AN ABSTRACT OF THE THESIS OF

Christine Marie King for the degree of Master of Science in Education presented on August 27, 1984.

Title: Classroom Teachers' Accuracy in Observing Students' Motor Performance

Abstract approved: John M. Dunn

The purpose of this study was to examine how accurate classroom teachers' observations are for identifying high and low motor performances.

In order to assess teacher accuracy in rating motor performance, an analysis was conducted on students' scores on the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency between those who were rated high as compared to those who were rated low in motor performance as determined by a teacher rating scale.

Ninety-one children were randomly selected from Fairplay Elementary School in the Corvallis, Oregon school district to serve as subjects in the study. The subjects ranged in age from 6 to 11 years. There were 44 female and 47 male subjects. Four classroom teachers participated in the study.

Subjects were initially observed by their classroom teachers and rated on six motor performance tasks on a rating scale. On the basis
of the teachers' ratings, the subjects were placed into upper and lower quartile groups. Those subjects who scored in the upper and lower quartiles on the rating scale were administered the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency (BOT-S).

Subjects were observed by their classroom teachers on a Monday and formally tested by 10 different volunteer testers on the following Thursday. Subjects were tested one class at a time, taking approximately one half-hour for 19-30 children. The test site was set up with 10 stations and subjects rotated around to each station until all items were completed. The order in which the children were assigned a station was randomly decided.

The two-tailed t statistic was utilized to determine if there was a significant difference between standard mean scores for upper and lower quartile performances on the BOT-S. The results of the data analysis indicated a significant difference between upper and lower quartile performances. The mean score for upper quartile performance (63.76) was significantly higher than the mean score for the lower quartile performance (54.08) at the .05 level. The obtained t value of 3.35 exceeded the tabled value of 2.01 at the .05 level of confidence.

While these results were significant at the .05 level, a degree of caution must be used when interpreting the data. The degree of variance within the high group was 66.77 whereas the degree of variance within the low group was 139.03. The results suggest that
classroom teachers are more accurate in observing high motor performances than they are in observing low motor performances.
CLASSROOM TEACHERS' ACCURACY IN OBSERVING STUDENTS' MOTOR PERFORMANCE

by

Christine Marie King

A THESIS

submitted to

Oregon State University

in partial fulfillment of the requirements for the degree of

Master of Science

Completed August 27, 1984

Commencement June 1985
APPROVED:

Redacted for Privacy

Professor of Physical Education in charge of major

Redacted for Privacy

Chairman of Department of Physical Education

Redacted for Privacy

Dean of School of Education

Redacted for Privacy

Dean of Graduate School

Date thesis is presented August 27, 1984

Typed by Mary Ann (Sadie) Airth for Christine Marie King
ACKNOWLEDGEMENTS

The author wishes to express sincere appreciation to Dr. John M. Dunn for his guidance throughout the course of this study. For their constructive comments and encouragement I extend gratitude to committee members Dr. David Andrews, Dr. Tom Grigsby and Dr. Richard Irvin.

I would also like to thank Dr. Kathleen Heath for her contributions in the final preparation of the text.

I extend a special thanks to Dr. A. Marie Boarman for her continuous support and humor.
DEDICATION

To my Mom and Dad, with Love
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Purpose of Study</td>
<td>3</td>
</tr>
<tr>
<td>Significance of the Study</td>
<td>3</td>
</tr>
<tr>
<td>Methodology</td>
<td>4</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>5</td>
</tr>
<tr>
<td>Delimitations</td>
<td>6</td>
</tr>
<tr>
<td>Limitations</td>
<td>6</td>
</tr>
<tr>
<td>Definitions</td>
<td>7</td>
</tr>
<tr>
<td>II REVIEW OF LITERATURE</td>
<td>9</td>
</tr>
<tr>
<td>Developments in the Study of Performance Rating</td>
<td>9</td>
</tr>
<tr>
<td>Factors which influence classroom teachers' decisions to refer students for special education assessment</td>
<td>16</td>
</tr>
<tr>
<td>Motor Proficiency Testing</td>
<td>23</td>
</tr>
<tr>
<td>Summary</td>
<td>27</td>
</tr>
<tr>
<td>III METHODS AND PROCEDURES</td>
<td>29</td>
</tr>
<tr>
<td>Subjects</td>
<td>29</td>
</tr>
<tr>
<td>Test Instrument</td>
<td>30</td>
</tr>
<tr>
<td>Procedures</td>
<td>34</td>
</tr>
<tr>
<td>Analysis of Data</td>
<td>36</td>
</tr>
<tr>
<td>IV RESULTS AND DISCUSSION</td>
<td>38</td>
</tr>
<tr>
<td>Results</td>
<td>39</td>
</tr>
<tr>
<td>Description of Subjects' Performance</td>
<td>39</td>
</tr>
<tr>
<td>Hypothesis Testing</td>
<td>40</td>
</tr>
<tr>
<td>Discussion</td>
<td>41</td>
</tr>
<tr>
<td>Performance Evaluations</td>
<td>45</td>
</tr>
<tr>
<td>Summary</td>
<td>46</td>
</tr>
<tr>
<td>V SUMMARY AND RECOMMENDATIONS</td>
<td>48</td>
</tr>
<tr>
<td>Summary</td>
<td>48</td>
</tr>
<tr>
<td>Recommendations</td>
<td>51</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>52</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>A. Use of Human Subject Approval</td>
<td>57</td>
</tr>
<tr>
<td>B. Teacher Rating Scale</td>
<td>58</td>
</tr>
<tr>
<td>C. Directions for the Short Form of the</td>
<td>59</td>
</tr>
<tr>
<td>Bruininks - Oseretsky Test of Motor</td>
<td></td>
</tr>
<tr>
<td>Proficiency</td>
<td></td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Description</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Comparison of Mean Scores Between Upper and Lower Quartile Performances</td>
<td>40</td>
</tr>
<tr>
<td>II</td>
<td>Comparison Between Observed and Expected Values Obtained from the BOT-S Between Groups Rated High and Low</td>
<td>44</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

The Education for All Handicapped Children Act of 1975, PL 94-142, mandates that all handicapped children receive a free appropriate public education in the least restrictive environment. Included in the law's definition of special education is physical education, the only subject area specifically mentioned. The law mandates assessment in physical and motor ability as one of nine areas of assessment of children suspected of having special needs. Public Law 94-142 also states that tests must be standardized and valid and be selected to evaluate specific areas of educational need (HEW, 1977, p. 42496).

Considerable research has shown that classroom teachers, as well as physical educators, feel both unprepared academically as well as technically with regard to the rules and regulations of PL 94-142 (Klesius, 1981, Carpenter and Robson, 1979, Ringlaben and Price, 1981, Marston and Leslie, 1983, Broadhead, 1981).

In order to assure compliance with the law, teachers must be aware of the referral process and have an understanding of the motor skill level at which students in their classrooms should be functioning.
To evaluate whether or not a child has a handicapping condition as defined in the law, it has been suggested that the students' performance should be compared with that of non-handicapped peers (Broadhead and Bruininks, 1982). On the basis of an individuals' performance, a program placement can be determined. One of the methods used to screen individuals to determine the presence of a handicapping condition is to test children using a standardized instrument (Pyfer, 1982).

Selecting a test in the motor domain requires that the items included cover a wide range of performance traits. Research supports the use of the Bruininks - Oseretsky Test of Motor Proficiency on the basis that it provides differentiated measures of gross and fine motor skills (Broadhead and Bruininks, 1982, Broadhead and Church, 1982, Krus, Bruininks and Robertson, 1981, Beitel and Mead, 1980). The test has also been validated for: 1) making decisions about educational placement, 2) assessing gross and fine motor skills, 3) developing and evaluating motor training programs, 4) screening for special purposes and 5) assisting clinicians and researchers (Bruininks, 1978).

A short form of the Bruininks - Oseretsky Test was constructed from data obtained in the standardization of the long form of the test. The short form allows for the test to be administered in a shorter period of time and to large numbers of children.

