This is to certify that the

thesis entitled

The Standardization of a Criterion-Referenced Test in Fundamental Motor and Physical Fitness Skills

presented by

Dale Allen Ulrich

has been accepted towards fulfillment of the requirements for

Ph.D. degree in Physical Education

Major professor

Date May 18, 1981

THE STANDARDIZATION OF A CRITERION-REFERENCED TEST IN FUNDAMENTAL MOTOR AND PHYSICAL FITNESS SKILLS

By

Dale Allen Ulrich

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Health, Physical Education
and Recreation

1981

ABSTRACT

THE STANDARDIZATION OF A CRITERION-REFERENCED TEST IN FUNDAMENTAL MOTOR AND PHYSICAL FITNESS SKILLS

By

DALE ALLEN ULRICH

The objectives of this study were: 1) identify and standardize criterion-referenced test items in the physical education domain based on the definition in Public Law 94-142; and 2) develop a set of norms for intellectually normal, educable mentally impaired, and trainable mentally impaired children in the age range of 36 months to 155 months.

Three major test functions were identified to guide in the test development process. A test user is advised to administer the test to students for the following purposes:

1. Screening for the identification of children with specific needs in the fundamental motor and physical fitness skill areas;
2. Aid teachers, administrators, and parents in making special education eligibility decisions in the physical education content area; and

3. Aid teachers, administrators, and parents in making placement and instructional programming decisions to meet the unique needs of the student in physical education.

The sample used in this study was comprised of 279 students, ages 36 to 155 months with normal intelligence or classified according to Michigan's state definitions as educable mentally impaired or trainable mentally impaired.

Three criteria were developed for the selection of specific skills within the locomotor, object control, and physical fitness skill areas. The criteria represented an effort to select skills that were relevant to the physical education content being taught in schools throughout the United States. Sixteen skills were selected to be measured by the criterion-referenced test (CRT).

This study utilized three content experts to investigate (1) content validity, (2) descriptive validity, and (3) criterion-selection validity. Two aspects of reliability were evaluated. The first indice studied was the internal consistency of the test using Cronbach's alpha coefficient and the second indice measured was the test-retest stability. The results obtained indicated excellent validity and reliability of the CRT.

The collection of student performance data was used as a field test under the same conditions in which it would be utilized. The initial analysis of the student performance data consisted of computing a three-way analysis of variance to test for differences between sex, age, student classification, and interaction effects.

Normative data were established by age and student classification. Student profiles were constructed to aid teachers in making nondiscriminatory decisions.

To my wife and colleague for her support and understanding during all those lost weekends and promised vacations.

To my mother for teaching me the meaning of sacrifice and compassion.

To my father for teaching me the meaning of hard work and honesty.

ACKNOWLEDGMENTS

The author wishes to recognize the invaluable contributions of several individuals without whose assistance this task could not have been completed.

To my friend and typist, Carol Brody, I owe a profound debt of gratitude. Without her knowledge and assistance this study would surely have been harder to accomplish.

To Dr. Janet A. Wessel and Dr. Paul G. Vogel for their professional advisement.

To Dr. D. Larry Carmichael for his professional support and friendship throughout my entire doctoral program.

To Ms. Barbara Keene for her assistance during the test development and student data collection process.

A special thank you is extended to several close friends: Crystal, Jack, Dick, and Jean.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

Page

Page

Page

Page

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

In the 1970's significant changes in public policy on education of handicapped children are reflected in the provisions of Public Law 94-142, Education for All Handicapped Children Act, and court actions. The major thrust in current legislation is to provide a free appropriate public education for all handicapped children. Two major provisions of the mandate, which are the specific concerns of the study, are the Individualized Education Program (IEP) and "Protection in Evaluation Procedures," nondiscriminatory assessment and evaluation.

Public Law 94-142, The Education for All Handicapped Children Act of 1975, mandates assessment for three purposes: to determine eligibility for special education services, instructional planning, and evaluating the effectiveness of the instructional plan. The mandate requires that state and local educational agencies shall ensure that assessment and evaluation procedures be nondiscriminatory. The rules and regulations (Federal Register, August 23, 1977) suggest the following minimum standards:

121a.532

> (a) Tests and other evaluation materials:
> (1) Are provided and administered in the child's native language or other mode of communication;
> (2) Have been validated for the specific purpose for which they are used; and
> (3) Are administered by trained personnel in conformance with the instructions provided by their producer;
> (b) Tests and other evaluation materials include those tailored to assess specific areas of educational need and not merely those which are designed to provide a single general intelligence quotient;
> (c) Tests are selected and administered so as best to ensure that when a test is administered to a child with impaired sensory, manual, or speaking skills, the test results accurately reflect the child's aptitude of achievement level or whatever other factors the test purports to measure, rather than reflecting the child's impaired sensory, manual, or speaking skills (except where those skills are the factors which the test purports to measure);
> (d) No single procedure is used as the sole criterion for determining an appropriate educational program for a child;
> (e) The evaluation is made by a multidisciplinary team or group of persons, including at least one teacher or other specialist with knowledge in the area of suspected disability; and
> (f) The child is assessed in all areas related to the suspected disability, including, where appropriate, health, vision, hearing, social and emotional status, general intelligence, academic performance, communicative status, and motor abilities (p. 42496).

The Council for Exceptional Children (CEC) and researchers in physical education also have shown concern for adequate nondiscriminatory testing. The CEC (March, 1977) established the following policy regarding nondiscriminatory evaluation:

Assessment instruments shall be appropriately adapted when used with children of impaired sensory, physical, or speaking skills and must consider each child's age and socioeconomic and cultural background.

Specialists implementing evaluation procedures must be familiar with local cultural, language, and social patterns and practices.

Tests and similar evaluation materials shall be administered in the child's primary language, wherever appropriate.

Interpreters, in the native language, and/or in sign language may be used throughout all phases of the evaluation.

All communication with parents and the child shall be in the native language of the home.

Local community norms shall be established when norm-referenced tests are used.

Criterion-referenced instruments should be used.

Developmental checklist(s) should be used where appropriate.

Instruments shall be administered only by trained personnel according to the producer's instructions.

Instruments shall assess specific abilities, not merely produce a single IQ score.

No one result shall determine placement (Exceptional Children, March, 1977).

Public Law 94-142 was designed to ensure the availability to all handicapped children of a free, appropriate education. Appropriate placement and effective education of the handicapped are ensured by the section of the Act which provides that an individualized education program (IEP) be developed for each eligible handicapped child. The requirement for formalization of goals, objectives, and procedures for evaluation provides a management tool designed to ensure and facilitate delivery of appropriate special education and/or related services to meet the unique needs of the handicapped child.

Although specific information in the IEP will vary from student to student, the regulations outlined by PL 94-142 describe the following minimum components (Federal Register, August 23, 1977):

1. A statement of the child's present level of educational performance;
2. A statement of the annual goals, including short-term instructional objectives;
3. A statement of the specific special education and related services to be provided to the child, and the extent to which the child will be able to participate in regular educational programs;
4. The projected dates for initiation of services and the anticipated duration of the services;
5. Appropriate objective criteria and evaluation procedures; and
6. Schedules for determining on at least an annual basis, whether the short-term instructional objectives are being achieved (p. 42491).

Currently, IEP's are written on approximately 155,000 students in the State of Michigan (Michigan Department of Education, 1977-1978). Projected across the United States, the figure increases to over six million. The need for appropriate, efficient assessment becomes readily apparent.

Teachers have stated the need for a means of assessing students to a) identify the presence of motor needs, b) determine eligibility for placement in the least restrictive environment (LRE) continuum, and c) aid in planning appropriate instruction. The effectiveness of the entire process of individualized educational programming hinges upon accurate assessment. Unless a teacher can pinpoint exactly what a student needs to learn, and how he learns most successfully, any attempt at individualization will fall far short of its intended goal. Assessment should be inseparable from instruction, characterizing teaching in a diagnostic-prescriptive way.

Review of current assessment instruments available to physical educators as to their conformance with the requirements of PL 94-142 and the Council for Exceptional Children shows tremendous need for revision and development of valid and reliable instrumentation in the physical education domain. Physical education is the only curricular area specifically addressed in PL 94-142 (Federal Register, August 23, 1977), and is defined as "Physical and motor fitness, fundamental motor skills and patterns, and skills in aquatics, dance and individual and group games and sports" (p. 42480). Authorities in the area of physical education for the handicapped (Cratty, 1975; Rarick, 1977, 1979; Wessel, 1980) have advocated development of valid and reliable assessment instruments for handicapped students. In the opinion of the same authorities, physical educators need to view assessment within the context of instructional placement and intervention. Also, assessment must be viewed clearly and simply as the process for collecting data for the purpose of making non-discriminatory decisions about students using the guidelines provided by PL 94-142 and CEC.

At a study conference on research and demonstration needs in 1969, physical educators listed the following two major concerns related to assessment in physical education:

1. The development of diagnostic and evaluative instruments which would effectively measure the motor performance of young children as well as children at low functional levels; and
2. The development of new testing instruments which would hopefully remedy the practice of modifying existing tools or using instruments which were originally designed for other purposes (Loovis & Ersing, 1979).

On January 20, 1977 at the State of the Arts Conference on Adapted Physical Education in Mississippi, participants expressed the following concerns:

1. The need for valid motor instrument methods;
2. The development of a motor assessment instrument to aid in determining proper placement of children; and
3. The necessity for assessments to aid in instructional planning.

On August 9, 1979 the National Consortium on Physical Education and Recreation for the Handicapped stated a priority need for the development and validation of motor assessment instruments.

Most existing instruments have the following limitations:

1. Only provide scores that are interpreted relative to a narrow, poorly described, student population;
2. Are limited in covering the scope of the motor domain;
3. Are not directly tied to curricular content;
4. Are inadequate for use with low-functioning students;
5. Are scored in a subjective manner causing low reliability estimates;

6. Are based entirely on a motor ability approach that identifies underlying strengths and weaknesses and has little research support (Salvia & Yesseldyke, 1978);
7. Are difficult to administer and interpret;
8. Provide inadequate data on validity and reliability; and
9. Measure only quantitative performance.

Assessment instruments and techniques must focus and emphasize eligibility, placement and instructional planning decisions to meet the demands of teaching in compliance with PL 94-142.

The two major diagnostic assessment approaches used in the physical education domain to provide information for systematically making the decisions referred to above are motor ability and criterion-referenced tests. Motor ability testing is an approach which attempts to identify general student strengths and weaknesses which may underlie the learning of specific motor tasks (Vogel, 1977). General weaknesses, when remediated, are presumed to facilitate the mastery of more specific tasks included in most physical education curriculums. The criterion-referenced strategy de-emphasizes assumed general abilities and emphasizes identifying the level (from a continuum of skill acquisition) a student has achieved on identified educationally relevant tasks (Vogel, 1977). Instruction is designed to move the student to the next skill level.

Both assessment approaches can identify strengths and weaknesses. The major difference in the two approaches is the interpretation as to what the observed strengths and weaknesses represent.

In the criterion-referenced approach, the interpretation of identified strengths and weaknesses is restricted to an evaluation of the present level of performance on specific target skills. The next level of performance on the skill learning continuum to be mastered and the behavioral components of that skill level become the emphasis of instruction.

Motor ability testing goes beyond observed performances and attempts to identify general abilities or deficiencies which may be the cause of obtained performance difficulties (Yesseldyke & Salvia, 1974), Instruction is then prescribed to remediate the general disabilities (e.g., coordination, balance) in hopes of improving specific educationally relevant skills. Support for this testing and interpretation approach is weak in that:

1. There is an abundance of data suggesting that skill learning is specific rather than general (Clark & Shelley, 1961; Gallagher, 1970; Henry, 1956).
2. Most standardized general ability tests do not meet acceptable reliability standards (Yesseldyke & Salvia, 1974).

3. It offers the teacher little information related to the effective conduct of daily activities (Hofmeister, 1974).

Data obtained in the motor ability approach are interpreted by comparing student scores with a standardized norm group (norm-referenced). Data obtained in the criterion-referenced approach are interpreted relative to a pre-identified criterion. The criterion can be in the form of qualitative and/or quantitative performance. On a criterion-referenced test (CRT), the teacher can report that some percentage of students can meet course objectives and interpret student status and progress on instructional content to be taught in the physical education program. Motor ability tests underrepresent the physical education content domain and are not designed to pinpoint instructional content relevant to student status and progress on course objectives (Mann, 1971; Yesseldyke, 1973; Yesseldyke & Salvia, 1974).

Glaser (1963) proposed the concept of CRT's which emphasized establishment of an individual's performance level as it relates to performance along a continuum of skill acquisition. Criterion-referenced test interpretations are useful in content areas that are cumulative and progressively more complex for the student, such as physical education where students have to reach some minimal level of proficiency or mastery before proceeding to tasks that are more advanced. A student who cannot grasp a ball should not be given instruction on a mature catching pattern. Any content where mastery is required should incorporate CRT's on an ongoing basis to verify when a student reaches the defined acceptable level.

When making decisions concerning the appropriate placement of a student along the continuum of skill acquisition, educators would be best served by CRT's directly related to the content to be learned. The present level or entry level of performance on the skills selected for inclusion in the program should be evaluated to facilitate student achievement of the next higher performance level. In this manner, CRT's aid in meeting the unique needs of the student.

Criterion-referenced tests directly linked to instructional content allows for frequent evaluation of student progress which results in facilitating appropriate changes in the day-to-day prescription of instruction. Continuous monitoring of student progress via CRT's facilitates the communication to students, parents, and administrators, and provides important data necessary for updating IEP's.

Eligibility decisions are best made by interpretation of test results in a norm-referenced manner. A school can develop eligibility guidelines by comparing the performances of large groups of students and setting cutoff scores such as the tenth percentile or minus two standard deviations below the mean when compared to the morm group. A norm-referenced test interpretation will facilitate the identification of students that are grossly deficient in the physical education domain.

