DOCUMENT RESUME

ED 241 166 PS 014 211

AUTHOR Hevern, Vincent W.; Geisinger, Kurt F.
TITLE An Initial Classification of Noncognitive Student Behavior Grading Items.
PUB DATE Aug 83
NOTE 43p.; Funded by a grant from the Graduate Studies Fund of the New York Province of the Society of Jesus.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Classification; Elementary School Teachers; Elementary Secondary Education; *Grading; Graduate Students; *Report Cards; Secondary School Teachers; *Student Behavior
IDENTIFIERS *Noncognitive Classroom Behaviors

ABSTRACT

Report-card marking items by which elementary school students' noncognitive classroom behaviors are graded were studied to establish a preliminary classification of behavioral assessment domains. Two studies were conducted. The first study employed a cluster-analytic approach to a pool of 136 noncognitive behavior-marking items that had been sorted and grouped into homogeneous sets by 16 psychologists and educators. In the second study, a total of 59 elementary school teachers rated a 32-item subset of the larger marking item pool on three scales. Specifically assessed were the degree to which items reflected (1) classroom or task-oriented adaptiveness, (2) interpersonal relationships and skills, and (3) personal adjustment and psychological self-concept. Cluster-analytic and analysis of variance results suggested a general dichotomy in the classification of items: quality of task orientation versus quality of interpersonal relational skills. Disagreement was found between the participants of the two studies in their classification of items dealing with students' self-regulatory behavior and acceptance or practice of normative patterns of socialized conduct. (Areas for further empirical study are discussed.) (Author/RH)

# An Initial Classification of Noncognitive Student Behavior Grading Items

Vincent W. Hevern*

Kurt F. Geisinger

Fordham University

Paper presented at the 91st Annual Convention of the American Psychological Association at Anaheim, CA, August 1983.





(* Vincent W. Hevern is now at Regis High School, 55 E. 84th St., New York, New York 10028. Please direct any comments or requests for additional copies to the first author.)

## Abstract

Elementary school report card marking items by which students' noncognitive classroom behaviors are graded were studied to establish a preliminary classification of their behavioral assessment domains. Two studies were undertaken: (1) 136 marking items were sorted by 16 psychologists and educators for homogeneity of content, and (2) 32 items were rated by 59 elementary school teachers for the degree to which the items' grading reflected task, interpersonal, or psychological adjustment behaviors of pupils. Cluster-analytic and ANOVA results suggested a general dichotomy in the classification of items: quality of task orientation versus quality of interpersonal relational skills. Areas for further empirical research are detailed.

The ongoing evaluation of the progress of children in elementary schools and the communication of such assessments comprise crucial functions within educational institutions. Among the differing means by which children are assessed and such evaluations are communicated, formal student progress reports ("report cards") rank with teacher-parent conferences as the most widely adopted means to fulfill these functions; upwards of 70% of American elementary-aged children receive report cards regularly (Kunder & Porwoll, 1977). The prevalence and importance of report cards have occasioned widespread research efforts, almost all of which have been dedicated to the further understanding of academic or subject-matter grading (Geisinger, 1982; Terwilliger, 1971). Nevertheless, most Americans remember that their report cards at the elementary school level contained a form of grades other than those given for academic subjects, namely, marks or ratings of attitudinal, affective, or noncognitive classroom behaviors and charateristics. The national survey of Kunder and Porwoll (1977) found that "separate ratings for behavior, work habits, or citizenship" were employed at 84% to 92% of schools which distribute report cards (p. 25). Thus, a curious and significant omission within the psychometric and educational literature involving grading practices has been the absence of any substantive research by which such noncognitive forms of report card assessment might be understood. The current investigators have failed to uncover a single empirical study in which such non-academic grades constituted the dependent variable(s) under review.

The importance of "noncognitive" student characteristics or behaviors has been recognized in terms of both their influence upon "[the] level and rate of subject-matter learning" (Messick, 1979, p. 281) and their predictive significance in the ongoing personal and psychological growth of young students (Achenbach, 1982, passim). Numerous research efforts have utilized teacher ratings of student behaviors in the form of behavioral rating checklists or scales by which to understand the factors or dimensions of normal and abnormal classroom behaviors (e.g., Achenbach & Edelbrock, 1978; Lambert, Hartsough, & Bower, 1979; Miller, 1972; Peterson, 1961; Schaefer, 1971). Furthermore, lists of desirable "outcomes" of schooling have been compiled or defined in terms of both academic and noncognitive aspects of student behavior (Krathwohl & Payne, 1971); notable among these efforts would be the educational Taxonomies developed by Bloom and his colleagues (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956; Krathwohl, Bloom & Massia, 1964; Simpson, 1966). Despite these efforts, a search for either clusters or dimensions of the noncognitive behaviors represented on report cards has not been made. Most basic psychometric characteristics of such grading practices remain unknown within the educational community, e.g., issues of reliability, validity, and grading distributions.

In addition to an absence of empirical data, a further question concerning noncognitive student behavior assessment emerged from the current authors' review of the literature involving teacher ratings of student behavior. It is evident from the diverse results of these studies that no uniform model exists for the dimensionality

underlying teacher perceptions of students. Investigators such as Armentrout (1971), Lessing, Williams, and Revelle (1981), and Miller (1972) have argued that a unifactor model of child behavior--Poor vs. Good Adjustment--underlies the ratings of classroom instructors. Achenbach and Edelbrock (1978) reviewed the rating scale literature involving significant adult figures in the social world of the child and advocated a two-factor behavioral model in the perception of such raters: an Externalizing vs. Internalizing dichotomy. Finally, a number of investigators who have focused explicitly upon the classroom as the context for assessment have proposed a three-factor model in the perception of student functioning (Kim, Anderson, & Bashaw, 1968 a,b; Lambert, Hartsough, & Bower, 1979; Schaefer, 1971). For example, Lambert and Nicoll (1977) provided "a conceptual model of nonintellectual behavior" by citing three general factors extracted from teacher rating scale data: Classroom Adaptiveness, Interpersonal Skills, and Intrapersonal (or Psychological) Factors. Since the completion of such behavioral rating scales or checklists functions as an analogue to the grading of students on report cards, the differences among the various studies caution against too facile an interpretation of what teachers might intend to communicate via noncognitive grading items.

In light of these studies as well as the lacunae within the literature, the objectives of the current report address two aspects of noncognitive grading practices, namely, (1) the gathering of a relatively comprehensive sample of actual noncognitive student behavior marking items from elementary school report cards and (2) an

initial categorization or grouping of these noncognitive marking items into relatively homogeneous (sub)sets of similar item stimuli. Preliminary in nature and intent, the current investigation was undertaken to establish an initial understanding of those report card items by which students are graded for other than subject-matter mastery.