It is the intent of this study to analyze student scores on the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency,
between those who were rated high as compared to those who were rated low in motor performance as determined by a teacher rating scale. In order to comply with PL 94-142, teachers must be perceptive in their observations of children's motor skills to assure that those in need of special services will receive them. As comprehensive training in physical education has not been a part of the classroom teachers' preparation, it is important to assess whether teachers are perceptive in their evaluations of students' motor performances. This study will provide information concerning the accuracy of classroom teachers' observations of students' motor performances.

PURPOSE OF THE STUDY

The purpose of this investigation, designed as a pilot study, was to examine how accurate classroom teachers' observations are for identifying high and low motor performances. In order to assess classroom teachers' accuracy in rating motor proficiency, an analysis was conducted on student scores on the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency between those who were rated high as compared to those who were rated low in motor performance as determined by a teacher rating scale.

SIGNIFICANCE OF THE STUDY

The Education for All Handicapped Children Act of 1975, PL 94-142, specifies that handicapped students receive a free appropriate education in the environment in which they can best benefit. Because
physical education is defined in the law as a direct service and must be provided for handicapped students, it is essential that those students who qualify as handicapped be identified and placed in the appropriate physical education class.

In compliance with the law, classroom teachers are expected to identify students with motor deficiencies. The results of this study will determine whether or not this is a realistic expectation. Further, the results will show potential implications for providing classroom teachers with in-service training to observe and evaluate children's motor skills.

METHODOLOGY

Ninety-one children from grades one, two and four at Fairplay Elementary School in Corvallis, Oregon, served as subjects for the study. The subjects were initially rated by their classroom teachers on a motor task observation form that was developed by the investigator. Those who scored in the upper and lower quartiles were later tested on the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency (BOT-S). Items on the observation form were designed to parallel the content found in the BOT-S and included running, ball skills, rope jumping and jumping jacks, balance, standing broad jump, and handwriting skills. Items on the BOT-S included running speed, balance, strength, coordination, upper-limb speed and dexterity, and visual-motor control.
The first test was administered by the classroom teachers to the students in their respective classes. The second test was administered to those subjects who scored in the upper and lower quartiles of the first test by ten different volunteer testers. The testers included five physical education graduates, one nurse, one school aide and two parents. Prior to testing, each volunteer met individually with the investigator to rehearse the specific directions and demonstrations to be used during the testing process. All subjects performed each test item in the presence of a group of peers and volunteer testers.

Test items on the teacher rating scale were rated on a scale of 1 to 5 points (1 = low, 5 = high). Point scores obtained from the Short Form of the Bruininks - Oseretsky Test were converted to standard scores. A t test was performed on the test data to determine if there was a significant difference between the means of standard scores obtained from the Short Form of the Bruininks - Oseretsky Test of students who were rated in the upper quartile compared to those who were rated in the lower quartile.

HYPOTHESIS

The following hypothesis was tested: There is no significant difference in motor proficiency between children who were rated high on motor performance compared to children who were rated low on motor performance by their classroom teachers.
DELIMITATIONS

The study was limited to investigating scores obtained from the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency using subjects who were rated high and low in motor performance as determined by a teacher rating scale. The test items selected for the teacher rating scale measure the following factors of motor proficiency; balance, agility, eye-hand coordination, upper-limb dexterity and strength. The standardized test used in this study was the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency. The subjects used in the study were elementary school aged children ages 6 - 11 who were enrolled in the Corvallis Public School System during the 1983-1984 school year. Classroom teachers who participated in this study were also from the Corvallis Public School System.

Students were asked to wear sneakers or rubber soled shoes but were not required to wear them. Students participated in their regular school clothing.

LIMITATIONS

The following conditions were limiting factors in this study:

1. Students were not assumed to have all had the same experience or exposure to physical education. It was assumed that the students would be familiar with certain test items and have the motor background to participate in all test items.
2. The Short Form of the Bruininks - Oseretsky Test of Motor Proficiency was administered in a group setting. The test results may reflect some effect of peer presence.

3. No guidelines were given on the teacher rating scale as to the kind of performance that would earn a 1 to 5 rating. The purpose was to allow as much teacher interpretation for performance as would be if no rating form were available.

4. No attempt was made to train the classroom teacher to observe students' motor performances. This was done to allow teachers to base their evaluations on their own knowledge of motor skill performance.

DEFINITIONS

For the purpose of this study, key terms were defined as follows:

1. **ASSESSMENT** - A procedure to measure and evaluate basic fundamental skills and motor development.

2. **IEP** - An individualized education program is a written statement developed by an educational team to identify the goals and objectives of education for a handicapped child in conformity with the Education for All Handicapped Children Act of 1975.
3. **LEAST RESTRICTIVE ENVIRONMENT** - Placement of a child where he or she can obtain the best education within the closest proximity to the regular educational setting.

4. **MAINSTREAMING** - The integration of handicapped students into regular educational programs.

5. **MOTOR PERFORMANCE** - The manner in which motor skills are carried out.

6. **MOTOR PROFICIENCY** - Performing movements with expert correctness.


8. **REFERRAL** - A process whereby children are directed to a specialist for formal testing to determine if a handicapping condition exists.

9. **SCREENING** - A process whereby children are informally observed for clues to functional problems in motor development.
CHAPTER II

REVIEW OF LITERATURE

The following review of literature will investigate performance ratings and methods used to improve accuracy in observation. The studies of some of the major contributors to research in performance rating will be reviewed. Factors which include classroom teachers' decisions to refer students for special education assessment will be examined. Finally, choosing a comprehensive motor proficiency test will be discussed.

DEVELOPMENTS IN THE STUDY OF PERFORMANCE RATINGS

The most widely used method of obtaining performance measures on individuals is rating scales. Unfortunately, rating scales have been shown to be prone to various types of systematic and random error. In an ideal sense, complete performance measurement would include the combination of objective, personnel, and judgmental indices (Landy and Farr, 1980). However, because of the difficulty in obtaining objective indices, most individuals concerned with performance measurement depend on judgmental indices alone.

Despite the widespread use of performance criterion scales, there has been a constant dissatisfaction with these measures due to both intentional and inadvertent bias. As a result, there has been a
tremendous amount of research done in an attempt to improve the validity of performance appraisals.

In 1979, Borman examined the effects of rater training and rating format on halo and accuracy in performance rating. A training program found to reduce a number of rating errors and five rating formats was developed for two jobs. Five-nine minute videotapes of actors performing these two jobs according to scripts depicting realistic levels of effectiveness provided the ratees for subjects to evaluate. One hundred twenty three college students were assigned randomly to one of ten cells defined by five format and two training treatments (trained and untrained), and asked to evaluate the effectiveness of the videotaped ratees performances. Results showed that training significantly reduced the halo effect, but this finding was uneven across the two jobs. Also, training did not improve accuracy. Borman concluded that there were two factors affecting accuracy in observation: 1) agreement on performance standards, and 2) that the significant dimension effect indicates possibly that certain kinds of dimensions are inherently more difficult than others for evaluating ratees accurately.

In a similar study, Borman (1978) sought to examine upper bound limits in reliability and validity of performance ratings by creating a nearly ideal environment for obtaining performance evaluations and then examining closely the precision of these ratings. Sixteen scripts describing persons performing two jobs - recruiting interviewer and manager - were prepared in such a way that the
performers' effectiveness on the various dimensions of performance approached a preset, realistic level. Five - seven minute performances were videotaped and fourteen raters evaluated the effectiveness of each performer on each performance dimension. Raters were knowledgeable about the two jobs, had ample time to observe the performers and used an advanced behavior scale type of rating form. Results showed that levels of convergent and discriminant validities were high, but that interrater agreement varied. In response to rater disagreement, Borman concluded that a close examination of the performance judgement process may provide some answers. Borman suggests that performance evaluation judgements be viewed in three steps: 1) observing work-related behavior, 2) evaluating each of these behaviors, and 3) weighting these evaluations to arrive at a single rating on a performance dimension.