One approach to assessment is not necessarily better or worse than another. Each simply serves a different purpose.

Need for the Study

Currently there are no valid and reliable criterion-referenced test batteries available for meeting the assessment, instruction and/or placement needs for delivering services to handicapped or nonhandicapped students in appropriately designed (PL 94-142) physical education programs. The construction of a valid and reliable CRT would provide physical educators with much needed standardized instrumentation. The following major functions would be served with a motor performance CRT:

1. Screening for identification of children with motor needs by specifying strengths and weaknesses in comparison to the norm group.
2. Provide input to determine eligibility for special education services.
3. Provide input to determine appropriate placement in LRE.
4. Diagnostic and prescriptive programming when tied to instructional content.

5. Evaluation for monitoring progress and revising the educational plan.

The CRT must be referenced to relevant physical education program content as defined in PL 94-142. It therefore must include items representative of:

1. Fundamental motor skills and patterns;
2. Physical fitness;
3. Aquatics;
4. Dance; and
5. Individual and group games and sports.

It must be suitable for populations of handicapped and nonhandicapped students to best serve the screening functions and for determining appropriate placement in the LRE.

Objectives of the Study

The objectives of this study were: 1) select and standardize criterion-referenced test items in the physical education domain based on the definition in Public Law 94-142; and 2) develop a set of norms appropriate for intellectually normal, educable mentally impaired, and trainable mentally impaired children in the age range of 36 months to 155 months.

Scope of the Study

The criterion-referenced test instrument was developed to assess fundamental motor skills and physical fitness, not the entire area specified by PL 94-142. The CRT items were constructed to assess the following three levels of qualitative performance: criterion, rudimentary, and assisted. The functional motor skill level was not considered in the development of the CRT. The criterion level was considered the essential prerequisite for students to progress to functional competence.

The sample used in this study was comprised of students ages 36 to 155 months with normal intelligence or classified according to Michigan's state definitions as educable mentally impaired or trainable mentally inpaired. The students enrolled in regular classes were classified as normal, learning disabled, or emotionally impaired. The educable mentally impaired students were enrolled in self-contained classes in the regular neighborhood school building. All trainable students were enrolled in self-contained classes in intermediate school districts.

Three criteria were developed for the selection of specific skills within the locomotor, object control, and physical fitness skill areas. The criteria represented an effort to select skills that were relevant to the physical education content being taught in Michigan school districts. Characteristics of well-designed CRT items were used as a guide in developing the test items

(Popham, 1978b). The following steps were followed in the item development process:

1. Select objectives from I CAN and Michigan Performance Objectives;
2. List potential item components for each objective;
3. Rate components; and
4. Write items.

Subsequent to item development, three testers were trained to reliably administer the test by viewing videotaped performances of educable mentally impaired students performing the 16 test items. Following the completion of the training, the three testers collected performance data on 279 students classified as normal, educable mentally impaired, or trainable mentally impaired. The performance data were used to construct normative tables across the three populations of students.

The validity of the test items was evaluated by the use of content experts in the physical education domain. Reliability of all test items was measured by a test-retest procedure. The internal consistency of the test was measured by computing the coefficient alpha.

Limitations of the Study

1. Sample size was unequal within groups.
2. The small sample used to evaluate test-retest stability (n = 7) was selected based on the constraints of time and facilities that occur in all school environments.
3. The selection of observable qualitative components for the criterion level of performance for each test item was not a data-based decision. Selection was based on the judgment of three persons knowledgeable in motor skill development.

Terminology and Definitions

Amplified Objective: An expanded statement of an educational outcome which provides boundary specifications regarding testing situations, response alternatives and criteria of correctness. Commonly used as a criterion-referenced test descriptive scheme that tells the user what the test is measuring.

Criterion-Referenced Test (CRT): A test designed to ascertain an individual's status on a set of pre-established educationally relevant tasks. To interpret a test in a criterion-referenced manner is to pinpoint target skills and determine whether a student has mastered or not mastered the tasks to be learned. Standardized procedures can be developed and normative data can be collected that will increase the utility of the CRT.

Standardized Test: A test designed to measure a sample of individual performance, administered according to uniform procedures, scored in conformance with uniform rules, and interpreted in reference to certain normative information and/or specific instructional content. Data on reliability and validity of the test must be provided.

Motor Ability Test: A test designed to identify general or process strengths and weaknesses (e.g., balance, eye-hand coordination, that are presumed to cause inadequate motor skill development) in order to prescribe interventions designed to remediate ability weaknesses and/or facilitate strengths. Scores are interpreted relative to well-described norm groups in which students can be compared to one another.

Educable Mentally Impaired (Michigan Requirements): A student identified by an educational planning and placement committee, based upon a comprehensive evaluation by a school psychologist, certified psychologist, or certified consulting psychologist, and other pertinent information as having all the following behavioral characteristics:

a) Development of a rate approximately two to three standard deviations below the mean as determined through intellectual assessment;
b) Scores approximately with the lowest six percentiles on a standardized test in reading and arithmetic; and
c) Lack of development primarily in the cognitive domain.

Emotionally Impaired (Michigan Requirements): A student identified by an educational planning and placement committee, based upon a comprehensive evaluation by a school psychologist and social worker, a certified psychologist, a certified consulting psychologist, or a certified psychiatrist, and other pertinent information as having one or more of the following behavioral characteristics:

a) Disruptive to the learning process of other students or himself in the regular classroom over an extended period of time.

Learning Disabled (Michigan Requirements): A student identified by an educational planning and placement committee, based upon a comprehensive evaluation by a school psychologist, or certified psychologist, or certified consulting psychologist, or an evaluation by a neurologist, or equivalent medical examiner qualified to evaluate neurological dysfunction, and other pertinent information as having all the following characteristics:

a) Disorder in one or more of the basic psychological processes involved in understanding or in using spoken or written language, which disorder may manifest itself in imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculation.

b) Manifestation of symptoms characterized by diagnostic labels such as perceptual handicap, brain injury, minimal brain dysfunction, dyslexia, or aphasia.

c) Development at less than the expected rate of age group in the cognitive, affective, or psychomotor domains.

d) Inability to function in regular education without supportive special education services.

e) Unsatisfactory performance not found to be based on social, economic, and cultural background.

Trainable Mentally Impaired (Michigan Requirements): A student identified by an educational planning and placement committee based upon a comprehensive evaluation by a school psychologist, certified psychologist, or certified consulting psychologist, and other pertinent information as having all the following behvaioral characteristics:

a) Development at a rate approximately 3 to 4.5 standard deviations below the mean as determined through intellectual assessment.

b) Lack of development primarily in the cognitive domain.

c) Unsatisfactory school performance not found to be based on his social, economic, and cultural background.

<u>General Needs Assessment</u>: The administration of criterion-referenced test representing the selected content to be learned in the physical education program which will provide the teacher with information on the student's strengths and weaknesses relative to the program objectives.

<u>Objective-Based Instructional System</u> (OBIS): An instructional program which systematically links instruction to assessed student need on stated performance objectives by providing:

a) Clearly stated goals.

b) Goal-related objectives.

c) A program organization built upon the appropriate placement of objectives from preschool through secondary levels.

d) Objective-related instructional activities and games prescribed for students based on their changing needs.

d) Objective-related student and program evaluation system.

<u>Screening and Referral</u>: A systematic process for determining the range of variability within a class or age level by comparing a student with established standardized test scores. A teacher can identify those students that are in the lowest portion of the range and refer them for a more in-depth evaluation in physical education.

<u>Nondiscriminatory Evaluation Instruments</u>: Testing and evaluation materials and procedures that are not racially and culturally discriminatory. Assessment with a single instrument, use of tests that are inappropriate for any of a variety of reasons, and testing by unqualified personnel, are but a few of the practices that allow for discrimination to occur. Perhaps the major solution to the problem is to make as few indirect assessments and predictions as possible to proceed more directly to assessments in the domain of instruction. Teachers themselves might well do most of the assessing and put the results to use immediately in their day-to-day instruction by selecting or developing criterion-referenced tests that are representative of the content to be learned. At the same time, it would be necessary and beneficial to collect normative data on the criterion-referenced test on the local level to minimize discriminatory decision-making practices.

<u>Qualitative Performance Level</u>: A mechanically mature movement pattern. The following four performance levels were used in the assessment of all 16 skills included in the criterion-referenced test:

a) Criterion level (C) - Student completes the item according to all stated criteria. Any quantitative criteria stating "consecutive trials" require performance of all qualitative criteria the stated number of times.

b) Rudimentary level (R) - Student responds according to some of the criteria but not all of the stated criteria (lacks quantitative or qualitative aspects).

c) Assisted (A) - Student needs some form of physical assistance to respond, such as manipulating the student, guiding a student's hand or tapping of student's limb. Through physical assistance, the student can perform a minimum of one qualitative criterion.

d) Other (O) - Student does not respond, responds inappropriately, resists assistance, or cannot perform a minimum of one qualitative criterion with physical assistance.

CHAPTER II

REVIEW OF RELATED LITERATURE

The objectives of this study were: 1) select and standardize criterion-referenced test items in the physical education domain relative to the requirements of Public Law 94-142; and 2) develop a set of norms for intellectually normal, educable mentally impaired, and trainable mentally impaired children in the age range of 36 months (3 years) to 155 months (12 years). The first section of the review of literature presents relevant information relating to criterion-referenced measurement. The second section presents a review of the literature pertaining to the standardization of tests. The final portion presents a review of motor skill and physical fitness assessment.

Criterion-Referenced Measurement: Nondiscriminatory Assessment and Evaluation

Evidence of bias in our educational system has long been present. While the more obvious examples of bias, such as separate schooling, are diminishing, questions of bias in assessment and placement of students with individual differences are now rising to the forefront.

Bias in schools occurs whenever educational decisions are inappropriately affected by a student's culture, race, economic background, or disability. Bias can occur in any phase of the educational program; however, the testing and placement of students in special education classes has become a major concern of those interested in equality in the schools (Bailey & Harbin, 1980).

Current attempts to reduce bias in assessment and placement include the design of new testing procedures, the use of adaptive behavior scales, the use of criterion-referenced measures, and the intepretation of assessment results using local or special group norms (Bailey & Harbin, 1980).

The emphasis of this study is on the use of criterion-referenced measures and the establishment of local or special group norms, where appropriate, in an effort to maximize nondiscriminatory educational decisions. Bailey (1979) contends that increased use of criterion-referenced tests in the evaluation process would force decision-makers to focus on the specific educational needs of children, as opposed to focusing on the labeling of students.

A criterion-referenced test (CRT) has been defined in a multitude of ways in the literature (Glaser & Nitko, 1971; Hambleton, Swaminathan, Algina, & Coulson, 1978; Harris & Stewart, 1971; Ivens, 1970; Kriewall, 1969; Livingston, 1972; Popham, 1978a). A very useful definition has been proposed by Glaser and Nitko (1971): "A criterion-referenced test is one that is deliberately constructed so as to yield measurements that are directly interpretable in terms of specified performance standards. The performance standards are usually developed by defining specific tasks that the student should perform. Representative samples of tasks from this domain are organized into a test. Measurements are taken and are used to make a statement about the performance of each individual relative to that domain" (p. 653). A criterion-referenced measurement interpretation allows a teacher to describe a student's competency on the content to be learned in absolute terms. A norm-referenced test (NRT) interpretation compares the performance of students with one another. Criterion-referenced tests supplemented with normative data also yield this potential.

A CRT approach should facilitate nondiscriminatory decision-making because the process then becomes one of 1) identifying basic skills that all students are expected to master, 2) assessing all students to determine which of these basic skills are present, and 3) designing appropriate instruction so the remaining skills can be learned.

According to Bailey and Harbin (1980), this process does have two problems. The current status of state and federal laws and funding systems requires a continuation of the labeling process which necessitates comparing one student with a norm group. It would be unwise and expensive for a school to ignore mandatory guidelines.

According to Bailey and Harbin (1980), CRT's can be very useful in making appropriate educational decisions. They can also adhere to the mandates for nondiscriminatory assessment when the following conditions are met:

1. The importance of the content measured by the CRT items and taught in the curriculum are agreed upon by culturally diverse groups within the school system.
2. Criterion-referenced items are constructed so as not to measure the skills of children from a particular cultural group unfairly (p. 593).

For a time, information from CRT's was considered inappropriate for making placement decisions because they were not supplemented with normative data. In actuality, they provide information on 1) the content to be learned, 2) the intensity of instruction needed to obtain high levels of on-task time to meet individual needs, and 3) the appropriateness of the instructional program to teach the desired content. They provide crucial information in determining the program that best meets a child's educational needs by pinpointing the appropriate content to be emphasized. The Council for Exceptional Children supports the use of CRT's as a strategy to reduce discriminatory educational decisions (Exceptional Children, 1977).

Reynolds and Birch (1977) suggest that the major solution to the problem of discriminatory assessment is to make as few indirect assessments and predictions as possible. They suggest we proceed to more direct assessments in the domain of instruction.

Local and Special Group Norms

Another approach to maximize nondiscriminatory evaluation relative to this study deals with the establishment of local and special group norms. The purpose of norms is to provide a reference for interpreting a student's performance. Therefore, a set of norms must provide a meaningful and relevant standard for comparison. A controversy over the use of local or special group norms exists concerning the conditions under which these norms are relevant (Oakland & Matuszek, 1977). Some professionals claim that many low-income and minority students are restricted by a process that judges their performance relative to the norms of the dominant culture, and that the use of special group or local norms allows comparison of the child to other children who have had similar experiences. Professionals on the other end of the controversy feel that a proper standardization sample is selected from different regions in the United States and is stratified by age, sex, socioeconomic status, and cultural groups in the same proportion as that in which they exist in society. The end result of comparing a child to special and local norm groups is that of confining him to those groups (Bailey & Harbin, 1980).