The investigation reported below commenced with the gathering of a comprehensive pool of noncognitive student behavior grading items from currently-used, student report cards. When this objective was reached, two separate studies were undertaken to meet the second or classificatory goal. Study 1 employed acluster-analytic approach to a pool of 136 noncognitive behavior marking items which had been sorted and grouped into homogeneous sets by 16 psychologists and educators. In Study 2, 59 elementary school teachers rated a 32-item subset of the larger marking item pool on three scales: the degree to which these items were reflective of the domains suggested by Lambert and Nicoll (1977), i.e., (1) classroom or task-oriented adaptiveness, (2) interpersonal relationships and social skills, and (3) personal adjustment and psychological self-concept. These data were then used, first, to test a set of hypotheses concerning the content of the classifications resultant from the cluster analysis of Study 1 and, secondly, were also subjected themself to a cluster analysis as a partial, cross-validation of the solution found in Study 1. The results of the item ratings by teachers were generally supportive of the classification produced in Study 1 though some important differences were noted between the two cluster-analytic

solutions.

## ASSEMBLING A POOL OF NONCOGNITIVE BEHAVIOR MARKING ITEMS

A total of 50 public, parochial, and private schools, systems and districts were contacted throughout the Metropolitan New York area. They were asked to provide a copy of their "student progress report" currently used in grades four through six as well as information about the school's geographical setting. The choice of these grade levels was made on an a priori basis since this entire investigation was preparatory to further study with students at such grade levels (cf., Discussion, below). Positive responses were received from 34 schools as summarized in Table 1.

---

Insert Table 1 about here

---

From among the schools responding, 26 different forms were received with one additional school indicating that no report card was used by that institution.

Each report card was examined to isolate any marking item on the card which appeared to grade a pupil's "noncognitive" classroom behaviors, traits, and attitudes. Excluded in this process were items which the report card itself labeled as a "subject" marking item. Hence, cues from within the report card were also employed which indicated noncognitive behavioral assessment (e.g., headings such as "SOCIAL DEVELOPMENT" or "WORK HABITS"). A total of 248 items were isolated from the 26 report cards. Within this pool of 248

items, 163 differently-worded statements were found.

Each of these 163 marking items was compared with every other statement for minor variants in wording, spelling, or formulation. This comparison was made to reduce the number of qualitatively redundant items in the total item pool. Items found to be very highly similar were then combined into a single item. For example, the items, "Accepts criticism" and "Takes criticism," were considered conceptually identical and were joined to form the statement "Accepts/takes criticism." Likewise, items differing only in singular and plural case endings were joined, e.g., "Assumes responsibility/responsibilities." However, any item with a specific conditional modifier was allowed to remain in the pool though it otherwise closely approximated another item. Hence, "Completes assignments" and "Completes assignments on time" remained separate statements within the item pool. Thus, the 163 differently-worded statements were reduced to 136, the number used in Study 1. A full listing of this item pool is found below among the results of Study 1 within Table 2.

## STUDY 1

### Method

Subjects. A group of 20 subjects, all of whom had doctoral degrees in psychology or education or were doctoral candidates in these fields, was asked to participate in a study of the 136 statements in the noncognitive behavior item pool. All subjects had had previous teaching experience with grading responsibilities and

were familiar to varying degrees with fundamental theories of test construction and validation, educational or psychological testing, and developmental psychology. Five subjects (who were among the Study's volunteers) also had experience as administrators within various educational settings. These subjects were intended to and may simply be characterized as a heterogeneous group of educators and psychologists. Completed responses were received from 16 of the solicited participants; 5 held doctoral degrees and 11 were ABD graduate students with "education" the principal field of study for 8 subjects and "psychology" the major field for the other 8 volunteers. Volunteers received no remuneration for their participation.

Procedure. Each subject received a packet of 136 3 x 5 in (7.6 x 12.7 cm) cards on which each of the statements was typed, one per card, followed by the number of that statement's placement within an alphabetical listing of the items. Subjects received two copies of the full alphabetical listing of items. Finally, each subject was given a report sheet upon which item groupings and a "summary label" could be entered.

The subjects were instructed to group the items as best they could so that the resultant groups represented "relatively homogeneous item clusters descriptive of the same behaviors of children in a classroom." Subjects were asked explicitly to consider two grouping criteria: (1) similarity of behavioral content and (2) a rough similarity of the statements' level of inference or generality. Each grouping was required to contain a minimum of two differently-worded items and no statement was permitted to be placed

in more than one grouping. Finally, if the item defied grouping, the subject was told to place that item in a grouping entitled "MISCELLANEOUS - UNCLASSIFIED." A request was made "to keep this MISCELLANEOUS grouping as small as possible." No further directives were given to guide subjects in assembling item subsets. When the subjects had finished their grouping of items and reviewed the different piles of cards in order to make any last changes, they were asked (1) to decide upon a "summary label" descriptive of each grouping and (2) to enter this label and the item numbers from each grouping on a "Statement Grouping Report." Descriptive labels were requested primarily as prompts for subjects to consider the bases upon which the items were grouped as well as the need for homogeneity of content within specific item groupings. This report was returned to the investigators.

## Results

Preparation of Data. Item groupings received from the subjects were first examined to determine the frequency by which each item was deemed "unclassified or MISCELLANEOUS." Eight statements were so categorized by at least 20% of the subjects, i.e., by more than three respondents. These items were eliminated from further analysis and, thus, the final item pool consisted of 128 grading statements.[1]

A raw dissimilarity score--$S(ij)$, $i$, $j$ = 1 to 128; $i \neq j$--was computed for each pair of statements in the item pool, first by counting the number of subjects who grouped items $i$ and $j$ together, and, then, by subtracting that sum from the total number of subjects, i.e., from 16. Thus, raw dissimilarity scores for any $S(ij)$ ranged

potentially from 0 to 16. In order to convert each raw dissimilarity score into a distance metric, the algorithm suggested by Rosenberg and Kim (1975) for sorting data was employed. Thus, the 128 x 128 matrix of raw dissimilarity scores $[S(ij)]$, called "disagreement scores" by Rosenberg and Kim (1975), was used to calculate a final 128 x 128 matrix of distances by means of the formula:

$$\delta_{ij} = \sum_{K=1}^{K} \left(S_{iK} - S_{jK}\right)^2$$

where K is the number of items in the matrix. Schneider (1973) noted that $\delta_{ij}$ was one of four frequently employed measures in the calculation of profile distances among traits in a "Trait x Perceiver" research design (p. 295).

Cluster Analysis. The data were submitted for cluster analysis to the "Clustr" program of the Analysis of Quantitative Data Collection (AQD, 7th Edition; Schlaifer, 1978), a suite of statistical programs developed and maintained by the Harvard Business School. The subdiagonal matrix of nonredundant $\delta(ij)$ distances for each pair of items was utilized in the analysis.