In a third study, Borman (1977) sought to determine the consistency of rating accuracy and rating errors of human performance. The study was designed to assess the intra-individual consistency of performance rating accuracy in regard to the perception of others' job performance. Results indicated that certain individual differences are consistently associated with accuracy. Within-rating task consistency in accuracy (rater tends to exhibit about the same level of that error on different parts of the task) was found to be higher than across-task consistency (rater consistency of accuracy or of a rating error across two or more different settings), indicating that, to some extent, individual differences or
"abilities" associated with rating performance accurately may be situation specific. The significance of this study was the focus on the performance-rating process and the effectiveness of rating individuals. Implied in the results is that the ability to perceive and rate accurately will vary according to each individual and each new setting.

Using procedures similar to Bormans' (1978), Murphy, Garcia, Kerkar, Martin and Balzar (1982), conducted a study to determine the relationship between observational accuracy and performance rating accuracy. Forty-four undergraduates viewed videotapes of four graduate student lecturers and rated each using scales designed to measure accuracy in behavioral observation, as well as standard teacher evaluation forms. Results of the study suggest that accuracy in observation was related to accuracy in performance evaluations. Raters who accurately discriminate between ratees in terms of the frequency of favorable teacher behaviors also were accurate in discriminating between ratees in their performance evaluations. They also were accurate in discriminating between ratees within each of their performance dimensions. In answer to the question of whether frequency ratings and performance evaluations are two distinct aspects of the rating process, Murphy et al. concluded that frequency ratings are tied to specific behaviors whereas the Performance Evaluation Scale requires complex, abstract judgements about the quality of performance. The data from this study support the relationship between accuracy in observing behavior and accuracy in
evaluating performance but leave three important questions unanswered: 1) the effect of the lag between observation and rating (i.e. annual appraisals), 2) the relationship between complex behaviors and an increased number of ratees, and 3) whether performance evaluations are actually increased as a result of observational training.

In 1970, Gordon conducted a study to determine the effects of two variables on the accuracy of ratings: 1) the correctness of the behavior rated, and 2) the amount of the raters experience with the particular rating device. One hundred eighteen managers rated nine specific dimensions of the performance of an "insurance agents" prospecting technique. The results indicated that the ratings were 88.0% accurate on behavior which was performed correctly, but only 73.8% accurate on behavior which was performed incorrectly. This finding was called the Differential Accuracy Phenomenon (DAP). Although experience with the ratings improved the overall accuracy, it had no effect on the DAP. Whereas this study was designed specifically to measure the DAP, any source of inaccuracy in ratings will affect the measurement of the relationships under study.

In an attempt to improve observation processes, Thornton and Zorich (1980) conducted a study to evaluate two training procedures hypothesized to improve observation accuracy. One hundred seventy male and female college students were trained with one of three sets of instructions. They then viewed a 45 minute videotape of a six-person leaderless group discussion. The first set of instructions
simply explained the task (control); the second encouraged careful observation and recording of specific behavior; the third included the instructions given in the second set plus training on the avoidance of eight systematic errors of observation. Observer accuracy was measured by an objective test of behavioral events in the videotape. Results showed that subjects trained under the third set of instructions were more accurate in observation than subjects trained under the second, both of which were more accurate than the control group. Clearly implied in the results is that increased accuracy in behavioral observation may lead to more accurate performance effectiveness ratings.

Latham, Wexley and Pursell (1975) conducted a similar study on the effects of training to reduce errors. Sixty managers in a large corporation were randomly assigned to a workshop, a group discussion, or a control group. The workshop and group discussion involved training directed toward the elimination of rating errors that occur in performance appraisal and selection interviews, namely, contrast effects, halo effects, similarity and first impressions. Six months after the training the managers rated hypothetical candidates who were observed on videotape. The results showed that: a) trainees in the control group committed similarity, contrast and halo errors, b) trainees in the group discussion committed impression errors, and c) trainees in the workshop committed none of the errors. The significance of the study demonstrated that training observers can minimize rating errors.
Bernardin and Walter (1977) also were interested in the effects of rater training and error in ratings. In their investigation the psychometric properties of ratings or Behavioral Expectations Scales (BES) were compared across four groups of student raters (n = 156). The groups differed with respect to amount of prior training (1 hour or more), the nature of psychometric errors, and the extent of exposure to scales (read scales and recorded observed critical incidents, discussed general scale dimensions, or no exposure to scales). Students (n = 12, 3 from each group) rated 1 of 13 instructors during the last week of a 10 week term. Results showed significantly less leniency error and halo effect, plus higher interrater reliability, for the group that had received the hour of training and full exposure to the BES. The group that received only training had significantly less halo error than the group that had received no training. Results from this study support the conclusion of Latham and others (1975) that training prior to observation will reduce error in ratings.

Based on this review of literature the following suggestions appear as viable alternatives to increase accuracy in observation and reduce errors in ratings: 1) determine specific performance objectives, 2) define kinds of behaviors vs. quality of performance, and 3) provide training for those making observational judgements.
FACTORS WHICH INFLUENCE CLASSROOM TEACHERS' DECISIONS TO REFER STUDENTS FOR SPECIAL EDUCATION ASSESSMENT

Public Law 94-142, the Education for All Handicapped Children Act of 1975, places a number of procedural demands on schools and their professional personnel. The law requires knowledges and skills of educators that go beyond procedural compliance with the regulations of PL 94-142. Clearly implied in PL 94-142 is that teachers, both regular classroom and special education, have skills that are related to the education of all students, including those who are exceptional. It was in this context that Lakin and Reynolds (1983) identified ten clusters of general understandings and skills necessary for all teachers. Cluster eight includes:

the ability to perform a major role in referral and child study processes (requiring knowledge of the roles and responsibilities of support staff), and the ability to collect objective data on student social and academic behavior and to accept shared responsibilities for each student (p. 15).

Klesius (1981) reports that teachers do not refer students because of their own lack of knowledge concerning the fundamentals of measurement and evaluation in the evaluation process.

In 1979, Carpenter and Robson found that two years after the initiation of PL 94-142, a majority of special education teachers still felt they had a low degree of knowledge about the law, i.e. the referral process, dates for compliance, least restrictive environment and due process.

Algozzine and Ysseldyke (1981) reported in a study of eligibility, classification, and placement decision making, that 51% of individual decision makers declared normal students eligible for special services. They concluded that teachers based their referral decisions on stereotypes despite assessment data of a contrary nature and that the teachers' expressed a "better safe than sorry" attitude.

The results of studies done by Marston and Leslie (1983), Johnson and Cartwright (1979) and DeGenaro (1975) conclude that teachers are inadequately prepared academically to identify, evaluate and teach those students in need of special education services.

Marston and Leslie (1983) conducted a study to determine if a difference existed in the perceptions of selected components of mainstreaming between those elementary school physical educators who were instructing handicapped students and those who were not. Issues selected for the study included: 1) teacher effectiveness, 2) popularity of PL 94-142, 3) mainstreaming benefits and 4) adequacy of professional preparation programs. Perceptions held by teachers with
handicapped students enrolled in their schools were compared to teachers who did not have handicapped student enrolled in their schools. Two hundred questionnaires were mailed out to teachers of Physical Education in Iowa's public elementary schools. One hundred sixty-six usable questionnaires were collected, 144 from teachers who had handicapped students enrolled in their schools and 22 from teachers who did not have handicapped students enrolled in their schools. Analysis of the data indicated that teachers who had handicapped students enrolled in their schools perceived the benefits of mainstreaming for the handicapped students as being significantly greater than those teachers who did not have handicapped students enrolled in their schools. No significant differences were found between those teachers who did and did not have handicapped students enrolled in relation to teaching effectiveness, the popularity of PL 94-142, and quality of current preparation of physical education majors in implementing the law. Their perceptions of the quality of professional preparation indicated inadequate preservice programs as a variable in the failure to adequately implement PL 94-142.