There is no easy solution to the complexities of this issue. Currently, the pros and cons of local or special group norms are philosophical in nature and not empirically supported. The need exists to systematically evaluate the effects of using local and special group norms, especially in the determination of whether all students are provided appropriate educational programs.

An interim solution may be to perform a test of significance across all stratification variables in the normative sample. If a significance does not exist, all data can be pooled and used as is. If a significance exists on any variable, then special group norms based on that variable should be used to decrease discriminatory decision-making. By providing norms based on several variables, the decision of which norms to use can be made at the LEA level.

Uses of Criterion-Referenced Tests

Millman (1974) indicated the following four general uses of CRT's: needs assessment, individualizing instruction, program evaluation, and teacher and personnel improvement. Figure 1 summarizes the specific uses of both criterion- and norm-referenced test data in education today.

Criterion-Referenced	Norm-Referenced
- Facilitate nondiscriminatory decision-making.	- Determining the range of variability in performance (individual differences) within a student, class, and/ or school.
- Instructional planning on a day-to-day basis.	- Identification of skill deficiencies.
- Designing and evaluating the IEP.	- Objective criteria for screening and referral decisions.
- Placement decisions within the continuum of skill development.	- Objective student data-based special education eligibility and LRE placement criteria.
- Provide specific content-related feedback to students, parents, and administration.	- Facilitate setting reliable student performance expectations.
- Program evaluation.	

Figure 1. Uses of criterion- norm-referenced test data in education today.

By identifying basic skills that most students are expected to achieve in the physical education motor skill and fitness domain; assessing all students, handicapped and nonhandicapped, to determine which of these skills are present, a CRT should facilitate nondiscriminatory decision-making. The criterion-referenced testing should be limited to direct assessment of the student on the content to be learned rather than trying to predict how a student will do by indirect assessment of the student's general abilities.

A CRT can be used as an evaluation device for the teacher to make effective decisions in the individualization of the student's physical education program. The individualization process should be based on the assessed needs of each student on the content to be learned. By using a CRT the teacher can determine precisely which components of a skill each student has mastered and those that are lacking and need instruction.

Another aspect of the individualization process that can be served by a CRT is determining the intensity or rate of instruction needed for the student to progress toward specified goals. If a student is deficient in prerequisite skills necessary to master the desired content, the student must be provided a higher intensity or on-task time to learn those skills. The placement of this student where a maximum on-task time can be achieved along the LRE continuum is suggested.

The final aspect of the individualization process is designing and evaluating the appropriateness of the instructional program to teach the skills that are needed. The instructional program must consist of activities, drills, games, and instructional sessions designed to take the student from the present level of performance to the expected level on the content of the program. It is necessary to continually evaluate the student's progress on the specified content and make modifications where needed based on the student's changing status. A CRT linked directly with the content of the program allows for continuous assessment on a day-to-day basis. If no progress is observed over a specified period of instructional time, the teacher must prescribe different activities and use a variety of techniques and procedures that are linked to the program content. When the CRT only represents a sample of the instructional performance objectives, the assessment will not be as specific, resulting in broader unit prescriptions.

School personnel responsible for developing and evaluating a student's IEP will gain valuable information from the use of CRT's. Teachers will be able to develop the IEP in terms of goals and instructional objectives based on the assessed needs of each student on the content of the program. The teacher will also have a simple instrument to continuously monitor the student's progress throughout the implementation of the IEP. Assessment will become part of instruction.

The use of a CRT to assess the student's present level of performance will provide the teacher with student data to aid in appropriately placing the student within the continuum of motor skill development. It will provide an entry point into the physical education program.

Another use of a CRT in education today is based on the mandate requiring educational accountability. The CRT can provide specific content-related feedback to students, parents, and administration on exactly what skills and subskills the student can perform (Millman, 1974).

The final use of CRT's is for program evaluation. A CRT will provide student data over an extended period of time to evaluate the effectiveness of the physical education program. Subsequent to an objective program evaluation, modifications can be implemented and re-evaluations can be performed (Hambleton & Gifford, 1977; Millman, 1974; Popham, 1975).

A review of the literature reports the development and implementation of a diverse collection of alternative educational programs that seek to improve the quality of education for students with individual differences by individualizing instruction (Gibbons, 1970; Gronlund, 1974; Heathers, 1972; Wessel, 1976). A common characteristic of many of the new programs is that they are goal-directed and defined in terms of instructional objectives and are generally referred to as "objective-based."

The overall goal of an OBIS is to provide an educational program which is maximally adaptive to the requirements of the individual learner. Among the best examples of objective-based instructional systems are Individually Prescribed Instruction (Glaser, 1968, 1970); Program for

<u>Learning in Accordance with Needs</u> (Flanagan, 1967; 1969); <u>The Individualized Mathematics Curriculum Project</u> (DeVault, Kriewall, Buchanan, & Quilling, 1969); and <u>The I CAN Objective-Based Physical Education Program</u> (Wessel, 1976, 1980).

One of the underlying premises of objective-based programs is that effective instruction depends on a knowledge of what specific skills the student has. The tests that measure student progress must be closely matched to the instruction. Over the years, standard procedures for testing and measurement within the conext of traditional educational programs (norm-referenced approach) have become well known to educators; however, the procedures are much less appropriate for use within objective-based programs (Glaser, 1963; Hambleton & Novick, 1973; Popham & Husek, 1969) because they are not closely matched to instructional content. As an alternative to the traditional norm-referenced techniques, criterion-referenced tests directly linked to the objectives of the program have been introduced to meet the testing and measurement requirements of objective-based programs.

Screening, placement, instruction, and evaluation can and should be based on the identified tasks of importance for students to learn and on their status and needs on those tasks (Duffy & Fedner, 1978).

Standardization Procedures for Criterion-Referenced Tests

According to Mehrens and Lehmann (1978), a standardized test is one that provides methods for obtaining samples of behavior under uniform procedures. The test is administered under uniform conditions with the same set of directions and equipment, and the scoring procedure is carefully delineated and uniform. Usually, a standardized test has been administered to a norm group or groups so that a student's performance can be interpreted in a norm-referenced fashion. Ebel (1979) states that it is imperative that a standardized test has explicit instructions for uniform administration and has tables of norms for score interpretation derived from administration of the test to a defined sample of students. A major trend in criterion-referenced test standardization is the collection of normative data to increase its utility in making appropriate educational decisions (Popham, 1976; 1978a).

The following general guidelines summarize the CRT characteristics that should be evaluated by educational personnel responsible for developing or selecting a standardized test for making effective educational decisions:

1) Function and purpose of the test (Kosecoff, Fink, & Klein, 1976);
2) Preparation or selection of objectives measured by the test (Hambleton & Eignor, 1978; Kosecoff et al., 1976; Popham, 1978a, 1978b);

3) Test item development (Hambleton & Eignor, 1978; Kosecoff et al., 1976);
4) Directions for administration and scoring (Hambleton & Eignor, 1978; Kosecoff et al., 1976; Mehrens & Lehmann, 1978);
5) Normative data (Hambleton & Eignor, 1978; Kosecoff et al., 1976; Mehrens & Lehmann, 1978; Popham, 1976, 1978a);
6) Test score interpretation (Hambleton & Eignor, 1978; Mehrens & Lehmann, 1978);
7) Reliability (Hambleton & Eignor, 1978; Kosecoff et al., 1976; Mehrens & Lehmann, 1978; Popham, 1978a, 1978b); and
8) Validity (Hambleton & Eignor, 1978; Kosecoff et al., 1976; Mehrens & Lehmann, 1978; Popham, 1978a, 1978b).

The remainder of this section will be devoted to a more specific explanation of the above guidelines for standardized criterion-referenced tests.

Function and Purpose of the Test

Criterion-referenced tests are most commonly used in two contexts: (1) instructional diagnosis and planning; (2) student and program evaluation (Kosecoff et al., 1976). In the first context, a CRT can diagnose specific content-related needs of each learner when it is directly linked to the instructional program. Identification of specific strengths and weaknesses on the objectives of the physical education program is used as an assessment of the present levels of performance on a continuum of development leading toward the acquisition of target objectives. The next level of performance to be mastered and the behavioral components of that skill level become the emphasis of instruction. When the objectives of an OBIS are arranged in an instructional sequence which ranges in ability from little competence to a mature or functional level, the objectives become the pool of potential items. Instruction is prescribed in accordance with the assessed needs of students on the objectives. A match between the physical education program and the individual student is made by selecting those objectives within the program that accommodate the unique needs of various students and constructing a test to fit the content.

In the second context, student achievement can be documented through reassessment during and at the end of instruction by the use of CRT's reflecting the content of the selected program objectives for each student. Continuous monitoring of each student's progress on target objectives is facilitated by incorporating an OBIS and a criterion-referenced test within the program organization. Decisions concerning the impact of instruction on each student and the work of the overall program can be made by documenting the learner outcomes on the objectives of the program measured by the CRT's.

Preparation or Selection of Objectives

The basic feature of CRT's is their foundation on clearly defined educational tasks and purposes which constitute the test's domain specifications. Objectives for CRT's can be developed or selected in at least three ways (Kosecoff et al., 1976):

1. Expert judgment. Domain experts assess the educational tasks that are the most relevant to measure and teach within a specific domain (Martuza, 1977).
2. Consensus judgment. Various groups such as community representatives, teachers, curriculum specialists, school administrators, and parents decide which educational tasks are the most important and form a pool of potential possibilities (Wilson, 1973).
3. Theories of learning and instruction. A literature review is conducted to formulate series or hierarchies of educational tasks (Keesling, 1974).

The initial and most important quality of a well-designed CRT is a descriptive scheme, that, with no ambiguity, spells out just what it is that students who take the test can or can't do. Sometimes these descriptive schemes are referred to as test specifications, item forms, or amplified objectives. The descriptive mechanisms are the verbal vehicles that render CRT's useful to educators. Good CRT's must be focused on a limited number of significant learned behaviors. At the same time, the small number of important behaviors being measured must still be described with sufficient clarity to communicate unambiguously what is being measured (Popham, 1978b).

It is important that a CRT be based on clearly defined objectives and to choose a representative sample of items. More useful than behavioral objectives are "amplified objectives." According to Millman (1974), "An amplified objective is an expanded statement of an educational goal which provides boundry specifications regarding standard testing situations, response alternatives (where appropriate) and criteria of correctness" (p. 335).

In light of the recent emphasis on objective-based instructional systems with documented goals and program objectives which specify the content to be learned, criterion-referenced tests can be constructed that represent the goals and objectives that represent physical education domains.

Hambleton and Eignor (1978) suggest the following questions concerning the objectives measured by a CRT:

1) Are the objectives stated in a clear and concise fashion?
2) Can a potential user "tailor" the test to meet local needs by selecting objectives of interest?
3) Is there a match between the content measured by the test and the situation where the test is to be used?
4) Does the set of objectives measured by the test serve as a representative pool from some content area of interest?

Test Item Development

Several approaches to the generation of criterion-referenced test items have been proposed by Anderson (1972), Bormuth (1970), Hively, Maxwell, Rabehl, Senison, and Lundin (1973), and Osborn (1968), but none is appropriate in the motor skill and physical fitness domains. The most popular and widely used of the newer techniques is the "amplified objective" approach of Popham and Baker (1973). This approach begins with an instructional objective and consists of a response description, content limits (essentially a rule for determining the content relevant to the achievement of the objective), a detailed description of the characteristics of the item and the appropriate means of responding to it, standard scoring criteria, and standard item directions and equipment. The popularity of this technique is a result of its wide applicability and because it is an extension of the well-formed objective. It appears to be the most practical approach to content definition in the motor skill and physical fitness domain.

It is imperative to maintain a nondiscriminatory approach to item development. The two most important aspects of the item development process that must be considered to maximize nondiscriminatory evaluation are (1) content and (2) wording. The content of a CRT item must be made up of skills that all students can be expected to learn. The wording of an item is offset by presenting it in the child's native language and presenting an appropriate demonstration of the skill according to the set criteria.

According to Hambleton et al. (1978), the quality of CRT items can be determined by the extent to which they reflect, in terms of their content, the objective from which they were developed.

A common approach used to determine the validity of the content of CRT items involves judging each item by content experts. The judgments that are made concern the match between an item and the objective that it is designed to measure.

Two strategies for the collection and analysis of the judgments of content experts were described by Rovinelli and Hambleton (1977) and others are offered by Popham (1975). Rovinelli and Hambleton (1977) asked content experts to rate test items relative to a set of objectives. The following three possible ratings of a test item were used:

+1 = definite feeling that an item matches the objective;

0 = undecided about the item-objective match;

-1 = definite feeling that an item does not match the objective.

A second strategy used by Rovinelli and Hambleton (1977) incorporated the use of a four-point rating scale. Content specialists are provided an objective and a set of test items. The task is to judge the appropriateness of each item as a measure of the objective. The data obtained are the average ratings across content experts.

Hambleton and Eignor (1978) suggest the following questions concerning the items in a criterion-referenced test:

1) Are the test items valid indicators of the objectives they were developed to measure?
2) Do the test items represent content that is important for students to learn?

3) Are the test items in an appropriate format to measure the objectives they were developed to measure?

Directions for Administration and Scoring

One factor strongly affecting a CRT's usability is the training necessary to administer the test reliably. Since few school systems have personnel specially trained to administer all testing programs, a CRT intended for use in a classroom context has greater utility if it can be administered by the student's teacher or by a paraprofessional.

A test is more practical if the instructions to the examiner and the student are clear, complete, and well organized (Walker, 1978). Uniform equipment and materials must be easily obtained in an educational setting. Elaborate and special equipment decreases the CRT's utility in a classroom setting. A test is more practical if it can be scored easily and objectively.

Good records of student performance are an important part of classroom management and meeting accountability requirements. A testing system is more practical when it has usable forms for recording students' test scores that are easily keyed to the objectives, easy to maintain, and easy to interpret (Walker, 1978).

