	
52. Harjoitusten organisaatio ei ollut riittävän hyvä	33.3	18.8	47.9	68.5	11.0	20.6	54.6	14.1	31.4	14.72	
53. Harjoitusten ohjaajan opetustaito ei ollut riittävän hyvä	47.9	37.5	14.6	83.6	11.0	5.5	69.4	21.5	9.1	17.43	1
54. Harjoituksissa selitettiin tehtävät selvästi	31.9	21.3	46.8	28.8	12.3	58.9	30.0	15.8	54.2	2.31	
55. Opin erottamaan opetusmallit tarkkaillessani ja luokitellessani palautetta	35.4	18.8	45.8	15.1	2.7	82.2	23.1	9.1	67.8	19.00	
56. Harjoitukset selvensivät luennoilla esitettyjä asioita	20.8	29.2	50.0	16.4	15.1	68.5	18.2	20.7	61.2	4.71	
57. Katson käsitykseni opetuskäyttäytymisestä laajentuneen	8.3	14.6	77.1	5.5	6.9	87.7	6.6	9.9	83.5	2.49	3
58. Tulen käyttämään todennäköisesti opetuksessani tietoisesti erilaisia esitettyjä opetusmalleja	8.5	12.8	78.7	4.1	13.7	82.2	5.8	13.3	80.8	1.01	2
59. Minulle selvisi kurssin aikana omassa opetuskäyttäytymisessäni ilmenevät puutteet ja virheet	19.2	12.8	68.1	12.3	9.6	78.1	15.0	10.8	74.2	1.54	

— = p < 0.05   ═ = p < 0.01   ≡ = p < 0.001

Appendix F

Liitetaulukko

OPISKELIJOIDEN ARVIOINNIT PIENOISOPETUKSEN KURSSISTA VUOSINA 1974 JA 1976

VÄITTÄMÄT	1974 N 48			1976 N 73			YHT. N=121			Ero 74-76 df=2 χ²	
	Eri mieltä %	Ei osaa sanoa %	Samaa mieltä %	Eri mieltä %	Ei osaa sanoa %	Samaa mieltä %	Eri mieltä %	Ei osaa sanoa %	Samaa mieltä %		
2. Kurssi esiteltiin siten, että alusta alkaen olin selvillä sen sisällöstä ja laajuudesta	52.1	10.4	37.5	60.3	8.2	31.5	57.2	9.1	33.9	.80	1
3. Pystyin alusta alkaen muodostamaan oikean kuvan luentosarjan tavoitteista	47.9	22.9	29.2	60.3	12.3	27.4	55.4	16.5	28.1	2.60	
4. Olin alusta lähtien selvillä harjoitusten tarkoituksesta	29.2	8.3	62.5	48.0	4.1	48.0	40.5	5.8	53.7	4.56	2
5. Kurssin pääkäsitteet on esitetty huonosti	56.3	33.3	10.4	43.8	17.8	38.4	48.8	24.0	27.3	12.12	
6. Kurssin opetus on herättänyt minussa kiinnostusta tähän aineeseen	31.3	18.8	50.0	17.8	15.1	67.1	23.1	16.5	60.3	3.91	
7. En ole oppinut luennoilla paljoakaan	37.5	23.0	39.6	48.0	19.2	32.9	43.8	20.7	35.5	1.28	
8. En ole oppinut harjoituksissa paljoakaan	68.8	6.3	25.0	82.2	2.7	15.1	76.9	4.1	19.0	3.05	2
9. Harjoitusten tehtävät ovat olleet järkeviä	18.8	23.0	58.3	41.1	8.2	50.7	32.2	14.1	53.7	9.25	
10. Opiskelijoiden käyttö oppilaina on ollut järkevää	70.8	10.4	18.8	39.7	4.1	56.2	52.1	6.6	41.3	16.9	2
11. Kurssin sisältö on mennyt tarpeettomasti päällekkäin sellaisen opetuksen kanssa, jota olen seurannut aikaisemmin	68.8	22.9	8.3	89.0	8.2	2.7	81.0	14.1	5.0	7.75	2
12. Kurssi on rakentunut aikaisemmin opetetuun järkevästi	27.1	35.4	37.5	21.9	19.2	58.9	24.0	25.6	50.4	5.94	
13. Kurssi on järjestetty hyvin verrattuna muihin vastaaviin kursseihin	43.8	33.3	23.0	32.9	38.4	28.8	37.2	36.4	26.4	1.50	
14. Luennon ja harjoitusten sisällöt eivät vastanneet riittävästi toisiaan	45.8	22.9	31.3	61.6	9.6	28.8	55.4	14.9	29.8	4.83	
15. Tämän kurssin olisi pitänyt kuulua opinnoissa varhaisempaan vaiheeseen	14.6	20.8	64.6	60.3	13.7	26.0	42.2	16.5	41.3	25.7	
16. Harjoituksiin kuului liian vähän erilaisia tehtäviä	27.1	27.1	45.8	53.4	6.9	39.8	43.0	14.9	42.2	12.9	
17. Harjoituksissa tahti on ollut liian tiivis	54.2	14.6	31.3	69.9	5.5	24.7	63.6	9.1	27.3	4.22	1
18. Liian vähän aikaa on käytetty palautteen analysointiin	29.2	22.9	47.9	50.7	4.1	45.2	42.2	11.6	46.3	12.08	
19. Luennoilla olisi saanut käyttää useammin audiovisuaalisia välineitä	20.8	49.7	31.3	46.6	16.4	37.0	36.4	28.9	34.7	15.47	
20. Luennoitsija on ollut liian persoonaton	50.0	41.7	8.3	78.1	20.6	1.4	67.0	29.0	4.1	11.27	2
21. Luennoitsija on puhunut riittävän selvästi	33.3	27.1	39.6	48.0	6.9	45.2	42.2	14.9	43.0	9.65	
22. Oli koko lukukauden ajan epätietoinen luentosarjan tavoitteista	68.8	16.7	14.6	68.5	5.5	26.0	68.6	9.9	21.5	5.42	1
23. Kurssin pääkäsitteet on opetettu riittävän selvästi	16.7	35.4	47.9	28.8	15.1	56.2	24.0	23.1	52.9	7.32	
24. Luennoilla ei annettu riittävästi aikaa kysymysten tekemiseen	37.5	35.4	27.1	49.3	30.1	20.6	44.6	32.2	23.1	1.69	
25. Opettajat ovat olleet kurssilla huolimattomia määräaikojen suhteen	66.7	14.6	18.8	75.3	9.6	15.1	71.9	11.6	16.5	1.17	
26. Yleensä olin luennoilla ikävystynyt	20.8	33.3	45.8	41.1	15.1	43.8	33.1	22.3	44.6	7.95	
27. Yleensä olin harjoituksissa ikävystynyt	79.2	4.2	16.7	74.0	5.5	20.6	76.0	5.0	19.0	.43	2
28. Monistetut luentojäsennykset olivat luennon tavoitteiden kannalta hyödyllisiä	4.2	4.2	91.7	6.9	1.4	91.8	5.8	2.5	91.7	1.27	4
29. Luentoa oli vaikea seurata	35.4	27.1	37.5	24.7	13.7	61.6	28.9	19.0	52.1	7.13	
30. Koko kurssi on liikunnanopettajain koulutuksessa hyödytön	83.3	6.3	10.4	79.5	11.0	9.6	81.0	9.1	9.9	.78	2

— = p < 0.05   ═ = p < 0.01   ≡ = p < 0.001