Research conducted by Johnson and Cartwright (1979) investigated whether information about and experiences with the handicapped would improve prospective regular education teachers' attitudes toward and knowledge about mainstreaming in relation to pupil placement. An experimental group of 29 prospective regular classroom teachers were enrolled simultaneously in two courses; 1) Educational Adjustments for Exceptional Children, designed to provide information about
mainstreaming and 2) Experience with Exceptional Children, designed to provide experience with the handicapped. Two control groups consisted of 27 prospective regular classroom teachers enrolled only in the Information course and 28 prospective regular classroom teachers enrolled only in the Experience course. Participants were assessed both pre and post using the Rucker-Gable Educational Programming Scale (RGEPS), designed to measure attitudes towards and knowledge about mainstream settings. Respondents were asked to choose the best educational setting for 30 children for services ranging from regular classroom placement to placement outside regular or special education. The RGEPS was administered prior to and at the conclusion of the ten week course. The results indicated that the teachers did not increase their knowledge about mainstreaming as a result of only information about or only experience with the handicapped. However, attitudes towards mainstreaming significantly improved as a result of a combination of information and experience.

In a study designed to document teachers' perceptions of factors influencing their referral decisions Christenson, Ysseldyke and Algozzine (1982) suggest that:

the decision to refer is seen as an interaction among estimates of the students likelihood of "success" in the regular education program, institutional constraints and external pressures. Teachers' estimates of the likelihood of success are based on an evaluation of student performance derived from student characteristics, individual differences in teachers, and educational goals (p. 341).
The subjects included in the study were 47 regular education and 5 special education teachers from Minnesota and Florida. Data were categorized by first identifying five institutional constraints and four external pressures. Institutional constraints included organizational factors, availability of service, hassle, teacher variables and attitudinal factors. Approximately twenty-three percent of the respondents stated that there were no barriers to referral in their districts. Those who did identify barriers most often listed organizational factors, availability of services and hassle. External pressures included agency influences, Federal and State guidelines, parental pressure and socio-political climate. Seventy percent of the respondents said there were no external pressures to referral decisions. Teachers listing pressures most often cited agency influences and Federal and State guidelines. Other concerns identified showed that twenty-three percent of the teachers expressed concern about the length of time from referral to team decision making; forty percent of the teachers indicated time and paperwork as barriers; and fifty percent cited the competence of the contact person as either a barrier or a facilitator for referral.

In a study conducted by Ysseldyke, Algozzine and Richey (1982), it was found that placement decisions were affected by stereotypes. Teacher variables for students included the child's sex, race, socio-economic status, appearance, perceived IQ and behavior. Ysseldyke et al. concluded that decisions were influenced by expectations and
preconceived notions for students who demonstrate particular kinds of handicaps.

In a similar study, Low and Clement (1982) sought to identify the relationship of race and socio-economic status (SES) to observed classroom behavior, academic achievement and special education referral. A sample of 109 fourth grade boys from varying racial (Anglo, black and Hispanic) and SES backgrounds was selected and observed systematically on four occasions for 12 behaviors. Data also were collected on academic achievement and special education referral. Results showed black and Hispanic children did not differ, but both were lower than Anglo children in achievement. A non-stepwise discriminant function analysis utilizing race, SES and observational data as predictors of referral for special education failed to make statistically reliable predictions. In contrast to the findings of Ysseldyke et al. (1982), Low and Clement concluded that the relative unimportance of race and SES as predictors challenges the validity of alleged "institutional racism" in referrals for special education. However, the results did suggest that classroom behavior was a factor in predicting referral of special education.

Classroom behavior was the focus of a study done by Forness and Esveldt (1975). The purpose of the study was to determine whether children in the process of being referred for such problems exhibit certain behaviors that are observably different from their classroom peers and whether such behaviors relate to subsequent special
education needs. Twenty-four boys under evaluation for school learning and behavior problems were observed in their school classroom over a period of six days each. Observable behavior was compared with that of male peers in the same reading or math group. Results indicated that target subjects were significantly different in on-task behavior and in frequency of teacher response to their behavior. Forness and Esveldt suggest the use of systematic classroom observation of all children early in the year as a screening technique to predict which children are headed for serious difficulty.

In an effort to assist the teacher and other school personnel to determine if a problem with a student exists, Tucker (1982) developed a 19-step model for evaluation. Level I, the classroom level involves steps 1-6: 1) detection of problem and gathering of initial data concerning pupils' behavior and academic progress, 2) collection of more in-depth data and work samples, also frequency counts of specific behaviors, 3) informal consultation with parents, 4) design lesson plan to incorporate what has been noted in steps 1-3; 5) provide instruction according to lesson plan and 6) evaluate pupil progress and determine if the problem has been remedied or show that the problem still exists despite classroom alternatives and move to step 7, referral.

Ysseldyke and Marston (1982), using a behaviorist approach, suggest the use of observation-based assessment techniques to record and measure behavior. This procedure includes: 1) anecdotal records:
a written description of everything the observer sees; 2) direct measurement of permanent products: grade books, attendance records and pupils' academic work; 3) frequency counting or event recording: the calculation of the number of occurrences of a specific behavior during a specified interval; 4) duration: the amount of time an event or behavior occurs; 5) interval recording and time sampling: the observer decides on a recording period and divides it into equal units; the observer then records whether or not a behavior occurs in a given time sample. This approach is based on dissatisfaction with making decisions based primarily on test scores alone.

**MOTOR PROFICIENCY TESTING**

Selecting a test in the motor domain requires that the test battery measure a wide range of performance traits. Few test batteries exist which can assess both gross and fine motor skills of non-handicapped as well as Mild and Moderate Mentally Retarded (MR) children, and which have been constructed and validated for the purpose of determining physical education placement (Broadhead, 1982). Oseretsky (1948) developed a test of motor proficiency that has recently been revised by Bruininks (1978). Of the original test, 60% of the items are new and the remainder are revisions of the Oseretsky test. In the examiners' manual for the Bruininks - Oseretsky Test of Motor Proficiency, the data from the test are presented as a common set of scaled values, developed from converted raw scores. This procedure was developed so that performance on
items measured by different units could be compared and totaled into composite scores. By adding the scaled values on all test items, a child's performance can be compared with the percentile ranks derived from the national norms. A short form (14 items) of the test was constructed from data obtained in the standardization of the long form (46 items) and allows for the test to be administered in a shorter period of time (Bruininks, 1978).

Basic traits of the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency (BOT-S) were discussed in a study by Broadhead and Bruininks (1982). Subjects for the study were those from the original national standardization sample and included non-handicapped boy and girls, ages 5 through 14. The purpose of the study was to determine if traits on the BOT-S were similar to those of other tests and instruments in the motor development literature. The results from an analysis of the raw data of the original standardization sample indicated that characteristics of the BOT-S by sex and chronological age, are similar to those of other test batteries reported in the literature. Over the age span of 5 to 14 years, mean performance curves for both boys and girls were linear for all the 14 test items. The results of a correlation analysis indicated considerable independence between test items within subtests, between items of one subtest to another, and among test items of one motor area and those in another. The authors concluded that the BOT-S is a comprehensive motor assessment profile that would provide information
either supportive of a need for regular physical education or for a specially designed (adapted) program.

In a second study, Broadhead and Bruininks (1983) describe the underlying structure of the motor abilities represented by the items on the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency (BOT-S), and report the consistency of emerging factors by sex at two Chronological Age (CA) levels. The subjects for the study were 765 non-handicapped boys and girls, aged 4 1/2 to 14 1/2 years, who comprised the original national standardization sample for the Bruininks - Oseretsky Test. The data for the children were categorized by sex at two CA groups; 4 years, 6 months to 9 years, 5 months; and 9 years, 6 months to 14 years, 5 months. In assessing consistency, rigorous criteria resulted in the extraction of from three to six common and specific factors for the four groups of children. Bilateral coordination, balance, running speed and agility, and strength were the most consistent across CA and sex. The authors concluded that while the shortened form of the BOT battery is not as precise as the long form, the short version is a useful instrument in mirroring the conceptually hypothesized structure originally presented.