The order in which the individual items that comprise the CRT must be administered has important consequences for a CRT's administration. For example, CRT's that require a prescribed order for testing have limited work with curriculums that follow another sequence (Kosecoff et al., 1976). A major consideration related to a CRT's administrative adequacy is the extent to which the instrument can be used to make educational decisions (Kosecoff et al., 1976; Walker, 1978). A promising practice in the last few years is the referencing of objectives and test items to specific instructional materials (Bagnato, Laub, & Kurtz, 1978; Loovis & Ersing, 1979; Wessel, 1976, 1980). Hofmeister (1975) states, "CRT's can reach their full potential only when they are integrated into the day-by-day functioning of the classroom" (pp. 77-78).

Hambleton and Eignor (1978) suggest the following questions concerning the administration of standardized CRT's:

1. Do the test directions include information relative to test purpose, equipment and materials, and scoring?
2. Are the test directions clear?
3. Is the test easy to score?

Collecting Normative Data

According to Popham (1976, 1978a), rather than denouncing normative data, supporters of CRT's should encourage designers of such measures to collect data regarding how various groups of students perform on the tests. With adequate comparative data, the administrative utility of the CRT is increased. Although CRT's permit educators to describe the extent to which a student possesses a specific skill, it doesn't automatically inform concerned teachers and parents how well a student should be expected to perform with respect to the skill in question.

The major reservation that some educators have about norm data for CRT's is that they will cause the test to lose its descriptive clarity. Popham (1976) says:

> You don't lose clarity of description by augmenting a test with comparative data, you merely pick up some information that's useful in setting reliable performance expectations (p. 594).

Initially, the use of comparative data will help teachers and others responsible for developing goals and objectives for students in physical education. If a student performs one and one-half to two standard deviations below the mean on fundamental motor skills when compared to student peers, a realistic goal for the student in physical education would be to develop competence in selected fundamental motor skills. It is important to select appropriate fundamental motor skills or teachers may not have enough instructional time to facilitate improvement in performance.

Normative data on a CRT will provide educators with decision-making criteria based on student performance data. The following major decisions can be made more efficiently in the presence of normative data representing student performance on the content of the local physical education program:

1. Special education eligibility in physical education;
2. Appropriate placement in the physical education program;
3. IEP development in terms of goals and objectives.

The decision that a student is in need of special education services in physical education can be based on how the student performs on the content of the program when compared to student peers. If the results of assessment show that the student is grossly deficient in motor skill and/or physical fitness, the student should be eligible for special education services in these areas, regardles of whether the student is considered handicapped or nonhandicapped. Local education agencies (LEA) can determine their own criteria for eligibility (i.e., one, one and one-half, or two standard deviations below the mean on the content when compared to their peers), or, where available, use the criteria established by the state.

If the student is grossly deficient on the content (depending on the LEA criteria for such a classification), then a placement within the physical education program where the student will get the most effective instruction (remedial class, additional instructional periods, small group instruction, or self-contained class with peer tutors, volunteers, or paid aides) can be made. An example of decision criteria for making placement decisions would be when an LEA decides that a remedial physical education class is for students deficient in only one or two skills with the expectation of remediating these skills in a relatively short time. An LEA might set criteria for regular education placement in physical education so that a handicapped student that performs within one standard deviation below the mean on the content of the program can participate in the regular program.

In the educational context, the term norms generally refers to the statistical information which describes the distribution of scores of a well-defined sample of students and it provides evaluative information about a student's level of performance when compared to the norm sample (Mehrens & Lehmann, 1978). The statistical information may be presented in a variety of forms including 1) summary statistics such as means and standard deviations, 2) conversion or norms tables which show the association in the norm sample between each possible raw score value and the matching values on derived scales (e.g., T, percentiles, stanine), and 3) student profiles which show at a glance the performances of the norm group on a number of simultaneous dimensions (Martuza, 1977). Most standardized tests tend to present normative data in several forms to maximize the interpretability and ease of use for varied consumer groups.

According to Mehrens and Lehmann (1978), normative data must be recent and representative of students found in most schools. As the content of a physical education program changes, not only the norms but the test itself becomes outdated. If the characteristics of the reference group have changed, then the normative data are obsolete and should result in the collection of new norms based on the present characteristics of the students.

If a normative sample is not representative of the general population being assessed, then sampling error occurs. According to Mehrens and Lehmann (1978) and Ebel (1979), a normative sample should generally be stratified by sex, age, and race when used in an educational setting. The relevance of the norm group is dependent upon the degree to which the population sampled is comparable to the group with which users of the test wish to compare their students (Mehrens & Lehmann, 1978).

Hambleton and Eignor (1978) suggest the following questions concerning standardized normative data:

1) Are the norms reported in an appropriate form?
2) Are the samples of students utilized in the norming process described?

Test Score Interpretation

The interpretation of test scores is important if reliable decisions are to be made in an educational setting. The two most important factors concerning the interpretation of a student's test scores for classroom use are 1) ease of interpretation; and 2) accuracy. According to Ebel (1979) and Martiza (1977), student test profiles are convenient ways of showing test scores; they are graphic devices enabling educators and parents to see the overall performance of a student at a glance.

They provide an excellent means for gaining a comprehensive picture of the strengths and weaknesses of a student or class. In general, profiles are used to show two or more scores for the same student.

Before a profile can be plotted, it is necessary to transform the scores to sets of comparable values. One approach is to convert the raw scores into some type of derived scores prior to plotting them. The most common method is to use either standard scores, percentile ranks, or stanines (Martuza, 1977). Another common approach is to scale the raw scores on the profile itself so that each scale has an equivalent mean and unit of measurement.

Walker (1978) feels that because tests are devices for making decisions about students, they should be constructed in a way that allows decisions to be made with confidence and ease. The information for decision-making should be easy to find, easy to use, and well justified. Although the decision criteria for special education eligibility and placement should be left up to the local test users, the developer should give an indication of the consequences of choosing different criteria.

The two most common uses of profiles are 1) diatnosis, and 2) planning (Martuza, 1977). For the purpose of diagnosis, educators and parents examine the profile for the most obvious skill strengths and weaknesses shown by the student's performance on the CRT.

For the purpose of planning, educators would prescribe instruction based on specific student weaknesses. Educators, parents, and other school personnel could base the development of a student's IEP on the performance profile. If the student was deficient on a certain number of fundamental motor skills (depending on LEA criteria), an appropriate goal area in physical education would be to develop competence in fundamental motor skills. This would be placed along with specific objectives in the student's IEP.

Hambleton and Eignor (1978) suggest the following questions concerning the score interpretations of a standardized CRT:

1. Are suitable guidelines included in the manual for interpreting individual and group objective score information?
2. Are appropriate guidelines offered in the manual for utilizing test scores to make descriptive statements, instructional decisions, program evaluation decisions, or other stated uses of the test scores?

Reliability

The extent to which a test measures with consistency is referred to as its reliability. The more consistently a test assesses whatever it is measuring, the more reliable it is (Popham, 1978a). It is imperative that, whatever the technique involved in estimating a CRT's reliability, there be at least a small amount of examinee response variance, or else the results of the analysis will be essentially worthless (Popham, 1978a; Swaminathan, Hambleton, & Algina, 1974).

According to Mehrens (1980) and Ebel (1980), traditional correlational strategies used to determine the reliability of norm-referenced tests are suitable for CRT's as long as there is at least a small amount of performance variability. Performance variability is not a necessary requisite for a good CRT. When no variability is present, new techniques to determine reliability must be used.

The most important types of reliability indices used for standardizing tests are: test-retest stability, equivalence and stability, and internal consistency (Popham, 1978a). According to Mehrens (1980), stability and internal consistency are the most appropriate for a CRT in the physical education area.

The test-retest procedure for assessing the reliability of a test in a particular examinee sample requires two administrations of the same test, separated by a reasonable period of time (Martuza, 1977). According to Popham (1978a), the actual duration of the delay is particularly crucial. If the between-testing period is too long, significant events may have occurred that would effect the test's ability to produce consistent scores. Generally, the interval between testing is between one and two weeks, long enough so that the students' recall of the initial testing will not significantly influence their second performance, but not too long so as to permit learning of the tested behaviors. Generally, the relationship between the test-retest scores is calculated via an available correlational technique such as the Pearson product-moment correlation coefficient with hopes of obtaining a stability coefficient of at least .80 (Mehrens, 1978). It is recommended that higher coefficients be used for making individual placement decisions.

With norm-referenced interpretations, internal consistency estimates constitute the most widely used procedure of measuring reliability (Popham, 1978a). Internal consistency estimates attempt to measure the amount of consistency among the test items. Internal consistency does not measure the reliability of decisions resulting from a test, but only the characteristics of the items themselves.

According to Ebel (1980), Martuza (1977), and Mehrens (1980), the alpha coefficient provides the best measure of internal consistency because of its application to any particular set of test data. Using the coefficient alpha (Cronbach, 1951), it is possible to obtain reliability estimates from only one set of test data. This is accomplished by using the mean of all split-half coefficients resulting from different splittings of the same test. Alpha provides a general method for assessing the reliability of a composite test using information on the component parts of the test.

Martuza (1977) and Mehrens (1980) offer one caution when using alpha as a measure of internal consistency of CRT's. They state that the values obtained are directly related to the variability in test scores. The amount of variability in such a distribution is typically quite low. As a result, an internally consistent test may yield a relatively low alpha coefficient. If this is the case, a newer technique such as Cohen's (1960) kappa statistic should be used. According to Mehrens (1980), as long as an acceptable reliability coefficient results, the amount of variability is unimportant.

The basis for much of the recent literature on reliability for CRT's is provided by Hambleton and Novick (1973), Popham (1978a), and Swaminathan, Hambleton and Algina (1974). Hambleton and Novick (1973) suggest "the traditional concepts of reliability and validity could be replaced by a complete decision-theoretic formulation" (p. 168). In cases where a CRT is used to categorize people into two groups, the metric of major interest for the test is a zero-one score where non-masters receive zero and masters a one. Reliability can be addressed in terms of the zero-one metric rather than the number-right score of another metric. Since product-moment correlations have undesirable properties for zero-one variables, Hambleton and Novick suggested that "an alternative measure of reliability might simply be the proportion of times the same decision would be made with repeated measurements" (p. 168).

Swaminathan, Hambleton, and Algina (1974) considered using the simple proportion of agreement between decisions on two administrations, but rejected this as an index because it "does not take into account the agreement that could be expected by chance alone" (p. 264). To adjust for chance agreement, Swaminathan et al. (1974) proposed the use of Cohen's (1960) kappa statistic. Kappa is the proportion of agreement uncontaminated by chance. Kappa can range from -1 to +1 with a positive 1 indicating perfect consistency and zero indicating chance agreement. Negative values imply that the observed agreement is worse than that which would be expected by change.

Moyer (1976, 1977) conducted a study comparing reliability results based on traditional strategies and those based on kappa's consistency of classification. She determined that the two techniques yielded very similar information for the tests used in the Michigan Assessment. Strasler and Raeth (1977) reported similar findings. Moyer (1976) concluded that for practical purposes the traditional measures were preferable since they are more familiar and yielded results similar to kappa.

An additional alternative presented in the literature is a coefficient of reproducibility. Cox and Graham (1966) suggested this coefficient for estimating the reliability of a decision-oriented measure. The coefficient of reproducibility is appropriate for some items that are sequentially scaled such that a person is expected to pass items up to a certain point and fail items beyond that point. The items should be expected to approximate a Gutman scale (Torgerson, 1958).

It is apparent by the above discussion on measuring the reliability of standardized CRT's that several potential procedures exist. Those procedures that are the most practical and traditional should be attempted initially. If low indices of reliability occur due to a lack of variability in scores, then the newer techniques based on the percent of agreement across several administrations may have more utility.

Validity

According to Ebel (1979) and Mehrens and Lehmann (1978), the degree of validity is the single most important aspect of a test. Validity can best be defined as the degree to which the test is capable of achieving the aims of the user (Mehrens & Lehmann, 1978). A system developed by Kosecoff et al. (1976) for describing and evaluating CRT's suggests several dimensions that can be used to validate a CRT: 1) content validity, 2) descriptive validity (item-objective congruency), and 3) criterion selection validity.

Content validity can be established by determining if the skills selected are representative of skills commonly taught in most school curriculums. This can be accomplished by surveying curriculums and/or having content judges evaluate the selected skills for their relevancy.

Item-objective congruence (descriptive validity) can be established by using judgmental data. Usually, content experts are given a variety of objectives and the items used to measure them, and are requested to comment on the appropriateness of the item-objective relationship. Popham (1978a) considers this notion of item-objective congruency as the test's descriptive validity (whether or not the items are congruent with the test specifications).

Popham (1978a) suggests the following procedure for determining the descriptive validity. Locate several content judges and ask them to first read the set of objectives and then judge, on an item-by-item basis, whether the item is congruent with its objective. All the judges need to do is go through the items and check those that are incongruent; then compute the percentage of congruent items as seen by the judges; then calculate the mean percentage across all judges. Popham goes on to suggest that congruency percentages of 90 or higher would appear to be satisfactory.

In certain content domains, such as math concepts or word recognition, it may also be necessary to evaluate the proportion of items representing each objective.

According to the system proposed by Kosecoff et al. (1976) and Popham (1978a), the third dimension is criterion selection validity. This is a procedure for verifying the importance of the behaviors used in a test item criterion. The criterion can be qualitative and/or quantitative in nature. This can be achieved by reviewing the research and selecting relevant behaviors and then asking a group of individuals knowledgeable in the content area to judge their importance. Kosecoff et al. (1976) feel that it is best when a CRT is based on objectives that are narrowly defined and operationally stated in such detail that developing items requires only transposing the objectives into test form. Criterion-referenced test score interpretations of objectives with these characteristics are meaningful because the objectives describe skills that can be measured directly by test items.

An example that only requires determining the relevancy or importance of the objectives selected to meet criterion validity is when objectives are selected directly from a validated curriculum. Criterion-referenced test score interpretations are significant for these objectives because the skills measured are being taught in classrooms using a specific curriculum.