The choice of an appropriate clustering algorithm for the present data was guided by a study by Blashfield (1976) which showed the "minimum variance method," proposed by Ward (1963), to provide the highest accuracy among four agglomerative hierarchical methods in the cluster analyses of randomly-generated data sets. The other three methods examined by Blashfield (1976) were single-, complete-, and average linkage; each of these demonstrated distinctly inferior

accuracy compared to Ward's method. Hence, the data of the current study were analyzed through the "minimum squared error" option of the AQD "Clustr" program.

The original dendrogram generated by this analysis contained 83 different inclusion or fusion levels for items within the clustering hierarchy. Because of both the number of items analyzed and steps in the agglomerative growth of the clusters, it was judged both impractical and unnecessary to reproduce the entire original dendrogram for this report. Rather, a more summational listing of items as they appeared in the results of the analysis are presented in Table 2.

A thorny problem in evaluating the results of any cluster analysis lies in the decision concerning the number of clusters present within the results (Blashfield, 1976; Everitt, 1974). One commonly suggested approach to the solution for this problem has been to "[examine] . . . the dendrogram for large changes between fusions" (Everitt, 1974, p. 59); another possible strategy calls for the cross-validation of the adequacy of the clustering with other data sets (Blashfield, 1976). While this second strategy constituted an important rationale for Study 2 reported below, the first approach mentioned above was applied to the cluster data reported in Table 2.

---

Insert Table 2 about here

---

The clustering results of the original dendrogram indicated that the largest increase in total squared error occurred at the last

inclusion step or fusion when a single cluster was formed out of the two large clusters which had been built up to that point in the analysis. This large change in total error seemed to suggest that two primary clusters existed among the 128 marking statements listed in Table 2: the first (labeled P1) ranged from item #11 to item #124 while the second (labeled P2) spanned item #64 to item #128. (Note that, within this entire report, all references to an item by number refers to that statement's position within the alphabetical sequence of the 128 marking statements in the original pool and not to placement or sequence within a particular table.) It was also observed that the second largest increase in total squared error occurred in the decrease from five to four groupings in the cluster fusions. For the purposes of this study it was considered appropriate to utilize these five groupings as items clustering at a "secondary" level in a hierarchy. Three of these secondary groupings formed the membership of the primary cluster P1 and two groupings comprised the primary cluster P2. Finally, on more intuitive and logico-linguistic bases, a tertiary division of items was made by the authors at an earlier point in the analysis as an aid to the interpretation of the primary and secondary clusters mentioned above. This tertiary set of item groupings was done on the basis of the relative homogeneity of content demonstated by items within these groupings. Interpretive labels were given to all three levels of item groupings and constitute the three divisions under which items are presented in Table 2.

Primary cluster P1 was interpreted as "Task Behaviors and

Intrapersonal Qualities." Items within P1 appear to deal with three aspects of student behavior within the school environment. The first aspect--labeled "Quality of Task-oriented Behavior"--assesses pupils along a hierarchy from very specific elements of classwork/clasrroom performance (e.g., neatness, accuracy, on-time work completion) to a broader set of concerns, particularly student efforts to "work to potential." A second dimension of this primary cluster is formed by items which evaluate how well children care for the physical environment of the school as well as their own possessions. The third aspect of this cluster has been termed "Quality of Intrapersonal Maturity and Self-Regulation." These items seem to assess a fairly broad range of behaviors and attitudes which characterize a student who profits from classroom instruction (e.g., attentive listening, openness to correction and criticism) and signify the level of personal autonomy or "maturity" of the student within the school environment (e.g., self-control, ability to work independently).

The label of "Quality of Social Skills and Interpersonal Relationships" has been given to items comprising primary cluster P2. Two general subdivisions appear to mark this cluster. The first, "Participation in School Activities and Discussions," is distinguished by items evaluative of the presence and extent of active involvement by students in different aspects of a school's life, e.g., classroom discussions and other types of activities by pupils in groups. The second subdivision, "Social Skills and Social Deportment," contains items assessing a range of interpersonal

skills both on a one-to-one basis (e.g., courtesy, respect, considerateness) and within groups (e.g., working with others in large groups).

## STUDY 2

The cluster solution described in Study 1 provided a categorization of the original pool of grading items with an apparent face validity. Nevertheless, Blashfield (1976) suggested that some form of cross-validating evidence from other data sets be sought to test the adequacy of the original cluster solution. Thus, to provide further empirical data by which to judge the adequacy of the categories resultant from the initial cluster analysis, Study 2 was undertaken, concurrently with Study 1, to gather a data set independent of that used in the first analysis. Teacher judgments about the topical content of a subset of noncognitive student behavior (NSB) marking items were sought along the three dimensions proposed by Lambert and Nicoll (1977) to characterize teacher ratings of pupil behavior, i.e., task-oriented or adaptive classroom behavior, interpersonal relationships, and psychological or intrapersonal factors.

A content analysis of the results summarized in Table 2 suggested three general hypotheses about the relative weight which teachers would assign to NSB items falling within various subgroupings. Three of these (sub)groupings were chosen to test these hypotheses: Quality of Task-oriented Behavior (I. A., henceforth called Group A); Quality of Intrapersonal Maturity and

Self-Regulation (I. C., labeled Group C), and QUALITY OF Social Skills and Interpersonal Relationships (primary cluster P2, labeled Group DE). It was hypothesized that: (1) significant differences would be found among the mean ratings of teachers for the degree of classroom adaptation or task orientation of NSB items so that Group A > Group C > Group DE; (2) Significant differences would be found among the mean ratings of teachers for the degree of interpersonal skill orientation of NSB items so that Group DE > Group C > Group A; and, (3) Group A items would be rated significantly lower for their personal adjustment content by teachers than either Group C or DE items (i.e., A < C = DE). Finally, as a second objective of Study 2, it was hypothesized that a cluster-analytic solution to the grouping of the subset of NSB items rated by the teacher-subjects would significantly correspond to the categorization of items found in Study 1.

Method

Subjects. A solicitation for subjects was made to 143 elementary and junior high school teachers from both convenience sources and mailings to school principals. A total of 59 teachers (49 female, 9 male, 1 unspecified) responded with completed data forms. The subjects indicated an average age of 43.5 years (SD = 12.2 years) with a mean teaching experience of 18.0 years (SD = 10.9 years). All but three of the subjects were White. The teachers were employed in public (N = 22; 38%), parochial (N = 35; 59%), and private (N = 1; 2%) schools with one teacher's employment unspecified. The locale described by the teachers for their work

ranged across urban (56%), urban-suburban (24%), and suburban (20%) settings. Teacher-raters completed the ratings voluntarily without receiving any gratuity.