Beitel and Mead (1980) examined the Bruininks - Oseretsky Test of Motor Proficiency to identify appropriate perceptual - motor measures for use with young "normal" children so that in future research developmental changes could be investigated. The test was administered to 24 children 3 - 5 years of age to assess its
potential for discriminating among ages and between sexes. The investigators also sought to determine if the short form accounted for a portion of the variability of the long form. Results of the study showed a significant relationship to age for 3-5 year olds and found no sex differences. The short form accounted for 96.3% of the variability of the long form, which would allow the short form to be substituted for the complete battery whenever appropriate.

Broadhead and Church (1982) conducted a study designed to determine whether a regular or specially designed physical education program would best suit the needs of handicapped and non-handicapped children. The specific goal was to describe the utilization of data of gross and fine motor proficiency in providing objective evidence upon which physical education placement decisions can be based. Two classes of handicapped and non-handicapped children (n = 67), between the ages of 5 - 12 years, were individually administered the Short Form of the Bruininks - Oseretksy Test of Motor Proficiency. Children had been initially placed in non-handicapped Kindergarten, non-handicapped First grade, handicapped Kindergarten, and handicapped First grade physical education classes. Overall, 75% of the subjects were correctly classified, but mis-classifications occurred in each class. While the results indicated a change of placement for certain children, support was demonstrated for the initial subjectively determined placements and pointed to the validity of the test in discriminating among four levels of motor performance.
SUMMARY

There seems little doubt about the growing concern that the regular school and regular classroom teacher need training and assistance in becoming more effective in serving students with special needs. Many children have needs that are not being identified and referred either due to the subtle nature of their condition or because teachers are not adequately prepared to discriminate between satisfactory and unsatisfactory performances.

Based on a review of the literature, several major concepts concerning student observation and referral have been identified. First, it is essential that specific performance objectives be defined. By doing so, stereotypes can be eliminated and the variability among raters can be reduced. Second, raters (or observers) must be able to discriminate between behavior and performance. Whereas behavior is concerned with conduct, performance is based on quality and skill. Finally, it has been shown that training increases observational accuracy in rating performances. If teachers are to comply with the law and assess students' motor skills, they must be provided with adequate in-service training and hands-on experience. As DeGenaro (1975) states:

...many teachers assume they are not qualified to diagnose children who are suspected of having a learning disability. Meanwhile, valuable time is lost. Also, procedures are needed by the classroom teacher for identifying children who should be referred for more extensive (formal) evaluations. Further, informal assessments place the teacher in the professional position of
accepting accountability for the pupils in his or her charge (p. 557).
CHAPTER III

METHODS AND PROCEDURES

The purpose of this study was to examine how accurate classroom teachers are in identifying high and low motor performances among students. The study was conducted during the 1983-1984 school year in a public elementary school in Corvallis, Oregon. The testing was conducted during the month of June, 1984. Ninety-one children and four teachers served as subjects in the study. Subjects scoring in the upper and lower quartiles on a teacher rating form were administered the Short Form of the Bruininks – Oseretsky Test of Motor Proficiency. A two-tailed t statistic was utilized to determine if a significant difference existed between subjects in the upper and lower quartiles.

SUBJECTS

One elementary school in the Corvallis school district was selected for use in this study. The school included grades K - 5 and had a population of 192 students and 9 teachers. Participation by classroom teachers and students was voluntary. The use of human subjects was approved by the Oregon State University Committee for the Protection of Human Subjects (Appendix A).

The children ranged in age from 6 years to 11 years, 7 months. There were 44 female and 47 male subjects. All subjects were free
from any handicapping conditions which might affect performance. Four classroom teachers took part in the study: two from grade one, one from grade two and one from grade four. All four teachers had five or more years of teaching experience, and were credentialed to teach elementary age students.

TEST INSTRUMENT

Two test instruments were utilized in this study: 1) a teacher rating scale, designed by the investigator to parallel the content found in the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency, and 2) The Short Form of the Bruininks - Oseretsky Test of Motor Proficiency (BOT-S). The BOT-S was selected on the basis that it provides differentiated measures of gross and fine motor skills. Because of the number of children tested, the Bruininks - Oseretsky Test was administered in a group setting.

The items on the teacher rating scale included: 1) running speed and ability to change directions quickly; 2) ball skills, catching and throwing; 3) rope jumping and/or jumping jacks; 4) balance: example - ability to walk on straight line; 5) ability to jump for distance; and 6) handwriting skills. Directions to teachers on the rating scale were as follows: Please write in the name of each student in your class. For each of the motor skills listed, rate each child 1 to 5 (1 = low, 5 = high). Teachers were asked to make performance evaluations based on the specific age group in their classroom. A sample is found in Appendix B. Items on the Short Form
of the Bruininks - Oseretsky Test of Motor Proficiency are described in the Bruininks - Oseretsky Test Examiner's Manual as follows:

Running Speed and Agility

**Equipment** - tape measure, masking tape, block, stopwatch

**Description** - A 1-yard piece of masking tape on the floor marks the start/finish line. A timing line is marked on the floor 1-yard in front of the start/finish line. The end line is marked at 15 yards away. A block is placed on the end line. On the signal, "On you mark, get set, go" the subject runs to the end line, picks up the block, and runs back across the start/finish line. The subject is timed between the first and last crossings of the timing line.

**Rules** -
1. allow 2 trials
2. record time to the nearest 0.2 second

Standing on Preferred Leg on Balance Beam

**Equipment** - target, masking tape, balance beam, stopwatch

**Description** - Target is placed on wall at subject's eye level. Balance beam is placed 10 feet away from wall. The subject stands on preferred leg on balance beam looking at target and must maintain this position for 10 seconds.

**Rules** -
1. administer second trial only if subject does not achieve maximum score on first trial
2. stop trial and record time before 10 seconds if subject drops leg so it touches floor, drops raised leg below 45 degrees after 1 warning or hooks raised leg behind supporting leg.

Walking Forward Heel-To-Toe on Balance Beam

**Equipment** - balance beam

**Description** - The subject walks forward on the balance beam heel-to-toe, with hands on hips. The subject must make 6 consecutive steps to correctly achieve a maximum score.

**Rules** -
1. administer second trial only if subject does not achieve a maximum score on first trial
2. record number of correct and incorrect steps
Tapping Feet Alternately While Making Circles with Fingers

**Equipment** - two chairs, stopwatch

**Description** - Two chairs are placed facing each other. Subject faces examiner. Subject's arms are held at shoulder height with elbows bent and index fingers pointing toward examiner. One finger is to move clockwise and the other counterclockwise. The subject taps feet alternately while making circles with fingers. The subject has 90 seconds to complete 10 consecutive foot taps correctly.

**Rules** - 1. 1 trial only  
2. record pass/fail

Jumping up and Clapping Hands

**Description** - Subject jumps as high as possible, clapping hands in front of face as many times as possible before landing. Subject must clap 5 times above chest level to achieve maximum score.

**Rules** - 1. administer second trial only if subject does not achieve a maximum score on first trial  
2. do not count claps made below chest level  
3. mark "0" if subject looses balance and touches floor with hands

Standing Broad Jump

**Equipment** - tape measure, masking tape

**Description** - A starting line is marked on the floor. The tape measure is pulled out and fastened to the floor. Subject starts with feet apart, behind the line. From a bent knee position, subject swings arms forward and jumps as far as possible.

**Rules** - 1. administer 3 trials  
2. record distance from take-off line to back of heel nearest take-off line.

Catching Tossed Ball with Both Hands

**Equipment** - standing mat, masking tape, tennis ball

**Description** - Place standing mat on floor with rough side down. Put masking tape 10 feet from mat. Subject stands on mat and with both hands catches a ball tossed underhand from 10 feet away.
Rules - 1. 1 practice, 5 recorded
2. catch is incorrect if subject misses ball or traps it against body, steps off mat, or catches with one hand.

Throwing a Ball at a Target with Preferred Hand

Equipment - target, tennis ball

Description - A line is marked on the floor 5 feet from the wall. Target is placed on wall at subject's eye-level. Subject stands behind line and using preferred hand, throws a tennis ball overhand at the target.

Rules - 1. 1 practice, 5 recorded
2. throw is incorrect is subject misses target, throws underhand or steps over the line.