Kosecoff et al. (1976) suggest a second step in establishing criterion validity. This involves the use of empirical means in determining whether students who perform well on the test have actually achieved the objective. This can be assessed by comparing results obtained by examinees who, prior to taking the CRT and using independent criteria, are judged to be masters or non-masters of the skills that the objective is intended to measure. To the degree that the CRT discrimination between these two groups of students, the CRT has criterion validity.

Summary

In an effort to maximize nondiscriminatory assessment and evaluation, educators have turned to the development and use of criterion-referenced tests that are directly linked to the content of instruction. A CRT approach facilitates nondiscriminatory decision-making by 1) identifying basic physical education skills that are important for students to learn, 2) assessing all students to determine which of these basic skills are present, and 3) designing appropriate instruction so the remaining skills can be learned. To increase the utility of a CRT, it appears most appropriate to collect normative data on specific groups of students so that norm-referenced decisions can also be made. The major use of normative data on CRT's is to provide educators and parents with student data-based decision criteria.

The following eight characteristics are important when developing or selecting a standardized CRT: 1) function or purpose of the test; 2) preparation or selection of objectives measured by the test; 3) test item development; 4) the directions for administration and scoring; 5) normative data; 6) test score interpretation; 5) reliability; and 8) validity.

The following conclusions related to standardizing CRT's have been drawn from this review:

1. Criterion-referenced tests must be directly related to the content to be learned in the physical education program.
2. Criterion-referenced test items must be based on clearly defined educational tasks.
3. Content experts' judgment of the educational tasks that are the most relevant to measure and teach in physical education is an acceptable validation strategy.
4. The CRT items should be based on a sufficiently limited focus of relevant and observable behaviors.

Motor Skill and Physical Fitness Assessment

Assessment in the physical education motor skill domain has become increasingly important with the passage of Public Law 94-142 (94th Congress, 1975). Special treatment given to physical education in this law is demonstrated in various components of the definition of special education (Section 121a.14 Federal Register, 1977):

> 121a.14 Special education
> (a)(1) As used in this part, the term "special education" means specially designed instruction, at no cost to the parent, to meet the unique needs of a handicapped child, including classroom instruction, instruction in physical education, home instruction, and instruction in hospitals and institutions.
> (b) The terms in this definition are defined as follows:
> (1) "At no cost" means that all specially designed instruction is provided without charge, but does not preclude incidental fees which are normally charged to nonhandicapped students or their parents as part of the regular education program.
> (2) "Physical education" is defined as follows:
> (i) The term means the development of;
> (A) Physical and motor fitness
> (B) Fundamental motor skills and patterns; and
> (C) Skills in aquatics, dance, and individual and group games and sports (including intramural and lifetime sports).
> (ii) The term includes special physical education, adapted physical education, movement education, and motor development.

A careful examination of the definition of special education reveals that physical education is the only curriculum area specifically mentioned in the law (Federal Register, 1977). The specific inclusion of physical education in the law necessitates that it be addressed in the Individualized Education Program (IEP) Committee (IEPC) meeting for each child identified as a possible candidate for special education. Therefore, if reliable decisions are to be made by the IEPC, they must be based on a clear picture of what the student can and cannot do (criterion-referenced interpretations) in the physical education motor skill domain and how these performances compare to their student peers (norm-referenced interpretations).

Section 121a.530 of PL 94-142 mandates the following seven guidelines to be followed to protect students during evaluation procedures (see Appendix A for a detailed list of suggestions and applications to physical education):

1) Be provided and administered in the student's native language;
2) Be administered by trained personnel;
3) Be provided with student data in all areas related to the suspected disability;
4) Be tailored to assess specific program areas;
5) Be selected and administered to ensure non-discrimination;
6) Not to be a single procedure; and
7) Be conducted by a multidisciplinary team.

Criterion-Referenced Measures Useful in Motor Skill Assessment

Recalling the discussion in an earlier section of this chapter pertaining to the characteristics of an objective-based physical education system, a reader is reminded that the center of an OBIS is found in its objectives. The most important characteristic of the objectives for the system to be useful is that the objectives must be measurable (Vogel, 1980). For objectives to be measurable, educators must be able to reliably assess whether a student has mastered the objective. According to Yelon (1979), a good objective has three major components: 1) the given conditions that are important to the behavior being measured must be identified; 2) the behavior being measured must be stated in observable terms; and 3) the standards for acceptable performance must be identified. Objectives with the above components represent instructional outcomes.

According to Vogel (1980), sequential objectives refer to the levels of performance within a specific skill, such as rudimentary level, mature or qualitative level, and the functional (usable) level. An explanation of these levels will be provided in the next section. The above levels will accommodate nearly all students from almost zero competence to the highly skilled individual. If a teacher has students that do not fit into one of these levels, it is suggested that a lower level or assistance level be added that requires the student to perform components of the quality level with physical assistance.

When objectives are stated in measurable terms and are subdivided into sequential levels, criterion-referenced tests can be designed to assess student status to determine their unique instructional needs on each objective. The educator's task becomes one of facilitating student achievement of the next performance level.

Criteria Selection for Performance Standards

Motor skill acquisition is a sequential and complex process during which attention is focused successfully from simple to more complex competency. The most common levels of motor skill acquisition used to describe the degree of competency in a specific skill, discussed in the preceding section, are: 1) rudimentary (nonmature) level; 2) mature or qualitative level; and 3) functional (qualitative and quantitative) level. The rudimentary level represents initial learning of the skill without having all of the required components. An example of a rudimentary overhand throw would be if a student has all the components of the skill except that he steps on the wrong foot when transferring his weight, or the student may lack only consistency to perform the throw three consecutive times with all the required components.

The mature or qualitative level represents a degree of competency exemplary of mastery of all the stated qualitative components of a skill. The only competency lacking is the competency for the components of accuracy and/or distance. An example of a functional overhand throw would be when a student can throw a ball with all the stated qualitative components and hit a specific target, i.e., a four-foot square target from a distance of 40 feet, three consecutive times.

Although highly competent performers demonstrate variations in style, these differences do not violate underlying skill or common components that are requirements for skilled movement. There is general agreement regarding the identification of these common skilled elements among those who are familiar with the biomechanical similarities and differences in skilled performance (Espenschade & Eckert, 1967; Seefeldt, 1976a, 1976b, 1976c; Seefeldt & Haubenstricker, 1974, 1975, 1976a, 1976b; Wickstrom, 1977).

Two common approaches used to determine the performance criteria established in motor skill and physical fitness assessment instruments are 1) factor analysis, and 2) biomechanical analysis.

Several studies have investigated the manner in which scores collected in batteries of motor ability tests tend to cluster into common factors indicative of unique and separate attributes. In a recent factor analysis of motor performance in normal children by Rarick and Dobbins (1975), the scores on 47 tests were analyzed. Six factors accounted for the major portion of the variance in both sexes:

1. Strength-power-body size, combining measures of height, weight, grip, and limb strength.
2. Gross limb coordination, including measures of throwing, running, and crawling.
3. Fine visual-motor coordination.
4. Fat, or dead weight.
5. Balance, including both static and dynamic balance.
6. Leg power and coordination.

Rarick and Dobbins (1972) investigated the factor structure of motor abilities of educable mentally retarded boys and girls in the age ranges 6 to 9 years and 10 to 13 years. Forty-seven tests were administered with the following factors occurring most often:

1. Muscular strength and power.
2. Visual-motor coordination (fine manipulative skill and hand-eye coordination).
3. Gross body coordination.
4. Dead weight or body fat.

They also concluded that the factor structures of normal and educable mentally retarded boys and girls are indeed quite similar.

In a similar study conducted by Rarick and McQuillan (1977), the following common factor structures were identified for trainable mentally retarded children:

1. Fine visual-motor coordination
2. Balance
3. Upper limb-eye coordination
4. Arm strength
5. Spinal flexibility
6. Leg-power coordination

In light of the above studies and results reported by others (Carpenter, 1940, 1941; Cumbee, 1957; Vandenberg, 1964), it appears that the following common factor structures exist across normal, educable and trainable mentally retarded children:

1. Eye-limb coordination
2. Gross body coordination
3. Upper body strength
4. Flexibility
5. Leg-power coordination

Most motor assessment instruments can be classified into two groups: 1) those that assess the quanitative aspects of motor performance; and 2) those that assess the qualitative performance. The vast majority of instruments surveyed fall into the first category. Figure 2 represents the tests reviewed in the literature.

Test	Physical, Motor and/or Fitness Emphasis	Target Population	Developmental or Chronological Age Range	Norm, Content or Criterion-Referenced	Primary Components Tested	Number of Test Items	Tester/Testee Ratio	Items of Equipment Needed	Administration Time (Minutes)	Standardized	Adequate Psychometric Properties	Matched to Instructional Programming
1. Bayley Schedules (The Motor Scale)	Motor	General Population	2 mos- 2.5 yrs.	Norm	Posture, Locomotion, Fine Motor Coordination	81	1/1	11	45	Yes	No	No
2. Denver Developmental Screening Test	Motor	General Population	2 wks- 6 yrs.	Norm	Pre-Locomotion, Locomotion, Balance, Eye-Limb Coordination	31	1/1	7	15	Yes	No	No
3. Early Intervention Developmental Profile	Motor	General Population	0-36 mos.	Norm	Locomotion, Balance, Eye-Limb Coordination, Reflexes	83	1/1	9	20	No	No	Yes
4. Gesell Developmental Schedule	Phys. & Motor	General Population	4 wks- 6 yrs.	Norm	Posture, Reflex, Locomotion, Balance, Eye-Limb Coordination, Fine Motor	155	1/1	22	25	Yes	Concurrent Validity	No
5. Learning Accomplishments Profile	Motor	General Population	1 mo- 6 yrs.	Norm	Locomotion, Balance, Eye-Limb Coordination, Rhythm	64	1/1	16	15	No	No	No
6. OSU - Sigma	Motor	General Population	3 yrs- 14 yrs.	Criterion	Locomotion and Object Control	11	1/1 or more	4	20	Inadequate	Reliability	Yes

Figure 2. Parameters of selected physical, motor and fitness tests for normal, mildly and moderately retarded children.

Test	Physical, Motor and/or Fitness Emphasis	Target Population	Developmental or Chronological Age Range	Norm, Content or Criterion-Referenced	Primary Components Tested	Number of Test Items	Tester/Testee Ratio	Items of Equipment Needed	Administration Time (Minutes)	Standardized	Adequate Psychometric Properties	Matched to Instructional Programming
7. DeOreo Fundamental Skills Inventory	Motor	General Population	3 yrs-5 yrs.	Norm	Locomotion and Object Control	11	1/1	5	20	No	No	No
8. Portage Project Checklist	Phys. & Motor	General Population	0-6 yrs.	Norm	Balance, Posture, Manipulation, Fine Motor, Locomotion, Eye-Hand Coordination, Strength	140	1/1	46	60	No	No	Yes
9. Tarc Assessment	Motor	Mentally Retarded	3 yrs-16 yrs.	Norm	Small Muscle Coordination, Large Muscle Coordination, Pre-Academic Skills	61	1/1	30	Untimed	Yes	No	No
10. AAHPER Special Fitness Test for the Mentally Retarded	Fitness	Mentally Retarded	8 yrs-18 yrs.	Norm	Arm-Shoulder Endurance, Abdominal Endurance, Agility, Leg Power, Speed, Coordination, Cardiorespiratory	7	1/1	7	Untimed	Yes	No	No
11. Lincoln-Oseretsky Motor Development Scale	Motor	General Population	5 yrs-15 yrs.	Norm	Motor Development	36	1/1	15	45	Yes	Reliability	No

Figure 2. (Cont.) Parameters of selected physical, motor and fitness tests for normal, mildly and moderately retarded children.

Quantitative measures of fundamental motor skills for primary and elementary age children were generated by a number of investigators. Taylor (1941) and Latchaw (1954) collected comparative data for activities found in the context of a physical education curriculum for the primary and elementary grades. Carpenter (1942), Glassow and Krause (1957), Govatos (1959), Johnson (1962), and Kane and Meredith (1953) investigated quantitative performance of what they termed "general motor ability." These studies of fundamental motor skills consisted of measuring distances and times with emphasis on age-appropriate behaviors.

Quantitative aspects of motor performance provide little, if any, information required in making programmatic decisions on the rudimentary and qualitative performance levels. They do provide valuable information once a student masters the qualitative level and moves toward the functional level in terms of distance and time expectancies.

In reviewing the available literature, it is apparent that quantitative methods are still used because little has been done to provide a qualitative guide in evaluating most of the motor skills. The factor analysis studies reviewed above used tests of a quantitative nature. If they had used qualitative tests as well, they may have found quite different, or at least more complete, findings.

Qualitative analysis of fundamental motor skills was first identified in detail when Wild (1938) studied the overhand throw of 32 children ranging in age from 24 months to 144 months. Hellebrandt, Lawrence, Glassow, and Carns (1961) studied the broad jump and were able to document the sequential development of that skill as the 47 subjects demonstrated various stages of the skill. Seefeldt and Haubenstricker (1972-1976), Milne (1972), and Wickstrom (1977) have attempted to qualitatively establish common sequences of motor skill development in several fundamental motor skills. According to Herkowitz (1978), a major strength of intraskill sequencing (rudimentary, qualitative, and functional) is its lack of emphasis on age-appropriate behaviors and focus on delivering instruction associated with actual instructional needs.

A review of the literature pertaining to the availability of materials developed with this developmental approach was conducted. Wessel (1976) has developed an objective-based curriculum through a task-analysis procedure in fundamental motor skills, body management skills, physical fitness, and aquatics. Wessel (1979) has recently finalized a leisure skill component of the objective-based system which task-analyzed various leisure skills and games. Wessel (1980) has also completed a component on preschool motor skills that are designed as prerequisite or lead-in skills to the primary motor skill components developed earlier.