Procedure and Instrument. Each teacher was given an introductory letter explaining the general purpose of the research and providing a promise of strict confidentiality concerning their responses. Attached to the letter was the "Teacher Item Rating Form" (TIRF) by which data were collected. The first page of the TIRF asked for the background and demographic information cited above.

The remainder of the TIRF consisted of three sets of ratings for a group of 32 NSB marking items. These items were chosen according to two criteria: (1) the 16 NSB statements most frequently appearing in the set of 26 report card forms originally gathered to assemble the pool of NSB items were identified and included on the TIRF and (2) 16 additional items were randomly chosen from among the remaining 112 NSB items. Note that selection of these items was made before the results of Study 1 were fully analyzed and one item, "Affective development," was included on the TIRF though subsequently dropped from the Study 1 item pool due to the high number of "MISCELLANEOUS" classifications it gathered. Hence, the results of Study 2 below are based ultimately upon 31 rather than the original 32 items.

Each teacher was asked to rate the 32 NSB items on three scales. Items were randomly sequenced on each scale. A 7-step, Likert-type rating format was adopted for the scales. The first rating scale requested teacher to judge "To what degree does an item reflect how well a child deals with classroom tasks?" This question was further

elaborated by noting that " 'Task-related' items would be those which evaluate how ADAPTIVE a child is to the classroom environment and the work he or she must do to profit from instructional material." Items on the "Task-related" scale (TASK in Table 3) were rated according to criteria by which a rating of "1" indicated that an item was "NOT AT ALL" reflective of the topic of the scale; a rating of "4" indicated that the item was "MODERATELY" reflective; and, a rating of "7" was made for those items judged "VERY HIGHLY" reflective. An identical scaling metric was adopted for the remaining two rating sets.

The second rating scale (INTER in Table 3) posed the question: "To what degree does an item reflect how well a child deal with his/her interpersonal relationships?" Teachers were told to consider how well items "evaluate[d] the quality of a student's social skills and interpersonal relationships . . . in the school environment, includ[ing] both peers and adults." Finally, teachers were asked to rate: "To what degree an item reflects a child's level of personal adjustment or self-concept?" on the third scale, labeled ADJUST in Table 3. Raters were told that "the notion of 'personal adjustment' refers to attitudes, behaviors, and traits indicating how a child feels about himself or herself . . . and may be inferred from different aspects of a child's personality and behavior, for example, degree of self-confidence or withdrawal, complaints and fears, mature or immature expressions of feeling, and so on." The TIRF concluded with a request to return the completed scales to the investigators by mail in an accompanying, stamped, self-addressed envelope.

The reliability of the scales forming the TIRF was examined by calculating intraclass correlations to determine inter-rater consistency on each scale. The results of this analysis indicated adequately high reliability levels: TASK = .93, INTER = .91, and ADJUST = .96. Among the 32 TIRF items, 11 were found to belong to Group A, 11 to Group C, and 8 to Group DE. A single item belonged to Study 1's subgrouping I. B.

Results

The mean ratings given by teachers to items belonging to each of the three cluster groups (A, C, and DE) are provided in Table 3. These mean ratings were used to test the hypotheses generated from the content analysis of the cluster solution in Study 1 and outlined above. Three one-way ANOVAs for unequal cell sizes were performed through use of the BMDP 7D program (Dixon, 1981) by entering the mean teacher ratings of Group A, C, and DE items for each of the three TIRF rating scales. All three ANOVAs indicated significant differences among the overall mean item ratings for the three groups: on scale TASK, $F(2, 27) = 11.65$, $p < .001$; on scale INTER, $F(2, 27) = 41.58$, $p < .001$; and, on scale ADJUST, $F(2, 27) = 12.95$, $p < .001$.

---

Insert Table 3 about here

---

Following these analyses of variance, the Newman-Keuls procedure, adapted for unequal cell sizes (Winer, 1971, pp. 215-218), was employed to test for the hypothesized significant differences in the ordering of the group means on each of the three scales. On

scale TASK, the hypothesized order of differences in mean ratings was found to be significant: Group A > Group DE ($M_{diff}$ = 1.97, $p < .01$); Group C > Group DE ($M_{diff}$ = 0.70, $p < .05$); and, Group A > Group C ($M_{diff}$ = 0.50, $p < .05$). On scale INTER, the predicated order of means was also found to be significantly different: Group DE > Group A ($M_{diff}$ = 2.55, $p < .01$); Group C > Group A ($M_{diff}$ = 1.72, $p < .01$); and, Group DE > Group C ($M_{diff}$ = 0.82, $p < .01$). Finally, on scale ADJUST the hypotheses that Group A would display a lower mean rating than either Groups C or DE (and that Group C would not differ significantly from Group DE) was also confirmed: Group A < Group DE ($M_{diff}$ = 1.01, $p < .01$); Group A < Group DE ($M_{diff}$ = 0.80, $p < .01$); and Group DE = Group C ($M_{diff}$ = 0.21, ns).

The second objective of Study 2 required a cluster analysis of the TIRF items based upon the teacher ratings. The mean ratings for each item on the three scales were used in this analysis. A Euclidean distance measure--D(ij)--was computed for each pair of items in the set according to the formula presented by Everitt (1974, p. 56):

$$D(ij) = \sum_{k=1}^{K} \left\{ \left[ X(ik) - X(jk) \right]^2 \right\}^{1/2}$$

where X(ik) is the value of the $i$th entry for the $k$th of the three variables (here, the three scales TASK, INTER, and ADJUST). Standaridized $z$-scores for each X(ik) were computed before

calculating D(ij) values in order to account for the variance between each of the three TIRF scales. A 31 x 31 matrix of distance scores was submitted to the AQD "Clustr" program for cluster analysis by means of the "minimum squared error" algorithm (Schlaifer, 1978).

The dendrogram produced by this analysis showed the largest increase in total squared error occurring at the fusion of two clusters into one. This indicated that two basic item groupings existed among the teacher-rated items. A listing of these results is given in Table 4. The first cluster, labeled Adaptive Task Behaviors (T1), included 15 items whose content centers about work-related behaviors by which students function adaptively in the classroom. The second cluster, called Interpersonal Relationships and Personal Maturity (T2), contained 16 items which appear to deal with various social skills and behaviors indicative of the personal maturity of pupils.

---

Insert Table 4 about here

---

A comparison was made between the analyses generated by Studies 1 and 2 and the categorizations of the 31 NISB statements common to both cluster solutions. It appeared that primary cluster P1 of Study 1 was most directly replicated by cluster T1 of Study 2. Parallel item placement across studies was less obvious for Study 1's cluster P2 and cluster T2 of Study 2. Hence, a 2 x 2 table was constructed by which to compare item placements across solutions and is reproduced as Table 5. The results of the Fisher Exact Probability

Test (Siegel, 1956, pp. 96-104) confirmed that item categorization was not independent of the two clustering approaches ($p < .01$).