Response Speed

Equipment - 2 chairs, masking tape, response speed stick

Description - Fasten a 1 foot strip of masking tape on wall slightly below subject's shoulders when seated. Subject places preferred hand flat on wall, next to speed stick. The examiner holds the stick against the wall and then drops it. The subject uses the thumb to stop the stick as it drops.

Rules - 1. 2 practices, 7 recorded
2. readminister test trial if subject fails to look when stick is dropped or touches stick before it has been released.

Draw a Line Through a Straight Path with Preferred Hand

Equipment - red pencil, test record sheet

Description - subject uses preferred hand to draw a line through a straight path

Rules - 1. 1 trial
2. record number of errors

Copying Circle and Overlapping Pencils with Preferred Hand

Equipment - pencil, test record sheet
Description - Subject uses preferred hand to copy a geometric shape

Rules - 1. allow as much time as necessary
2. do not allow subject to rotate test sheet more than 45 degrees

Sorting Shape Cards with Preferred Hand

Equipment - shape cards, stopwatch

Description - With preferred hand, subject sorts a mixed deck of red and blue cards into two piles, separating them by color.

Rules - 1. number of cards correctly sorted in 15 seconds is recorded.

Making Dots in Circles with Preferred Hand

Equipment - red pencil, test record sheet, stopwatch

Description - Subject makes a pencil dot inside each of a series of circles. The number of circles dotted correctly in 15 seconds is recorded.

Rules - 1. 1 practice, 1 recorded

PROCEDURES

One week prior to formal testing, the investigator went to the school and explained the pilot study to the principal. The principal discussed the study with her teachers and asked for volunteers. Four teachers volunteered to participate. The teachers were given motor skill rating forms on Friday and asked to return them by the following Monday afternoon. The teachers were asked to observe and rate their students' motor performances on a scale of 1 to 5 on six separate items. No attempt was made to describe the criterion for a 1 to 5 performance. They were told only that a rating of 5 was high
and a rating of 1 was low. This was done to allow the teachers to make their own performance evaluations based on their own knowledge and familiarity with physical education skills. Data from the motor skill rating form were analyzed by an independent person to determine upper and lower quartiles. The results were not known by either the investigator or the volunteer testers. This was done to eliminate any bias in the administration or scoring of the Bruininks - Oseretsky Test of Motor Proficiency.

Prior to formal testing, volunteer testers met individually with the investigator. Each volunteer was assigned a test or tests he or she would administer. The investigator described and demonstrated each test item for each volunteer. The volunteers then rehearsed the test item. All volunteers were given a 5 X 7 card for their test or tests with specific written instructions to say to each child. The specific directions and verbal cues followed the guidelines established by Bruininks and are found in Appendix C. Each volunteer tester recorded scores for their assigned test or tests.

Formal testing took place on Thursday on the same week. On this day, the investigator was introduced to the children one class at a time. A list of names was given to the classroom teacher and those students went with the investigator to the test site. Subjects were tested one class at a time, taking approximately one half-hour for 19 to 30 children. The subjects were given a test record sheet, sent to one of ten stations and rotated around the test site until all stations were completed. The order in which the children were
assigned a station was randomly decided. All test areas were in a
group setting. Before each test item was administered, the
volunteers gave a verbal description of the test items and a
demonstration. When conducting the testing, the children were
reminded to model the volunteers' demonstrations and then move
quietly to the next station. After completing all test items the
children were accompanied back to their classroom. This procedure
was repeated until all subjects had been tested.

All subjects were tested in the same area with the same set of
equipment for the entire study. Tables and chairs were provided for
use by the school and all other necessary equipment was brought to
the test site. The testing instrument used was the Short Form of the
Bruininks - Oseretsky Test of Motor Proficiency which is contained in
a kit and includes: test manual, tape measure, stopwatch, balance
beam, tennis ball, block, masking tape, black pencils, red pencils,
response speed stick, shape cards, standing mat and target. Extra
tennis balls, pencils and stopwatches were provided by the
investigator.

ANALYSIS OF DATA

Two sets of data (test scores from a teacher rating form and
test scores from the Short Form of the Bruininks - Oseretsky Test of
Motor Proficiency) were gathered. Students scoring in the upper and
lower quartiles of the teacher rating form were administered the
Short Form of the Bruininks - Oseretsky Test of Motor Proficiency.
A two-tailed t test was utilized for analyzing the data from the Short Form of the Bruininks - Oseretsky Test. The following hypothesis was tested: There is no significant difference in motor proficiency between children who were rated high on motor performance compared to children who were rated low on motor performance by their classroom teachers.

To determine the statistical significance at which the hypothesis was to be rejected the .05 level of significance for a two-tailed test was selected.
CHAPTER IV

RESULTS AND DISCUSSION

The purpose of this study was to examine how accurate classroom teachers' observations are for identifying high and low student motor performances. In order to assess classroom teacher accuracy, two test instruments were used in the study. The first test, a teacher rating scale, was designed by the investigator to provide classroom teachers with a guideline to evaluate students' motor performances. Efforts were made to match specific skills evaluated on the teacher rating scale with those items included on the Short Form of the Bruininks – Oseretsky Test of Motor Proficiency (BOT-S). This was done to ensure that subjects were being evaluated on similar tasks. The second test which was administered was the Short Form of the Bruininks – Oseretsky Test of Motor Proficiency. The BOT-S was designed to assess the motor proficiency of children between the ages of 4 1/2 to 14 1/2 years. The subjects in this study were assessed in a group setting. Although test results may be affected by the presence of others (Hillery and Fugita, 1975; Martens and Landers, 1969; Zajonc, 1965), group testing has proved reliable for some investigators (Bowman and Dunn, 1978). A statistical analysis of the results of the study and an interpretation of these results are presented in this chapter.
RESULTS

DESCRIPTION OF SUBJECTS' PERFORMANCE

This study examined two separate conditions for the performance of elementary school-aged childrens' motor skills.

Ninety-one subjects were observed initially for six motor skill tasks by their classroom teachers and rated on a performance rating scale. Subjects were then ranked and placed into upper and lower quartile groups. There were 26 subjects in each group. However, on the day of formal testing, two subjects were absent and one subject was eliminated due to an incomplete rating by the classroom teacher. The upper quartile was then composed of 25 subjects and the lower quartile included 24 subjects. The total number of points on the rating scale was 30. Up to 5 points were awarded for each of six items on the rating scale. Subjects' scores in the upper quartile ranged from 24 - 29 with a mean score of 26.28. Subjects' scores in the lower quartile ranged from 8 - 17 with a mean score of 13.75. Those children who scored in the upper and lower quartiles on the rating scale were administered the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency. The 25 subjects in the upper quartile had performances ranging from the 27th to the 99th percentile and from the 4th to 9th stanines. The 24 subjects in the lower quartile had performances ranging from the 1st to the 99th percentile and from the 1st to 9th stanines.
HYPOTHESIS TESTING

The following hypothesis was tested: There is no significant difference in motor proficiency between children who were rated high on motor performance compared to children who were rated low on motor performance by their classroom teachers. The Statistical Interactive Programming System at the Oregon State University Computer Center was used to analyze the data. A two-tailed t statistic was used to test the hypothesis of the study. A two-tailed t test requires that only two sets of data be compared at one time. Therefore, in comparing upper and lower quartiles, it was necessary to have one set of data to represent the upper quartile and one set of data to represent the lower quartile. Mean scores for upper and lower quartile performances on motor skills were computed. The results of the data analysis are presented in Table I.

| TABLE I |

Comparison of Mean Scores Between Upper and Lower Quartile Performances

<table>
<thead>
<tr>
<th>GROUPS</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
<th>t</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPPER QUARTILE</td>
<td>63.76</td>
<td>8.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOWER QUARTILE</td>
<td>54.08</td>
<td>11.79</td>
<td>3.35</td>
<td>.0016a</td>
</tr>
</tbody>
</table>

a  t value at .05 = 2.01
   t value at .01 = 2.68
A significant difference existed between upper and lower quartile performances. The mean score for upper quartile performance (63.76) was significantly higher than the mean score for the lower quartile performance (54.08) at the .05 level. The obtained t value of 3.35 exceeded the tabled value of 2.01 at the .05 level of confidence. These results allow the hypothesis to be rejected.