The State of Michigan (1979) has recently developed minimal performance objectives in physical education with the intent of assessing the competence of students in grades three, six, and nine. These minimal performance objectives have been generated through task analyzing motor skills and listing both qualitative and quantitative aspects of skilled performance levels.

To determine performance criteria in motor skill and physical fitness assessment, a review of the available literature for established components was conducted. Literature pertaining to the fundamental motor skills and physical fitness skills selected for inclusion in this study has been reviewed in an attempt to identify basic skill criteria or elements of mastery to set standards for performance on both qualitative and quantitative levels. Figures 3 through 14 present a summary of common qualitative components of the mature pattern of the skills selected for inclusion in this test. A more detailed review can be found in Appendix B.

Physical Fitness Parameters and Common Test Items Used in Assessment

According to the factor analysis studies reported above (Rarick & Dobbins, 1972; Rarick & Dobbins, 1975;

	Arms move in opposition elbows bent	Period of nonsupport	90^{o} leg flexion of non support leg	Foot placement near or on line	Heel toe and/or toe-heel foot placement
Cratty (1979)	X				
Espenschade & Eckert (1967)		X			
I CAN (Wessel, 1976)	X	X	X	X	X
Minimal Performance Objectives in Physcial Education (Michigan Department of Education, 1979)	X	X	X	X	X
Seefeldt (1976)	X		X		X
Seefeldt, Reuschlein & Vogel (1972)	X		X		X
Wickstrom (1977)	X	X	X		X

Figure 3. A summary of the qualitative components of a mature run.

	A coordinated lift of the arms	A step forward with bad foot followed by a step with near foot	Brief periods of nonsupport	A steady rhythmical pattern	Trailing foot does not cross in front of lead foot at floor contact
Espenschade & Eckert (1967)	X				
Fuller (1973)				X	
I CAN (Wessel, 1976)	X	X	X	X	X
Minimal Performance Objectives in Physical Education (Michigan Department of Education, 1979)	X	X	X	X	X
Nester (1977)	X				
Sapp (1980)		X		X	X
Sinclair (1971)			X	X	

Figure 4. A summary of the qualitative components of a mature gallop.

	Take off on one foot and land on the same foot	Carriage of non support leg slightly flexed	Synchronized arm swim forward and upward and elbows bent slightly	Upright trunk carriage over the support leg	Foot of non support leg is carried in back of body	Pendular leg swing to aid in force production
Espenschade & Eckert (1967)	X					
Haubenstricker & Seefeldt (1975)		X	X			X
I CAN (Wessel, 1976)		X	X			
Loovis (1975)			X			
Minimal Performance Objectives in Physical Education (Michigan Department of Education, 1979)		X	X	X		
Nester (1977)			X	X	X	

Figure 5. A summary of the qualitative components of a mature hop.

	Take off on one foot and land on the opposite foot	Forward reach with arm opposite the lead foot	An in-flight forward trunk lean of 80^{o} or less	Land on bad foot without losing balance
Broer (1973)	X	X		
Espenschade & Eckert (1967)	X			
Godfrey & Kephart (1969)	X			
I CAN (Wessel, 1976)	X	X	X	X
Latchaw (1954)	X	X		
Minimun Performance Objectives in Physical Education (Michigan Department of Education, 1979 draft)	X	X	X	X
Milne (1972)	X			
Nester (1977)	X	X	X	
Schurr (1967)	X	X		

Figure 6. A summary of the qualitative components of a mature leap.

	Two-foot take-off & two foot landing	Preparatory movement includes $90^{o}(\pm 20^{o})$ flexion of both knees with arms extended behind body	Forceful thrust of both arms & full extension of the legs at take-off	Take off angle at $45^{o}(\pm 5^{o})$	Feet make contact with floor ahead of body mass.	Thighs near parallel to the floor at touch-down; simultaneous forward arm action during landing	Arms extend vigorously forward and upward upon take-off, reaching full extension above the head	Crouching & swinging the arms backward & forward	Knees bent at impact
Cratty (1979)	X								
Espenschade & Eckert (1967)	X								
I CAN (Wessel, 1976)	X	X	X	X	X	X			
Minimal Performance Objectives in Physical Education (Michigan Department of Education, 1979 draft)		X	X	X	X	X	X		
Nester (1977)	X		X				X		
Seefeldt (1976)			X			X	X		
Wickstrom (1977)			X			X		X	X

Figure 7. A summary of the qualitative components of a mature horizontal jump.

	A step-hop pattern on alternate feet	Arms move in opposition to legs and bent slightly	A period on nonsupport with each step-hop	A smooth low flexion on nonsupport leg near surface	A smooth flowing transfer of body weight	Reduced arm action during transfer of weight phase	Foot of supporting leg carried near surface during hopping phase.
Espenschade & Eckert (1967)	X						
Godfrey & Kephart (1969)	X	X					
I CAN (Wessel, 1976)	X	X					
Minimal Performance Objectives in Physical Education (Michigan Department of Education, 1979 draft)	X	X	X				
Nester (1977)	X	X		X	X		
Seefeldt & Haubenstricker (1974)	X				X	X	X
Sinclair (1973)		X					

Figure 8. A summary of the qualitative components of a mature skip.

	Trunk maintained in an upright position	Weight transfer from the trailing foot to the lead foot along a straight line	Body turned sideways to desired direction of travel	Slide to the right & to the left	A long sideways step to the lead foot	A period of non-support as the trail foot is brought forward	Arms aid in balancing
I CAN (Wessel, 1976)	X	X	X	X			
Latchaw (1969)		X			X		
Minimal Performance Objectives in Physical Education (Michigan Department of Education, 1979 draft)	X	X	X		X	X	
Nester (1977)		X			X	X	
Schurr (1967)		X			X		
Sinclair (1971)		X			X		X

Figure 9. A summary of the qualitative components of a mature slide.

	One hand contacts ball	Ball contact on upward portion of the bounce	Contact ball at about hip height	Push ball with fingers of either hand	Flex wrist and extend elbow to impart force to the ball	Ball contacts floor in front of the foot on the side of the bounding arm	Smooth integration of components
Espenschade & Eckert (1967)	X	X					
I CAN (Wessel, 1976)			X	X	X	X	X
Minimal Performance Objectives in Physical Education (Michigan Department of Education, 1979 Draft)			X	X		X	X

Figure 10. A summary of the qualitative components of a mature bounce.

	Elbows bend as ball makes contact with hands	Elbows are to the sides of the body	Hands are cupped with thumbs or little fingers together	Hands in front of body, elbows flexed near sides	Extension of the arms in preparation for ball contact	Contact the ball with hands only	Smooth integration of components
Cratty (1975)	X						
Espenschade & Eckert (1967)		X	X				
I CAN (Wessel, 1976)	X			X	X	X	X
Minimal Performance Objectives in Physical Education (Michigan Department of Education, 1979 draft)	X			X	X	X	X
Seefeldt (1976)				X		X	
Wickstrom (1977)			X	X	X	X	

Figure 11. A summary of the qualitative components of a mature catch.

	Running approach to ball	A single continuous motion	Arm-foot opposition	Full leg backswing with concomitant forward body lean	Follow through of kicking leg	Step forward on the non-kicking leg with foot placement next to the ball	Hip extension and knee flexion	Contact center of ball with toes or instep	Smooth integration of components	Leap just prior to the kick	The trunk is inclined backward prior to and during contact	A hop on support leg just after kick
Cratty (1975)	X	X										
Espenschade & Eckert (1967)		X	X	X	X							
I CAN (Wessel, 1976)			X		X	X	X	X	X			
Minimal Performance Objectives in Physical Education (Michigan Department of Education, 1979 draft)			X		X	X	X	X	X			
Seefeldt & Haubenstricker (1975)							X			X	X	X
Wickstrom (1977)			X			X	X					

Figure 12. A summary of the qualitative components of a mature kick.

	- Step forward with foot opposite the throwing arm	Hip-trunk & shoulder rotation	Shift weight to the forward foot	Almost complete extension of the throwing arm to initiate windup	Side orientation toward direction of throw	Follow through well beyond ball release	Smooth integration of the components	A downward arc of the throwing arm initiates windup	Derotation of the hips, spine & shoulder	Extension of the contralateral at the knee	Ball release
Cratty (1975)	X	X	X								
Espenschade & Ecker (1967)	X	X									
I CAN (Wessel, 1976)	X	X		X	X	X	X				
Minimal Performance Objectives in Physical Education (Michigan Department of Education, 1979 draft)		X	X	X	X	X	X				
Seefeldt & Haubenstricker (1976)	X	X				X		X	X	X	
Wickstrom (1977)	X	X				X		X			X

Figure 13. A summary of the qualitative components of a mature overhand throw.

	Dominant hand gripping bat above nondominant hand	Side orientation toward direction of travel	Bat is held behind dominant shoulder prior to strike	Hip and spine rotation	Weight transfer from back foot to front foot during swing	Follow through well beyond point of contact	Smooth integration of components	Shift of weight to forward foot occurs while bat is still moving back	Bat is kept near the body at the initiation of forward movement	Arm swing around and forward
I CAN (Wessel, 1976)	X	X	X	X	X	X	X			
Minimal Performance Objectives in Physical Education (Michigan Department of Education 1979 draft)	X	X	X	X	X	X	X			
Seefeldt & Haubenstricker (1976)					X			X	X	
Wickstrom (1977)				X	X					X

Figure 14. A summary of the Qualitative Components of a Mature Two-Hand Strike.

Rarick & McQuillan, 1977), the following factor structures relate to the physical fitness area and exist across normal, educable, and trainable mentally retarded children:

1. Upper body strength;
2. Flexibility; and
3. Leg-power coordination.

Recently, a joint committee representing the Physical Fitness, Measurement and Evaluation and Research Councils of the American Alliance for Health, Physical Education, Recreation and Dance (AAHPER Joint Committee, 1977) identified three areas of physiological function that are related to positive health: 1) cardiorespiratory function, 2) body composition, and 3) abdominal and low back musculoskeletal function.

Based on the above, four parameters were selected to represent the common factor structures and health-related fitness areas as defined by the Joint Committee. These four parameters were: 1) arm and shoulder strength, 2) abdominal strength, 3) trunk and leg flexibility, and 4) cardiorespiratory endurance.

Literature pertaining to various test batteries used to measure the four selected parameters was reviewed. Figures 15 through 18 present a summary of the most common test items used to measure these parameters. A more detailed review can be found in Appendix C.

	Sit-ups	Flexed leg sit-ups	Bent leg sit-ups in 30 seconds	Perform continuous bent leg sit-ups for specified minimal performance criteria based on age and sex	One bent leg sit-up	Five quality bent leg sit-ups
AAHPERD Health Related Physical Fitness Test (1979)		X				
AAHPER Special Fitness Test for Mildly Mentally Retarded Persons (AAHPER, 1976)	X					
AAHPER Youth Fitness Test Hunsicker & Reiff, 1976)		X				
Fitness & Work Capacity Testing (Sharkey, 1977)			X			
I CAN (Wessel, 1976)				X		
Kraus-Weber Minimum Muscular Fitness Tests (Kraus & Hirschland, 1954)					X	
Minimal Performance Objectives in Physical Education (Michigan Department of Education, 1979 draft)						X
Motor Fitness Test for the Moderately Mentally Retarded (Johnson & Londeree,) 1976				X		
Physical Fitness for the Mentally Retarded (Hayden, 1964)				X		

Figure 15. A summary of common test items used to measure abdominal strength.

	Flexed arm hang for time	Flexed arm hang (females) pull-ups (males)	Push-ups in 60 seconds	Three quality push-ups	Straight arm hang for time
AAHPER Special Fitness Test for Mildly Mentally Retarded Persons (AAHPER, 1976)	X				
AAHPER Youth Fitness Test (Hunsicker and Reiff, 1976)		X			
Fait Physical Fitness Battery for Mentally Retarded Children (Fait, 1972)	X				
Fitness and Work Capacity Test (Sharkey, 1977)			X		
I CAN (Wessel, 1976)	X				
Minimal Performance Objectives in Physical Education (Michigan Department of Education, 1979 draft)				X	
Motor Fitness for the Moderately Mentally Retarded (Johnson and Londeree, 1976)	X				
Physical Fitness for the Mentally Retarded (Hayden, 1964)					X

Figure 16. A summary of common test items used to measure arm and shoulder strength.

	300 yard run/walk for time	600 yard run/walk for time	1.5 mile run for time	Jog/walk continuously for 5-15 minutes depending on age	Run/walk continuously for 1 mile	300 yard run for time	1 mile run	9 minute run for distance
AAHPERD Health Related Physical Fitness Test (AAHPERD, 1979)							X	X
AAHPER Special Fitness test for Mildly Retarded Persons (AAHPER, 1976)	X							
AAHPER Youth Fitness Test (Hunsicker & Reiff, 1976)		X						
Fait Physical Fitness Battery for Mentally Retarded Children (Fait, 1972)	X							
Fitness and Work Capacity Testing (Sharkey, 1977)			X					
I CAN (Wessel, 1976)				X				
Minimal Performance Objectives in Physical Education (Michigan Department of Education, 1979 draft)					X			
Motor fitness test for the Moderately Mentally Retarded (Johnson & Londeree, 1976)						X		

Figure 17. A summary of common test items used to measure cardiorespiratory endurance.

	Sit and Reach	Sitting, bending, reaching	Standing floor touch
AAHPERD Health Related Physical Fitness Test (1979)	X		
Frostig Movement Skills Test Battery (Orpet, 1972)		X	
I CAN (Wessel, 1976)	X		
Minimal Performance Objectives in Physical Education (Michigan Department of Education, 1979 draft)	X		
Motor Fitness Test for the Moderately Mentally Retarded (Johnson and Londeree, 1976)	X		
Physical Fitness for the Mentally Retarded (Hayden, 1964)			X

Figure 18. A summary of common test items used to measure trunk and leg flexibility.