Insert Table 5 about here

In order further to test the hypothesis that both cluster solutions would yield significanltly similar item groupings, a kappa statistic was computed as suggested by Fleiss (1973, pp. 145-147). The resultant kappa was .492 ($SE_K = .15$) which was highly significant ($z = 3.48$, $p < .001$). Though demonstrating statistical significance, the kappa value obtained only suggests a moderate degree of congruence between the two clustering solutions. A divergence in placement was evident for eight items of cluster T2 of the second study. These items, noted by an asterisk (*) in Table 4, were categorized under the heading "Interpersonal Relationships and Personal Maturity" in the solution of Study 2 while they were more highly associated with the "task-oriented" cluster (P1) of Study 1. It would appear that items dealing with distinctly social skill or task-oriented behaviors were more readily distinguised from each other than items which assess more "psychological (intrapersonal) adjustment" or "personal maturity" types of behaviors. Elementary school classroom teachers tended to see the assessment of maturity or self-regulatory behaviors as more similar to items evaluative of interpersonal functioning and school participation than did the subjects in Study 1.

## DISCUSSION

The introduction to the current investigation enunciated two objectives: the first sought to gather a relatively comprehensive subset of NSB grading items in actual use on report cards while the second was directed toward an preliminary categorization of those NSB statements. The pool of 163 NSB items assembled from 26 student progress report instruments at the fourth- to sixth-grade level was relatively large. A redundancy in NSB item formulation can be discerned in the listing of Table 2, even after minor variants in wording were eliminated. Such a redundancy, in turn, appears to support a judgment that the assembled item pool probably represents a fairly comprehensive subset of the universe of actual NSB items. As noted earlier, the current investigators are unaware of any published listing of NSB statements comparable to that found in Table 2. Hence, the first objective of this investigation seems to have been generally met.

In light of the NSB item pool assembled for this report, a serious question can be raised in regard to the potential application of these statements as rating stimuli. The very number of such discrete marking items as opposed, for example, to the possible range by which academic grading categories can be worded suggests a possible source of psychometric unreliability in the use of these items. Teachers are required to evaluate pupils on only a small subset of NSB statements (typically, from the assembled report cards, 3 to 10 items). Further, each subset differs across either school systems or specific institutions. Hence, one might question the

adequacy of the general preparation of teachers for the evaluative task of NSB grading within university or other forms of training. One only has to examine, as have the investigators, a few of the standard texts which focus on educational evaluation to discover that little or no attention is given to this regular assessment responsibility of elementary school teachers. More substantive justification for this caution is indicated by the results of the dual studies described above.

The data reported here demonstrated concurrence by the two differing subject groups on a dichotomy between two sets of NSB items somewhat strictly defined. The first set consisted of those task-oriented marking items by which a pupil's work output and task motivation are assessed. The underlying concept among items of clusters P1 and T1 seems to include the notion of "classroom industry" or "work engagement." The second set of items for which agreement was evident consisted of items from clusters P2 and T2 which related to pupils' interactions with others and, thus, the quality of interpersonal relationships expressed both within the working processes of the classroom and toward peers and adults at school. Nevertheless, a distinctive disagreement was encountered between the participants of the two studies in their classification of items dealing with students' self-regulatory behavior and acceptance or practice of normative patterns of socialized conduct, i.e., vis-a-vis authority, individual and property rights, regulations, discipline and self-control. These items chiefly appear within Study 1's subgrouping "Quality of Intrapersonal Maturity and

Self-Regulation"; further, they seem to require relatively high levels of inference on the part of a classroom instructor concerning the degree of personal maturity of students.

The discrepancy between the results of the two cluster-analytic solutions may possibly be understood by reference to the impact of differing implicit personality theories of the rater-subjects involved in this study. One might raise the question of the degree to which the data provided in this report reflect a factor such as "illusory correlation" (Chapman, 1967). Study 1's subjects may have been sorting or categorizing items largely, perhaps, on the basis of linguistic considerations and/or psychological theory. Schneider (1973) has indicated this category as one facet of the "realism" issue within implicit evaluative frameworks (pp. 301-302). Further, Schneider (1973) has pointed to the question raised by D'Andrade in 1970 concerning the influence of memory upon the behavioral rating of individuals. Linguistic similarity among items may influence memory ratings of behavior more highly than would be found if such ratings were done soon after a particular social interaction. Primarily, such topics pose the problem of whether "what ought to go together" rather than "what actually goes together" in behavioral evaluation items informs the structuring either of data such as those reported here or in actual NISB grading.

Further, a range of studies dealing with implicit personality theory, cited in the review of Schneider (1973), demonstrated crucial phenomena of relevance to the current data. For example, caution in adopting any taxonomic or classificatory schema derived soley from

sorting or rating data is suggested by findings of perceiver variance in the application of category content and relevance across differing situations (Schneider, 1973, pp. 306-307). Likewise, an important distinction can be made between (1) the categorization of items by "laboratory" versus "field" methods (sorting or rating versus actual grading data) and (2) the assessment of NSB with results communicated to a researcher as opposed to results sent to parents and pupils. Actual NSB grading practices may not correspond directly with the suggested dimensionalities uncovered in this study. For example, in unpublished data involving actual NSB marks gathered by the first author for an empirical follow-up to this report (Hevern, undated), there is evidence that only the very general dichotomy mentioned above influences the grading by teachers of fifth- and sixth-grade boys. Thus, the more subtle distinctions among items classified by the five subgroupings found in Study 1 (Table 2) may exist in theory, but, nonetheless, may be illusory in actual practice.

These data indicate that further empirical studies, particularly with actual NSB grades, are clearly needed: first, to understand the degree of psychometric reliability of these items when used in report card assessment; secondly, to define more precisely the dimensions underlying teacher perceptions of students when NSB items are employed as the rating stimuli; and, finally, to establish psychometric validity (concurrent, predictive, and construct) for both NSB items and any classificatory schema used to categorize them. This investigation attempted to begin the process by which a glaring deficiency within the psychometric and educational assessment literature might be addressed and remedied.

## References

Achenbach, T. A. Developmental psychopathology (2nd ed.). New York: John Wiley & Sons, 1982.

Achenbach, T. A., & Edelbrock, C. S. The classification of child psychopathology: A review and analysis of empirical efforts. Psychological Bulletin, 1978, 85, 1275-1301.

Armentrout, J. A. Parental child-rearing attitudes and preadolescents' problem behaviors. Journal of Consulting and Clinical Psychology, 1971, 37, 278-285.

Blashfield, R. K. Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical models. Psychological Bulletin, 1976, 83, 377-388.

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. Taxonomy of educational objectives. Handbook I: Cognitive domain. New York: Longmans, Green, 1956.