**DISCUSSION**

Ninety-one children (44 female and 47 male), ranging in age from 6 years to 11 years, 7 months, served as subjects in this study. Children and teachers (3 female and 1 male), were from grades 1, 2 and 4. On the teacher rating form, children demonstrated a wide range of motor performance levels, and had test scores with a mean of 20.12. Subjects in the upper quartile had scores that ranged from 24 to 29. Subjects in the lower quartile had scores that ranged from 8 to 17.

When the mean scores of subjects in the upper and lower quartiles were compared with the scores obtained from the BOT-S it was determined that there was a significant difference between the two groups.

While these results were significant at the .05 level, a degree of caution must be used when interpreting the data. The degree of variance within the high group was 66.77 whereas the degree of variance within the low group was 139.03. Although there was no reason to expect different conditions for the two groups, the low
group was subject to some unknown factor(s). One explanation for the variation of student performances among grades is that teachers' performance expectations changes from grade to grade. In this study, two first grade classes, one second grade class and one fourth grade class were utilized. The first grade classes and the second grade class rated 13 students in the upper quartile. The fourth grade class also rated 13 students in the upper quartile. However, 23 students from grades one and two were rated in the lower quartile while only 1 student from grade four was rated in the lower quartile. Recognizing that the four teachers were directed to base their evaluations on the specific age group they were teaching, this rating distribution brings up two important questions: 1) are teachers prepared to make age-related motor skill performance evaluations, and 2) is it reasonable to assume that within each class, there will be a normal performance curve, with highs, lows, and a majority in the middle?

Another factor concerning the variance in the low group appears when comparing individual scores on the teacher rating form against scores obtained from the BOT-S. Out of 24 subjects rated in the lower quartile (1-4 staines) by their classroom teachers, only 6 scored in the lower quartile on the BOT-S. This suggests that teachers were able to correctly observe and identify only 25% of the subjects rated in the low group. It should be noted, however, that with the exception of 1 subject, those who scored in the lower quartile on the BOT-S, were correctly rated by the classroom teachers.
as belonging in the low group. The upper quartile (5 - 9 stanines) had a correct identification of 24 out of 25 subjects; a correct identification of 96%. A frequency distribution between scores obtained from the BOT-S and the teacher ratings are presented in Figure I.

FIGURE I. Stanines Derived From The Bruininks - Oseretsky Test of Motor Proficiency Compared to Teachers' Perceptions of Students' High and Low Performances (n = 49).

Subjects rated high = X
Subjects rate low = 0

As indicated in Figure I, those subjects who were rated in the lower quartile had performances that spread from the 1st to the 9th stanines on the BOT-S. However, students rated in the upper quartile had performances that spread from the 4th to the 9th stanines on the BOT-S.

Because of the large amount of variance within the low group, a chi square analysis was conducted to determine if a relationship existed between teacher ratings and scores obtained from the BOT-S. Utilizing the \( \chi^2 \) table with 1 degree of freedom and with \( \alpha = .05 \) the table value = 3.841. Since the obtained value of 23.09 exceeded 3.841, the null hypothesis was rejected. The conclusion was that the
teacher ratings and scores obtained from the BOT-S were not independent, even though the relationship was weak. The results of the chi square analysis support the findings derived from the t test. The complete contingency table with expected frequencies in parentheses is shown in Table II.

**TABLE II**

<table>
<thead>
<tr>
<th></th>
<th>AGREE</th>
<th>DISAGREE</th>
<th>TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH</td>
<td>24 (15)</td>
<td>1 (9)</td>
<td>25</td>
</tr>
<tr>
<td>LOW</td>
<td>6 (14)</td>
<td>18 (9)</td>
<td>24</td>
</tr>
<tr>
<td>TOTALS</td>
<td>30</td>
<td>19</td>
<td>49</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 23.09 \ (P \ .01) \]
\[ df = 1 \]

The frequency distribution among the low group poses two additional questions. Investigators (Forness and Esveldt, 1975; Thornton and Zorich, 1980; Low and Clement, 1982; Murphy et al., 1982) have determined that behavior is a factor that influences performance related judgements. Because no criterion was established for making motor performance evaluations, student ratings may reflect some teacher bias. In addition, the process used by the classroom teachers to administer the test may have influenced the test results. On the first test it was assumed that classroom teachers observed the
children in an informal environment, i.e. the playground or while participating in the physical education class, without the subjects' awareness. The second test, however, was administered formally and the children had been prepared by their classroom teachers to perform as best they could. The influence of a structured environment and the knowledge that performance scores were being recorded, may have encouraged higher performances from children than those previously observed by the classroom teachers.

PERFORMANCE EVALUATIONS

The results of this study demonstrate that classroom teachers are able to distinguish between high and low motor performances. The results also lend support to Gordon's (1970) findings that there is a higher percentage of accuracy in observing behavior that is performed correctly as opposed to behavior performed incorrectly. Within this study 96% of the high group were correctly observed, while only 25% of the low group were correctly observed. Although teachers were not specifically asked to indicate which students would be referred for special education assessment, it appears reasonable to assume that those students scoring in the lower quartile would be identified as those in need of special attention. Based on this assumption, teachers in this study had a tendency to overrate low performances, which were not substantiated by the results of the BOT-S. They were however, able to observe and correctly rate those who would qualify for further testing as indicated by scores obtained from the BOT-S.
In terms of efficiency and cost effectiveness, the results of this study suggest that school district administrators consider the following alternatives to the present system: 1) rely on data from a standardized screening instrument, such as the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency, to determine referral decisions and 2) provide in-service programs designed to assist classroom teachers in observing and assessing students' motor performances.

**SUMMARY**

In the present study, a significant difference was found on students' scores on the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency (BOT-S), between children who were rated high on motor performance compared to children who were rated low on motor performance by their classroom teachers. However, the results reveal a wide frequency distribution among students who were rated in the low group. Although teachers were able to correctly identify those students who scored low on the BOT-S, teachers also demonstrated a tendency to overrate low performances which were not substantiated by results obtained from the BOT-S. These findings appear to point out the need for established performance criterion as well as the need to train teachers to observe students' motor performances.

In conclusion, a response to the question which concluded Chapter III: "Is there a difference in motor proficiency between children who were rated high on motor performance compared to children who were
rated low on motor performance by their classroom teachers" will be presented.

Significantly different results were obtained from the groups of children who were administered the BOT-S. These results indicate that classroom teachers were able to discriminate high and low differences in students' motor performances. However, due to a wide variance within the low group, these results must be viewed with caution when using test data to determine student placement.

Results of this study lend support to Borman's (1978) hypothesis that performance evaluations be viewed in 3 steps: 1) observing work related behavior, 2) evaluating each of these behaviors, and 3) weighting these evaluations to arrive at a single rating on a performance dimension. Clearly implied from the results of this study was that teachers need training to observe and evaluate students' motor performances.

Three additional factors must be considered when analyzing the results of this study: 1) the effect of group testing, 2) the influence of behavior as a factor in the teachers' ratings, and 3) that only 4 teachers and 91 children were included in this pilot study.
CHAPTER V

SUMMARY AND RECOMMENDATIONS

SUMMARY

Since the implementation of Public Law 94-142, The Education for All Handicapped Children Act of 1975, the role of the classroom teacher has changed dramatically. Teachers who have been prepared to work with normal performers are now expected to identify, evaluate and teach those students who may have exceptional needs. The range of knowledges and skills required of the classroom teacher extends beyond the traditional boundaries of the classroom. Accurately observing and evaluating a child suspected of having a handicapping condition is a challenging task.

Because classroom teachers are expected to evaluate students' motor performances, it is essential to examine the skills that are necessary to implement the law. The purpose of this study was to examine how accurate classroom teachers' observations are for identifying high and low motor performances. To assess teacher accuracy in rating motor performance, an analysis was conducted on student scores on the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency between those who were rated high as compared to those who were rated low in motor performance as determined by a teacher rating scale.