Summary

The review of the literature was guided by the objectives of this study: 1) select and standardize criterion-referenced test items in the physical education domain based on the definition in Public Law 94-142; and 2) develop a set of norms for intellectually normal, educable mentally impaired and trainable mentally impaired students in the age range of 36 months to 155 months. The three major content areas for this chapter are: 1) criterion-referenced measurement; nondiscriminatory assessment and evaluation; 2) standardization procedures for criterion-referenced tests; and 3) motor skill and physical fitness assessment.

Current attempts to reduce bias in assessment and placement include the design of new testing procedures, the use of adaptive behavior scales, the use of criterion-referenced measures, and the interpretation of assessment results using local or special group norms. A criterion-referenced testing and instructional approach facilitates nondiscriminatory decision-making because the process becomes one of 1) identifying basic skills that all students are expected to master, 2) assessing all students to determine which of these basic skills are present, and 3) designing appropriate instruction so that remaining skills can be learned.

Recent interest in objective-based instructional systems generates a need for CRT's. The implementation of an objective-based instructional system facilitates the use of CRT's to assess the entry status of the students on the objectives, the progress of students on the objectives, and the exit abilities of the students on the objectives. The use of CRT's facilitates a continuous evaluation process.

A review of the literature provides eight guidelines that were followed during the CRT standardization process:

1) Function or purpose of the test.
2) Preparation or selection of objectives to be measured by the test.
3) Test item development.
4) Standardization of directions for administration and scoring of the test.
5) Collection of normative data.
6) Test score interpretation.
7) Reliability of the test.
8) Validity of the test.

The content of the test should include test items that measure the common factor structures of the motor domain and qualitative skill analysis. The importance of qualitative skill analysis is its lack of emphasis on age-appropriate behaviors.

CHAPTER III

METHODS AND PROCEDURES

The objectives of this study were: 1) select and standardize criterion-referenced test items in the physical education domain based on the definition in Public Law 94-142; and 2) develop a set of norms for intellectually normal, educable mentally impaired, and trainable mentally impaired children in the age range of 36 months to 155 months.

Design Overview

The first objective of this study was reached by following a six-step process. First, the developer had to decide what functions the items would serve. Second, objectives had to be developed or selected from previously constructed objective-based programs. Third, criterion-referenced test items had to be designed to measure the objectives. Fourth, directions for administering and scoring each test item had to be developed. Fifth, data had to be collected on the validity of the test items; and last, reliability of the test items had to be evaluated.

Functions of the Test

The following functions will be served by this standardized CRT:

1) Screening for identification of children with motor needs by specifying strengths and weaknesses in comparison to the child's norm group;
2) Provide input to determine eligibility for special education services in physical education; and
3) Instructional assessment to aid in decisions regarding appropriate placement and instructional programming to meet the unique needs of the student in physical education.

The above three purposes were used as a guide in the test-development process. Any decisions that were made during this process were made relative to their impact on meeting these purposes.

Development or Selection of Objectives

If the above functions of the test were to be met, it was necessary to develop or select objectives that were representative of physical education skills most commonly taught in American schools.

Fundamental motor skills and physical fitness skills were selected as being representative of skill areas most commonly taught to students in the age range of three to 12.

These two skill areas also were specifically selected from the definition of physical education in Public Law 94-142. Three criteria were used during the selection of specific skills within each of the two skill areas. These criteria represented another effort to maximize the selection of skills most commonly taught in Michigan schools. It was assumed that Michigan was representative of most states. The criteria used were:

1. The skill had to be listed as a minimal performance objective in physical education in the State of Michigan within the object control, locomotor, or physical fitness skill areas (Michigan Department of Education, 1979 draft);
2. The skill had to be a program objective in the I CAN Objective-Based Instructional System within the object control, locomotor, or physical fitness skill area (Wessel, 1976); and
3. The skill had to be determined as content valid by surveying selected physical education literature.

Figure 19 was used as a checklist in selecting those fundamental motor skills that met the above three criteria. Figure 20 was used as a checklist in selecting physical fitness parameters that met the same criteria.

	Cratty (1976, 1979)	Dauer (1971)	Espenshade and Eckert (1967)	I CAN Instructional System (1976)	Ohio State Sigma (1979)	Seefeldt and Haubenstricker (1972-1976)	State of Michigan Min. Perf. Obj. (1979)	Wickstrom (1977)	Meets Selection Criteria
Climb									
Gallop									
Hop									
Horizontal Jump									
Leap									
Run									
Skip									
Slide									
Walk									
Backhand Strike									
Bounce									
Catch									
Forehand Strike									
Kick									
Overhand Strike									
Overhand Throw									
Punt									
Side-Arm Strike									
Underhand Roll									
Underhand Strike									
Underhand Throw									

Figure 19. Fundamental motor skill selection checklist.

	AAHPER Special Fitness Test for Mildly Mentally Retarded Persons (1976)	Daur (1971)	Fait (1972)	Hunsicker & Reiff (1976)	I CAN (Wessel, 1976)	State of Michigan Minimal Performance Obj. (1979)	Safrit (1981)	Sharkey (1979)	Meet Selection Criteria
Abdominal Strength									
Agility									
Arm & Shoulder Strength									
Balance									
Cardiorespiratory Endurance									
Explosive Muscular Power									
Flexibility									
Relaxation									
Speed									
Weight Maintenance									

Figure 20. Checklist for selecting physical fitness parameters.

Subsequent to specific skill selection, amplified objectives were developed from the program objectives contained in the I CAN Objective-Based Physical Education System (Wessel, 1976) and the Minimal Performance Objectives for Physical Education in the State of Michigan (Michigan Department of Education, 1979 draft). The amplified objective for each specific skill was used as the test item's descriptive scheme to guide in the CRT item construction. A descriptive scheme constitutes the "criterion" to which the test is "referenced." It provides a clear description of the behaviors that the student can or cannot perform. The major purpose of the descriptive scheme is to communicate to the test users what it is that the test is measuring. The amplified objectives can be reviewed in Appendix D.

Criterion-Referenced Test Item Development

The following two characteristics of well-designed CRT items were used as a guide in developing the test items (Popham, 1978b):

1) An established descriptive scheme in the form of an amplified behavioral objective; and
2) A sufficiently limited focus on those behaviors that are judged to be the most relevant and observable in the everyday physical education class.

The major task in item development following the specification of amplified objectives was to determine which behaviors were relevant and observable. This task was accomplished by 1) listing the most common qualitative components of each of the selected fundamental motor skills and physical fitness skills, 2) selecting three content experts, and 3) having each content expert rate each qualitative skill component for a) ease of observability in the physical education setting, and b) its compatibility with the research that describes a quality performance. Those qualitative components that received a negative rating on observability and/or compatibility by two or more raters were deleted. Those components that survived the selection process were placed in the criterion level of each test item. It is important to note that the selection of observable behaviors was not a data-based decision.

Standardization of Directions for Administration and Scoring

During the development of standard directions for administration of each CRT item it was necessary to keep in mind the requirements of Public Law 94-142 with regard to nondiscriminatory evaluation practices. The following four stimulus attributes were used in the directions for administering each test item in an attempt to minimize any discriminatory practices:

1. Preceded by a demonstration and verbal request;
2. Two practice trials to assure that the student understands what to do;
3. One additional demonstration is provided where the student does not know what to do on the first trial;
4. Instructions are provided in the student's native language or mode of communication (e.g., sign language).

Four response alternatives were used in an attempt to include the majority of possible performance levels that might exist when assessing normal, educable mentally impaired and trainable mentally impaired students in the age range of 3 through 12 years. The following response alternatives represent a range of performance from no ability to a quality pattern:

1. Criterion Level (C) - Student completes the item according to all the stated criteria. Any quantitative criteria stating "consecutive trials" require performance of all qualitative criteria the stated number of times.
2. Rudimentary Level (R) - Student responds according to some of the criteria, but not all of the criteria (lacks quantitative or qualitative aspects).

3. Assisted Level (A) - Student needs some type of physical asistance to respond, such as manipulating the student, guiding a student's hand or tapping the student's limb. Through physical assistance the student can perform a minimum of <u>one</u> qualitative criterion.
4. Other Level (O) - Student does not respond, responds inappropriately, resists assistance, or cannot perform a minimum of one qualitative criterion with physical assistance.

Validity of the Test Items

Three content experts were utilized in this study. These experts investigated 1) content validity, 2) descriptive validity, and 3) criterion selection validity. Based on the purpose of the investigation, the criteria for selection of these experts included:

1. A minimum of 18 credits in motor development beyond the master's degree;
2. A minimum of three years experience in teaching physical education skills to children or youth; and
3. A minimum of three years experience in observing and assessing the qualitative motor performance of children or youth.

These criteria were used to help ensure the view of persons knowledgeable in the systematic study of motor development and the practitioner.

Content Validity

Two aspects of content validity were investigated:

1. The degree to which the objectives represent the fundamental motor skill (locomotor and object control skills) and physical fitness skill domains; and
2. The degree to which the objectives represent the following factors (subareas) identified in most motor ability studies of normal, educable mentally retarded and trainable mentally retarded children:
    (a) Balance
    (b) Flexibility
    (c) Limb-eye coordination
    (d) Leg power
    (e) Upper body strength

Content validity was measured by having the three experts read a clear description of each skill selected for the study. They were then given a chart identical to Figure 21. Their task was to independently mark an "X" in the appropriate cell if they felt that the objective (skill) was representative of that particular domain and/or subarea. If they felt the objective was not representative of a specific domain or subarea, they were asked to mark a

Objectives	Locomotor Skill	Object Control Skill	Physical Fit-ness Skill	Balance	Flexibility	Leg Power	Limb-Eye Coordin-ation	Upper Body Strength
Gallop								
Hop								
Horizontal Jump								
Leap								
Run								
Skip								
Slide								
Bounce								
Catch								
Kick								
Overhand Throw								
Sidearm Strike								
Sit-Ups								
Push-Ups								
5-Minute Run/Walk								
Sit and Reach								

Directions: Place an "X" in the cell if you feel that the objective is representative of that specific domain or subarea. Place an "0" in the cell if you feel that the objective is not representative of that specific domain or subarea.

Signature of rater:____________________________Percent of agreement_____

Figure 21. Content Validity Form.

"0" in the space provided. The percentage of agreement between raters was calculated and used as the measure of content validity. A percentage of .90 was used as an indicator of satisfactory content validity (Popham, 1976). Where a .90 percentage agreement was not reached, a discussion session was encouraged. At this time, each rater was asked to provide a rationale for his/her rating. Following the discussion, the content experts were asked to rerate the content validity. A .67 percentage of agreement was used for the rerating of content validity.

Descriptive Validity

Descriptive validity was measured by judging the item-objective congruency. Item-objective congruency was defined as "the ability of a test item to measure a specific objective." The same content experts were first asked to read the "amplified behavioral objectives" that constitute the CRT specifications and then read the test items. Each rater was then requested to judge on an item-by-item basis whether the item was congruent with its objective. The following rating described by Kosecoff et al. (1976), Martuza (1977), and Popham (1978a) was utilized.

+1 - if the item was definitely congruent with its objective;

0 - if the rater was not sure of the congruency; or

-1 - if the item was definitely not congruent with its objective.

Figure 22 represents the item-objective congruency form that was used in the descriptive validity study. Following the independent ratings, the responses were tallied and the index of item-objective congruency was calculated. This computation was the mean score obtained for each item. An item had to receive an index of +1 to be accepted as having descriptive validity. Those items that received less than +1 were redesigned until a +1 was reached.

Criterion or Domain-Selection Validity

Two aspects of criterion validity were measured: 1) whether the components of the criterion level of each test item were compatible with what experts consider a quality pattern, and 2) whether the components of the criterion level could be reliably observed in the physical education class. Each content expert was given a copy of the components selected for the criterion level of each item. The raters were requested to independently rate each item by placing an "X" in the appropriate space if the component was compatible and another "X" in the appropriate space if the component was observable by teachers in the physical education class. If a rater gave a component a "0", meaning it was not compatible or was not observable, the raters were asked to write a rationale for their decision. The written rationales were reviewed to make any potential changes in item content.

Objective	Rating
Run	
Gallop	
Hop	
Skip	
Horizontal Jump	
Slide	
Leap	
Abdominal Strength	
Overhand Trhow	
Trunk and Leg Flexibility	
Catch	
Bounce	
Arm and shoulder strength	
Kick	
Two-hand strike	
Cardiorespiratory Endurance	

Directions: (1) Read the amplified objective
(2) Rate the item-objective congruency using the following:

+1 Item is definitely congruent with objective
0 Unsure of congruency
-1 Item is definitely not congruent with objective

Figure 22. Item objective congruency form.

It was anticipated that minor discrepancies would exist and be tolerated due to the fact that the test developer was looking for a compromise between 1) what content authorities consider the components of a quality pattern, 2) what components could reliably be observed in physical education class, and 3) administrative utility (as few components as necessary). As could be expected, some were willing to compromise more than others. Those components receiving two or more "0's" were deleted or revised based on the rationale provided. Figure 23 represents an example form used in the evaluation of criterion validity.

Reliability

The two reliability measures investigated in this study were: 1) internal consistency and 2) test-retest stability. The internal consistency of the CRT was measured by calculating Cronbach's (1951) alpha coefficient. According to Ebel (1979), a coefficient of .80 or higher is acceptable. The standard used in this study was .85.

The test-retest procedure was used to measure the stability of the decision-making process. An available sample of educable mentally impaired students (N = 7) were administered the test (16 items) on a Friday morning by three independent raters. Two weeks later the same three raters readministered the same test. A Pearson Product

Objective	Components	Compatible	Observable	Rationale
Run	A period when both feet are off the floor			
	Toe-heel or heel-toe foot contact			
	Arms move in opposition to legs, elbows bent			
	Smooth pattern for 50 feet			
Gallop	Brief periods where both feet are off the floor			
	Trailing foot does not cross in front of lead foot at floor contact			
	Smooth, rhythmical pattern			
	5 consecutive gallop strike leading with each foot			
Hop	foot of nonsupport leg remainds behind body			
	Maintain upright body position, elbows bent			
	5 consecutive hops forward on each foot			

Directions: (1) Place an "X" in the appropriate box if the component is compatible with a quality pattern and a "O" if it is not.
(2) Place an "X" in the appropriate box if the component is observable and a "O" if it is not.
(3) Provide a rationale for any components receiving a "O"

Figure 23. An example form used in evaluating criterion validity.