Chapman, L. J. Illusory correlation in observational report. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 151-155.

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

Everitt, B. Cluster analysis. New York: John Wiley & Sons, 1974.

Fleiss, J. L. Statistical methods for rates and proportions. New York: Wiley-Interscience, 1973.

Geisinger, K. F. Marking systems. In H. E. Mitzel (ed.), Encyclopedia of educational research (5th ed.). Vol. 3. New York: Macmillan and The Free Press, 1982.

Hevern, V. W. A preliminary construct validation of teacher grading of nonintellective student behavior among fifth- and sixth-grade males (Doctoral dissertation, Fordham University, in preparation).

Kim, Y., Anderson, H. E., & Bashaw, W. L. Social maturity, achievement, and basic ability. Educational and Psychological Measurement, 1968, 28, 535-543. (a)

Kim, Y., Anderson, H. E., & Bashaw, W. L. The simple structure of social maturity at the second grade level. Educational and Psychological Measurement, 1968, 28, 145-153. (b)

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. Taxonomy of educational objectives: The classification of educational goals. Handbook II: Affective domain. New York: David McKay Co., 1964.

Krathwohl, D. R., & Payne, D. A. Defining and assessing educational objectives. In R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D. C.: American Council on Education, 1971.

Kunder, L. H., & Porwoll, P. J. Reporting pupil progress: Policies, procedures, and systems (ERS Report). Arlington, VA: Educational Research Services, Inc., 1977.

Lambert, N. M., Harsough, C. S., & Bower, E. M. A process for the assessment of effective student functioning: Technical report. 1979. Available from Publishers Test Service, CTB/McGraw-Hill, Monterey, CA.

Lambert, N. M., & Nicoll, R. C. Conceptual model for nonintellectual behavior and its relationship to early reading achievement. Journal of Educational Psychology, 1977, 69, 481-490.

Lessing, E. E., Williams, V., & Revelle, W. Parallel forms of the IJR Behavior Checklist for Parents, Teachers, and Clinicians. Journal of Consulting and Clinical Psychology, 1981, 10, 347-362.

Messick, S. Potential uses of noncognitive measurement in education. Journal of Educational Psychology, 1979, 71, 281-292.

Miller, L. C. School Behavior Check List: An inventory of deviant behavior for elementary school children. Journal of Consulting and Clinical Psychology, 1972, 38, 134-144.

Peterson, D. R. Behavior problems of middle childhood. Journal of Consulting Psychology, 1961, 25, 205-209.

Rosenberg, S., & Kim, M. P. The method of sorting as a data-gathering procedure in multivariate research. Multivariate Behavioral Research, 1975, 10, 489-502.

Schaefer, E. S. Development of hierarchical, configurational models for parent behavior and child behavior. In J. P. Hill (Ed.), Minnesota symposium on child psychology. Vol. 5. Minneapolis: The University of Minnesota Press, 1971.

Schlaiefer, R. User's guide to the AQD Collection (7th ed.; No. 9-173-095). Cambridge, MA: Harvard Business School, 1978.

Schneider, D. J. Implicit personality theory: A review. Psychological Bulletin, 1973, 79, 294-309.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Simpson, E. J. The classification of educational objectives: Psychomotor domain. Illinois Teacher of Home Economics, 1966, 10,

110-144.

Terwilliger, J. S. Assigning grades to students. Glenview, Illinois: Scott, Foresman & Co., 1971.

Ward, J. H. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 1963, 35, 160-165.

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

## Footnote

The authors wish to express their appreciation to the Graduate Studies Fund of the New York Province of the Society of Jesus for the grant to the first author which supported part of this research. Appreciation is also extended to the subjects who willingly participated in the studies reported in this paper.

1 The items eliminated from the final pool were: "Aesthetic development, Affective development, Attention, Behavior needs improvement, Comes for help, Cultural growth, Demonstrates/shows good sportsmanship, and Safety-Obeys safety rules."

TABLE 1

RESPONSES OF SCHOOLS TO REQUEST

FOR STUDENT PROGRESS REPORT FORMS

| | Public | Parochial | Private | Total |
|---|---|---|---|---|
| SOLICITED | 25 | 17 | 8 | 50 |
| RETURNED | | | | |
| Urban | 6 | 4 | 3 | 13 |
| Urban-Suburban | 2 | 1 | 2 | 5 |
| Suburban | 8 | 5 | | 13 |
| Suburban-Rural | 2 | 1 | | 3 |
| TOTAL (% of solicited) | 18 (72) | 11 (65) | 5 (63) | 34 (68) |

TABLE 2

STUDENT BEHAVIOR GRADING ITEMS:

CLUSTER-ANALYTIC GROUPINGS

| Summary Label | # | Item |
|---|---|---|
| I. TASK BEHAVIORS AND INTRAPERSONAL QUALITIES (P1) | | |
| A. Quality of Task-oriented Behavior | | |
| 1. Timely Work Completion | | |
| | 11 | Begins work promptly |
| | 25 | Completes homework assignments on time |
| | 20 | Completes assignments on time |
| | 22 | Completes classroom assignments on time |
| | 27 | Completes study and written work on time |
| | 30 | Completes work on time |
| | 31 | Completes work promptly |
| 2. Preparation for and completion of classwork | | |
| | 21 | Completes classroom assignments |
| | 23 | Completes classwork |
| | 24 | Completes homework (assignments) |
| | 19 | Completes assignments |
| | 26 | Completes make-up work |
| | 44 | Does not complete assignments |
| | 60 | Homework preparation |
| | 17 | Comes prepared |
| | 18 | Comes prepared for class |
| | 14 | Class preparation |
| | 85 | Preparation of assignments |
| | 116 | Unprepared |
| 3. Attention to various details of tasks | | |
| | 52 | Follows directions |
| | 71 | Listens (to) and follows directions |
| | 51 | Expresses himself clearly in all written work |
| | 108 | Speaks clearly and correctly |

Note.- Table 2 continued on the following page.

TABLE 2 (cont'd.)

STUDENT BEHAVIOR GRADING ITEMS:

CLUSTER-ANALYTIC GROUPINGS

| | |
|---|---|
| 28 | Completes work accurately |
| 7 | Accuracy in basic processes |
| 42 | Does accurate work |
| 45 | Does work carefully |
| 29 | Completes work neatly |
| 43 | Does neat work/does work neatly |
| 112 | Takes pride in neat and accurate work |
| 74 | Makes good use of time |
| 118 | Uses time wisely |
| 119 | Work habits need improvement |

4. Effort and seriousness of purpose

| | |
|---|---|
| 46 | Effort |
| 47 | Effort and application |
| 48 | Effort and attitude |
| 109 | Strives for improvement |
| 87 | Puts forth best effort |
| 122 | Works to ability |
| 123 | Works to potential |

B. Respect for and Care of Property and Materials

| | |
|---|---|
| 95 | Respects the rights and property of others |
| 92 | Respects own and others' property |
| 93 | Respects property of others |
| 12 | Cares for possessions |
| 70 | Is responsible for personal property |
| 13 | Cares for/Takes care of property |
| 99 | Shows care for school property and materials |
| 110 | Takes care of books and materials |
| 117 | Uses equipment carefully |

C. Quality of Intrapersonal Maturity and Self-Regulation

1. Attentive listening

| | |
|---|---|
| 63 | Is a good listener |

Note.- Table 2 continued on the following page.