Ninety-one children from Fairplay Elementary School in the Corvallis, Oregon school district serve as subjects in the study.
The subjects ranged in age from 6 to 11 years. Forty-four subjects were female and forty-seven subjects were male. Four classroom teachers, each with five or more years of teaching experience, took part in the study.

Subjects were observed initially by their classroom teachers and rated on six motor performance tasks on a rating form. Those subjects who scored in the upper and lower quartiles on the rating form were then administered the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency, which includes 14 test items of motor proficiency.

Subjects were observed by their classroom teachers on a Monday and formally tested by 10 different volunteers testers on the following Thursday. Subjects were tested one class at a time, taking approximately one half-hour for 19 to 30 children.

The test site was set up with 10 stations and subjects rotated around to each station until all items were completed. The order in which the children were assigned a station was randomly decided.

Point scores on the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency were converted to standard scores. Subjects who were rated high and subjects who were rated low on the teacher rating form were divided into two groups, providing two sets of data for the Bruininks - Oseretsky Test. Since the two sets of scores came from the same test, the two-tailed t statistic was utilized in analyzing the data and testing the hypothesis of the study. Significance at the .05 level was required to reject the null hypothesis.
Significantly different results were obtained from the group rated high as compared to the group rated low. However, subjects in the low group demonstrated a wide range of performance abilities, ranging from the 1st to the 99th percentile on the Short Form of the Bruininks - Oseretsky Test of Motor Proficiency. The results suggest that classroom teachers are more accurate in observing high motor performances than they are in observing low motor performances.

When analyzing the results, two major variables must be considered: 1) by design, specific performance objectives were not defined on the rating scale and this may have allowed teacher bias to influence student ratings, and 2) that only four classroom teachers and ninety-one children (91 children observed; 49 children tested on BOT-S) participated in the study. These numbers represent a small sample of the total population in the Corvallis School District.

CONCLUSIONS

Within the experimental limits of this investigation the following conclusions seem justified:

1. Classroom teachers are more accurate in observing high motor performances than they are in observing low motor performances.

2. Specific performance objectives should be defined in order to eliminate teacher bias from influencing students' ratings.
RECOMMENDATIONS

The following are recommendations for further study:

1. A study similar to the present study should be conducted utilizing a greater number of children and classroom teachers.

2. A study similar to the present study should be conducted which would compare the observational accuracy of teachers of Grades Kindergarten, One and Two to teachers of Grades Three, Four and Five to determine if a relationship exists between teacher perceptions and performance levels of young compared to older children.

3. A study similar to the present study should be conducted in which specific performance objectives are defined in order to eliminate behavioral characteristics or stereotypes that might influence referral decisions.

4. A study similar to the present study should be conducted to determine if sex differences, both the teachers and students, are a factor in the number of referrals to special education.

5. A study should be conducted to examine the effects of a training session in motor skill observation with classroom teachers.
BIBLIOGRAPHY


APPENDICES
APPENDIX A

OREGON STATE UNIVERSITY
USE OF HUMAN SUBJECT APPROVAL
Committee for the Protection of Human Subjects

Chairman's Summary of Review

Title: Teacher Accuracy in Observing Student Motor Performances

Program Director: Dr. John Dunn (grad. student Christine King), Phys. Ed.

Recommendation:

X Approval*  

Provisional Approval  

Disapproval  

No action

*The informed consent forms obtained from each subject need to be retained for the long term. Archives Division of the OSU Department of Budgets and Personnel Service is willing to receive and archive these on microfilm. At present at least, this can be done without charge to the research project. Please have the forms retained in archives as well as in your files.

Remarks:

Date: August 6, 1984  
Signature

If the recommendation of the committee is for provisional approval or disapproval, the program director should resubmit the application with the necessary corrections within one month.
APPENDIX B

TEACHER RATING SCALE

DIRECTIONS: Please write in the name of each student in your class. For each of the motor skills listed, rate each child 1 - 5 (1 = low, 5 = high). Please make performance evaluations based on the specific age group that you are teaching.

GRADE: __________________________

<table>
<thead>
<tr>
<th>NAME</th>
<th>Running Speed &amp; Ability to Change</th>
<th>Directions Quickly</th>
<th>Ball Skills - Catching &amp; Throwing</th>
<th>Rope Jumping and/or Jumping Jacks</th>
<th>Balance (Ex: Ability to Walk on Straight Line)</th>
<th>Ability to Jump for Distance</th>
<th>Handwriting Skills</th>
<th>Total Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXAMPLE - MARY SMITH</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C

DIRECTIONS FOR THE SHORT FORM OF THE BRUNININKS - OSERETSKY TEST OF MOTOR PROFICIENCY

1. Running Speed and Agility - "When I say ,On your mark, get set, go,' run as fast as you can to the block, pick it up, and bring it back across this line. Don't slow down; run fast across this line. On your mark, get set, go."

2. Standing on Preferred Leg on Balance Beam - Stand on the beam on your (right/left) leg and raise your other leg like this. Place your hands on your hips and look at the target. Stand like this until I tell you to stop."

3. Walking Forward Heel-To-Toe on the Balance Beam - "Place your feet on the beam like this. Place your hands on your hips. When you walk down the beam, hit the toe of your back foot with the heel of your front foot. Walk to end of the beam. Remember, keep your feet on the beam and your hands on your hips as you walk. Ready, begin."

4. Tapping Feet Alternately While Making circles with Fingers- "First tap one foot and then the other foot like this. At the same time you tap your feet, hold your arms in front of you and close your hands, pointing your first fingers to me like this."
Make circles with just your fingers; try not to move your hands, wrists or arms. Keep tapping your feet and making circles with your fingers until I tell you to stop. Ready, begin."

5. **Jumping up and Clapping Hands** - "When I tell you to begin, jump straight up as high as you can. As you jump, clap your hands in front of your face as many times as you can before you land. Ready, begin."

6. **Standing Broad Jump** - "Stand behind this line with your feet spread about as far apart as your shoulders. Bend your knees, lean forward, and swing your arms at your sides a few times. when I say go, put your arms back and jump as far as you can, letting your arms swing forward, and land on both feet. Remember, bend your knees, swing your arms back, and jump as far as you can. When you jump, let your arms swing forward, and try to land on both feet. If you lose your balance, try to fall forward. Ready, go."

7. **Catching A Tossed Ball with Both Hands** - "Stand on the mat and catch this ball with both hands when I throw it to you. Catch the ball with both hands each time I throw it to you."

8. **Throwing A Ball At A Target with Preferred Hand** - "Stand behind this line. You are to throw the ball overhand at the bulls-eye. Throw from behind this line. Ready, begin."
9. **Response Speed - Stick** - "We are going to find out how fast you can stop a falling stick. Let me show you what to do. Put your (right/left) hand against the wall next to the red line on the stick. Watch the red line on the stick. When the red line moves, stop the stick as fast as you can with your thumb. Just before I let the stick fall, I will say 'Get set' then when you see the red line move, stop the stick with your thumb as fast as you can.

10. **Drawing A Line Through A Straight Path with Preferred Hand** - "This is a road. Take the red pencil and draw a line from here to the end of the road, here. Stay inside the lines - try not to go off the road. Take as much time as you need. Ready, begin."

11. & 12. **Copying a Circle with Preferred Hand and Copying Overlapping Pencils with Preferred Hand** - "Look at the (name shape) in this box. With your (right/left) hand make one just like it in the empty box below. Take as much time as you need. Ready, begin."

13. **Sorting Shape Cards with Preferred Hand** - "When I say go, put all the red cards here and all the blue cards here. Use your (right/left) hand to sort the cards one at a time as fast as you can. Hold the cards in your other hand. Now you try it. Keep sorting the cards with your (right/left) hand until I tell you to stop. Ready, go."
14. Making Dots in Circles with Preferred Hand - "When I say go, take the red pencil in your (right/left) hand and make one dot in each circle as fast as you can. Now you try it here. Make one dot in each of these circles. Put a dot in as many circles as you can as fast as you can. Ready, begin."