TABLE 2 (cont'd.)

STUDENT BEHAVIOR GRADING ITEMS:

CLUSTER-ANALYTIC GROUPINGS

| | |
|---|---|
| 68 | Is developing listening skills |
| 57 | Has good listening skills |
| 72 | Listens attentively |
| 73 | Listens carefully |
| **2.** | **Obedience to and Respect for Authority and Rules** |
| 2 | Accepts and respects authority |
| 69 | Is obedient and respects authority |
| 32 | Complies with school regulations |
| 75 | Obeys rules and regulations |
| 76 | Obeys/observes school rules |
| 90 | Respects class and school rules |
| **3.** | **Openness to correction and criticism** |
| 3 | Accepts constructive criticism |
| 5 | Accepts suggestions and criticism |
| 6 | Accepts/takes criticism |
| 86 | Profits from corrective suggestions |
| **4.** | **Coping with responsibility** |
| 4 | Accepts responsibility/responsibilities |
| 1 | (Assumes and) carries out responsibilities |
| 9 | Assumes responsibility/responsibilities |
| 67 | Is dependable |
| **5.** | **General deportment, attitudes, and personal hygene** |
| 15 | Classroom discipline |
| 33 | Conduct |
| 10 | Attitudes |
| 58 | Has positive attitudes |
| 98 | Shows a positive attitude toward school |
| 65 | Is careful of personal appearance |
| 59 | Health habits |
| 84 | Practices good health habits |
| 115 | Tries to obey health rules |

Note.- Table 2 continued on the folllowing page.

TABLE 2 (cont'd.)

STUDENT BEHAVIOR GRADING ITEMS:

CLUSTER-ANALYTIC GROUPINGS

6. Self-discipline and control

| | |
|---|---|
| 8 | Acts with self-direction and control |
| 49 | Exercises/practices self-discipline |
| 50 | Exhibits/practices/shows self-control |
| 56 | Growing in self-discipline |
| 96 | Self-control |

7. Personal initiative and independence in work

| | |
|---|---|
| 53 | Gains satisfaction from work |
| 111 | Takes pride in accomplishments |
| 113 | Takes pride in personal habits |
| 97 | Self-motivation |
| 104 | Shows self-confidence |
| 41 | Demonstrates/displays/shows initiative |
| 114 | Thinks for himself |
| 120 | Works well independently |
| 124 | Works well alone |

II. QUALITY OF SOCIAL SKILLS AND INTERPERSONAL RELATIONSHIPS (P2)

D. Participation in School Activities and Discussions

| | |
|---|---|
| 64 | Is active and helpful in school activities |
| 62 | Involves himself in classroom activities |
| 78 | Participates in class (activities) |
| 79 | Participation in classroom and school activities |
| 16 | Classroom Participation |
| 35 | Contributes to (class) discussions |
| 77 | Participates in activities and projects |
| 80 | Participates in discussion |
| 81 | Participates in discussions and activities |
| 36 | Contributes to group discussions and activities |
| 82 | Participation |

Note.- Table 2 continued on the following page.

TABLE 2 (cont'd.)

STUDENT BEHAVIOR GRADING ITEMS:

CLUSTER-ANALYTIC GROUPINGS

E. Social Skills and Social Deportment

1. Courtesy, respect, and considerateness

34 Considerate
100 Shows consideration
66 Is courteous
83 Practices courtesy
101 Shows courtesy to others
88 Recognizes and respects the rights of others
91 Respects others
94 Respects rights of others
103 Shows respect

2. Cooperativeness and ease in peer and adult interaction

38 Cooperates willingly
39 Cooperates with others
30 Cooperates/cooperation
54 Gets along well with others
55 Gets along well with peers/other children
89 Relates well with peers and adults
61 Influence on others
102 Shows leadership
105 Social behavior
106 Social development
107 Social growth

3. Ability to work well within groups

37 Cooperates in group situations
121 Works (well) with others
125 Works well in groups/a group
126 Works well with a group
127 Works with others in small groups
128 Works with others in large groups

TABLE 3

TEACHER RATINGS OF NONCOGNITIVE STUDENT BEHAVIOR MARKING ITEMS: AVERAGE MEANS AND STANDARD DEVIATIONS

| [a] # | [b] Item | Teacher Ratings [c] TASK | INTER | ADJUST |
|---|---|---|---|---|
| I. A. | QUALITY OF WORK AND TASK-ORIENTED BEHAVIORS | | | |
| 17 | Comes prepared | 5.441 (1.74) | 3.305 (1.64) | 5.224 (1.49) |
| 21 | Completes classroom assignments | 5.542 (1.61) | 3.136 (1.62) | 4.121 (1.80) |
| 24 | Completes homework (assignments) | 5.203 (1.64) | 3.172 (1.78) | 4.328 (1.61) |
| 28 | Completes work accurately | 5.271 (1.68) | 2.983 (1.54) | 4.333 (1.67) |
| 29 | Completes work neatly | 4.339 (1.52) | 2.814 (1.60) | 4.448 (1.58) |
| 30 | Completes work on time | 4.966 (1.63) | 2.915 (1.55) | 4.123 (1.67) |
| 46 | Effort | 5.552 (1.76) | 3.793 (1.79) | 5.193 (1.70) |
| 51 | Expresses himself clearly in all written work | 4.864 (1.72) | 2.559 (1.36) | 3.690 (1.65) |
| 52 | Follows directions | 5.707 (1.77) | 4.220 (1.59) | 4.456 (1.55) |
| 74 | Makes good use of time | 5.356 (1.56) | 3.695 (1.65) | 4.379 (1.69) |
| 122 | Works to ability | 5.362 (1.74) | 3.390 (1.77) | 5.466 (1.44) |
| | OVERALL GROUP IA ITEM RATINGS | 5.237 (0.39) | 3.710 (0.48) | 4.524 (0.54) |

Note.- Table 3 continued on the following page.

TABLE 3

TEACHER RATINGS OF NONCOGNITIVE STUDENT BEHAVIOR MARKING ITEMS: AVERAGE MEANS AND STANDARD DEVIATIONS

| [a] # | [b] Item | [c] Teacher Ratings TASK | INTER | ADJUST |
|---|---|---|---|---|
| I. C. | QUALITY OF INTRAPERSONAL MATURITY AND SELF-REGULATION | | | |
| 2 | Accepts and respects authority | 4.220 (1.51) | 6.069 (0.93) | 5.379 (1.40) |
| 3 | Accepts constructive criticism | 4.746 (1.42) | 5.492 (1.54) | 5.448 (1.42) |
| 4 | Accepts responsibility/ responsibilities | 4.915 (1.43) | 5.293 (1.31) | 5.655 (1.22) |
| 15 | Classroom discipline | 5.172 (1.74) | 5.448 (1.31) | 5.140 (1.64) |
| 50 | Exhibits/practices/shows self-control | 4.712 (1.39) | 5.610 (1.35) | 5.569 (1.39) |
| 53 | Gains satisfaction from work | 5.000 (1.71) | 3.780 (1.77) | 5.759 (1.48) |
| 58 | Has positive attitudes | 5.085 (1.68) | 5.552 (1.43) | 6.086 (1.44) |
| 72 | Listens attentively | 5.627 (1.52) | 4.559 (1.56) | 4.649 (1.64) |
| 76 | Obeys/observes school rules | 4.203 (1.52) | 5.271 (1.51) | 4.793 (1.39) |
| 84 | Practices good health habits | 3.085 (1.77) | 4.153 (1.74) | 4.931 (1.67) |
| 120 | Works (well) independently | 5.339 (1.60) | 3.690 (1.73) | 5.105 (1.60) |
| | OVERALL GROUP 1C ITEM RATINGS | 4.737 (0.70) | 4.992 (0.81) | 5.319 (0.44) |

Note.- Table 3 continued on the following page.

TABLE 3

TEACHER RATINGS OF NONCOGNITIVE STUDENT BEHAVIOR MARKING ITEMS:

AVERAGE MEANS AND STANDARD DEVIATIONS

| [a] # | [b] Item | [c] Teacher Ratings TASK | INTER | ADJUST |
|---|---|---|---|---|
| II. D. E. | PARTICIPATION IN SCHOOL ACTIVITIES & SOCIAL SKILLS AND SOCIAL DEPORTMENT | | | |
| 35 | Contributes to (class) discussions | 4.458 (1.44) | 4.914 (1.50) | 5.483 (1.23) |
| 54 | Gets along well with others | 3.983 (1.35) | 6.441 (0.97) | 6.103 (1.07) |
| 61 | Influence on others | 3.789 (1.26) | 5.741 (1.28) | 5.228 (1.27) |
| 66 | Is courteous | 3.483 (1.45) | 5.814 (1.12) | 5.000 (1.24) |
| 82 | Participation | 4.684 (1.45) | 5.276 (1.30) | 5.474 (1.44) |
| 103 | Shows respect | 3.741 (1.51) | 6.000 (1.23) | 5.259 (1.31) |
| 106 | Social development | 3.724 (1.20) | 6.017 (1.15) | 5.719 (1.29) |
| 121 | Works (well) with others | 4.458 (1.51) | 6.322 (1.20) | 6.000 (1.20) |
| | OVERALL GROUP WIDE ITEM RATINGS | 4.040 (0.44) | 5.816 (0.51) | 5.533 (0.39) |
| 95 | Respects the rights and property | 3.759 (1.64) | 6.254 (1.17) | 5.707 (1.28) |

The number of teachers rating each item fell in the range: 57-59.

[a] The number printed to the left of each item corresponds to that item's alphabetical placement within the original pool of 128 items.

Note.- Table 3 continued on the following page.

TABLE 3

TEACHER RATINGS OF NONCOGNITIVE STUDENT BEHAVIOR MARKING ITEMS:

AVERAGE MEANS AND STANDARD DEVIATIONS

b This list of items is divided into three groupings based upon an item's membership within the secondary subclusters defined by Table 2 of this study. At the end of each subcluster of items presented here, an average mean and standard deviation for that subgroup's items is reported.

c Teachers evaluated each item on a scale with values from 1 ("NOT AT ALL" reflective) to 7 ("VERY HIGHLY" reflective). The heading TASK refers to the teacher rating scale on which they were asked to assess the degree to which an item reflects how well a child deals with classroom tasks or is adaptive to the classroom environment. INTER refers to ratings of the degree to which an item reflects how well a child deals with his/her interpersonal relationships at school. The column headed ADJUST indicates those ratings of the degree to which an item is reflective of a child's level of personal (psychological) adjustment or self-concept.

TABLE 4

CLUSTER ANALYSIS OF NONCOGNITIVE STUDENT MARKING ITEMS

BASED UPON ELEMENTARY SCHOOL TEACHER RATINGS

| # | LABEL Item[a] |
|---|---|
| | **ADAPTIVE TASK BEHAVIORS (T1)** |
| 122 | Works to ability |
| 17 | Comes prepared |
| 120 | Works (well) indepedently |
| 46 | Effort |
| 53 | Gains satisfaction from work |
| 72 | Listens attentively |
| 52 | Follows directions |
| 74 | Makes good use of time |
| 30 | Completes work on time |
| 28 | Completes work accurately |
| 24 | Completes homework (assignments) |
| 21 | Completes classroom assignments |
| 51 | Expresses himself clearly in all written work |
| 29 | Completes work neatly |
| 4 | Practices good health habits |
| | **INTERPERSONAL RELATIONSHIPS AND PERSONAL MATURITY (T2)** |
| 121 | Works (well) with others |
| 54 | Gets along well with others |
| 106 | Social development |
| 95 | Respects the rights and property of others * |
| 2 | Accepts and respects authority * |
| 103 | Shows respect |
| 61 | Influence on others |
| 66 | Is courteous |
| 82 | Participation |
| 4 | Accepts responsibility/responsibilities * |
| 50 | Exhibits/practices/shows self-control * |
| 3 | Accepts constructive criticism * |
| 15 | Classroom discipline * |
| 58 | Has positive attitudes * |
| 76 | Obeys/observes school rules * |
| 35 | Contributes to (class) discussions |

Teacher-Raters: $N$ = 59.

[a] An asterisk following items in cluster T2 indicate those items which belonged to cluster P1 of Study 1 (cf., Table 2)

TABLE 5

COMPARISON OF ITEM GROUPINGS

BETWEEN CLUSTER SOLUTIONS OF STUDY 1 AND STUDY 2

| | Study 2 | |
|---|---|---|
| | T1 | T2 |
| Study 1 | | |
| P1 | ADAPTIVE TASK BEHAVIORS | INTERPERSONAL/ MATURITY |
| TASK/ INTRAPERSONAL | 15 | 8 |
| P2 | | |
| INTERPERSONAL/ SOCIAL SKILLS | 0 | 8 |

Fisher Exact Probability Test: $p < .01$

Kappa = .492, $p < .001$