Moment Correlation was calculated on the two groups of test scores with a .90 coefficient used as a standard of acceptable test stability. Failure to get a .90 coefficient due to inadequate variability in test scores would result in the calculation of one of the newer methods of determining CRT reliability. Live performance ratings (as contrasted to videotaped) were used to get a more realistic reliability measure similar to the physical education class.

Collection of Student Performance Data

To meet the functions of the test, namely, 1) screening, 2) determining eligibility for special education services in physical education, and 3) instructional assessment for placement and programming, it was necessary to collect student performance data (norms) to which scores could be referenced.

Training the Testers

Three testers were employed to collect student performance data. Two of the testers were master's degree students in physical education. The other tester was a doctoral student in physical education for the handicapped. The three testers were given a copy of the test and administration manual for their review. The testers were trained through repetition and practice over a one-week period of time by administering each test item to each other and to several available children.

Establishing the Performance Standards

A videotape of a representative sample of students available from the Lansing School District was developed. Students were selected who represent all skill levels on the motor skill test items. The researcher administered standard procedures for each test item while a technician videotaped three trials per test item. The following standard test administration procedures were used:

1. Use equipment specified in the criterion-referenced test description;
2. A verbal request in the child's native language and a demonstration of the skill;
3. Three trials per student;
4. A physical prompt when the student fails to perform at a rudimentary level (R) in the first two trials.

Three expert raters were employed to rate the training videotape. The expert raters were selected on the following criteria:

1. A minimum of two years teaching physical education to children or youth in a school or clinical setting;
2. A minimum of two years experience assessing the qualitative motor performance of children or youth; and

3. A minimum of a master's degree in physical education.

The raters were given an opportunity to study the test items prior to rating the videotape. They were asked to rate the performance of all the children on all the items as a group, rather than an independent rating. They were encouraged to discuss any discrepancies and allowed to replay the tape as many times as needed. The final performance ratings were used as the standard to train the three testers.

Prior to any formal data collection, each tester had to meet a 1.0 level of mastery when compared to the expert ratings (performance standards) on the videotape for each student and item. Subsequent to reading the mandatory level of assessment accuracy, the testers evaluated the motor performance of 279 children.

Sampling Design

A 2x10x3 fixed effects research design was employed in this study. There were 60 possible levels across the tnree independent variables. The sampling technique required that each of the 60 cells have a minimum of three subjects which would result in 180 total. Several schools would not allow students to be sampled from classes but insisted that all students in a given class be used. This resulted in a total sample size of 279.

Subsequent to reaching the mandatory level of assessment accuracy, student performance data were collected on a stratified sample (N = 279) of normal, learning disabled and emotionally impaired students placed in the regular physical education program (N = 117), educable mentally impaired students placed in special classes in the regular neighborhood school (N = 96), and trainable mentally impaired students placed in special classes in intermediate school districts (N = 66).

The age range of the sample was 36 months to 155 months with sample sizes by age including: 3-year-olds (N = 18), 4-year-olds (N = 21), 5-year-olds (N = 29), 6-year-olds (N = 29), 7-year-olds (N = 33), 8-year-olds (N = 39), 9-year-olds (N = 46), 10-year-olds (N = 26), 11-year-olds (N = 21), and 12-year-olds (N = 17). The sample was also stratified by sex: females (N = 134) and males (N = 145). The purpose of stratifying the sample was to assure representation of various groups in at least minimum proportions.

Each tester administered all test items on an addividual basis. The tester was responsible for completing a score sheet for each student with necessary information concerning subject's name, age, sex, classification (normal, educable, trainable), date, test administrator's name, and starting and ending times. The data concerning age and classification were obtained from the student's teacher. Following the completion of daily assessment activity, a code number was substituted for the student's name.

Analysis of the Student Performance Data

The student performance data were initially analyzed by computing a three-way analysis of variance to test for a significant interaction effect. If an interaction effect was significant, the means were plotted to determine if it was ordinal or disordinal. Where disordinal interaction was present, the main effects were tested. If main effects were significant, the Tukey Multiple Range Test was conducted.

If the interaction was not significant, the main effects were evaluated.

The next level of analysis consisted of computing item percentiles (cumulative). Subtotal and total test score cumulative percentiles were then constructed. The final level of analysis was to compute the mean and standard deviation for each item, subtotal, and total test score to aid in the construction of student performance profiles.

Test Score Interpretations

Subsequent to student performance data analysis, guidelines for making special education eligibility and placement decisions were developed. The guidelines were based on the construction of student performance profiles. These profiles were developed by using the mean and standard deviation data available for each item, subtotal score, and total test score by age and classification. Figure 24 represents an example of a student profile.

It would be realistic for a school district to provide special education services in physical education to a child that performs consistently below the mean on fundamental motor skills and/or physical fitness skills. If a student performs two or three standard deviations below the mean in one particular skill area, it might be realistic to place that student in an adapted class or provide more physical education classes until the student reaches an acceptable level. A trainable mentally impaired student that performs as well as normal students should realistically be placed in the regular physical education class. These and other decisions will be aided by consulting the student performance profiles.

Pat - Age 10

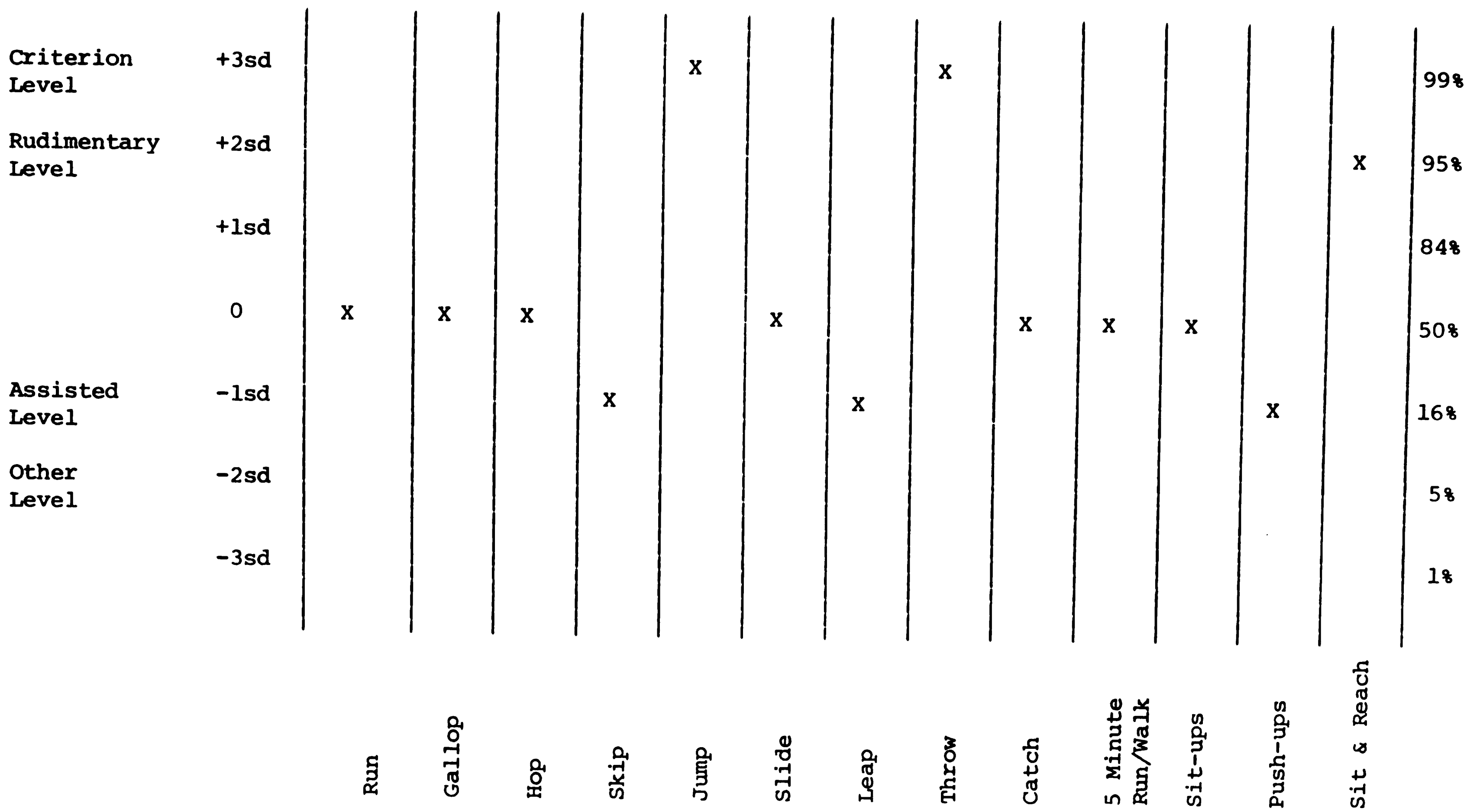


Figure 24 Student Profile

CHAPTER IV

RESULTS AND DISCUSSION

It was the objective of this investigation to: (1) select and standardize criterion-referenced test items in the physical education domain based on the definition in Public Law 94-142; and (2) develop a set of norms for intellectually normal, educable mentally impaired, and trainable mentally impaired children in the age range of 36 months to 155 months. This chapter includes the findings of this study. The chapter is divided into the following two major sections with eight subsections:

1. Selecting and standardizing criterion-referenced test items.
    - 1.1 Determining the functions to be served by the test.
    - 1.2 Selecting objectives to be measured by the test items.
    - 1.3 Development of criterion-referenced test items to measure the selected objectives.
    - 1.4 Standardizing the directions for administration and scoring of each test item.
    - 1.5 Evaluating the validity of the test.

1.6 Evaluating the reliability of the test.

2. Developing a set of norms.

2.1 Collecting student performance data.

2.2 Developing guidelines for test score interpretation.

Identifying and Standardizing Criterion-Referenced Test Items

The Functions of the Test

Three major test functions were identified to guide in the test development process. Any subsequent decisions that were made were based on their impact on these functions.

Selection of Objectives

Fundamental motor skills and physical fitness skills were selected as being representative of the Public Law 94-142 definition of physical education and of skill areas most commonly taught to students in the age range of 3 through 12.

Figure 25 presents the fundamental motor skills and Figure 26 presents the physical fitness skills that were selected to be used in the CRT.

	Cratty (1976,1979)	Dauer (1971)	Espenshade and Eckert (1967)	I CAN Instruction-al System (1976)	Ohio State Sigma (1979)	Seefeldt and Haubenstricker (1972-1976)	State of Michigan Minimal Perf. Obj. (1979)	Wickstrom (1977)	Frequency	Meets Selection Criteria
Climb			X		X				2	
Gallop	X	X	X	X		X	X		6	X
Hop	X	X	X	X	X	X	X		7	X
Horizontal Jump	X	X	X	X	X	X	X	X	8	X
Leap		X		X			X		3	X
Run	X	X	X	X	X	X	X	X	8	X
Skip	X	X	X	X	X	X	X		7	X
Slide		X		X			X		3	X
Walk	X	X	X	X	X	X		X	7	
Backhand Strike				X			X		2	
Bounce			X	X			X		3	X
Catch	X	X	X	X	X	X	X	X	8	X
Forehand Strike				X			X		2	
Kick		X	X	X	X	X	X	X	7	X
Overhand Strike				X						
Overhand Throw	X	X	X	X	X	X	X	X	8	X
Punt						X	X		2	
Side-Arm Strike		X	X	X	X	X	X	X	7	X
Underhand Roll				X			X		2	
Underhand Strike				X					1	
Underhand Throw				X			X		2	

Figure 25. Fundamental motor skill selection checklist.

	AAHPER Special Fitness Test for Mildly Mentally Retarded Persons (1976)	Daur (1971)	Fait (1972)	Hunsicker & Reiff (1976)	I CAN (Wessel, 1976)	State of Michigan Minimal Performance Obj. (1979)	Safrit (1981)	Sharkey (1979)	Frequency	Meet Selection Criteria
Abdominal Strength	X	X		X	X	X	X	X	7	X
Agility	X	X	X	X				X	5	
Arm & Shoulder Strength	X	X	X	X	X	X		X	7	X
Balance		X	X					X	3	
Cardiorespiratory Endurance	X		X	X	X	X	X	X	7	X
Explosive Muscular Power	X	X		X				X	4	
Flexibility		X			X	X	X	X	5	X
Relaxation					X				1	
Speed		X	X	X				X	4	
Weight Maintenance					X				1	

Figure 26. Physical fitness skill selection checklist

<u>Development of the Criterion-Referenced Test Items</u>

The major task in CRT item development was to determine which qualitative behaviors would be used in each test item. Those qualitative components listed in Appendix B were placed on a rating form and evaluated for 1) their observability in a physical education class, and 2) their consistency with what research described as a mature pattern. Three content experts were used in this rating and the subsequent result was a sufficiently limited focus on relevant and observable behaviors. Figure 27 presents a summary of those specific qualitative components selected to represent each skill. Appendix E provides the content expert ratings for each qualitative component.

Subsequent to skill selection and the determination of the specific qualitative behaviors for each selected skill, amplified objectives were developed from the program objectives contained in the I CAN Objective-Based Physical Education System (Wessel, 1976) and the Minimal Performance Objectives for physical education in the State of Michigan (Michigan Department of Education, 1979 draft). The amplified objective for each skill can be found in Appendix D.

It is important to note that the I CAN format of including the qualitative behavior, "smooth, rhythmical pattern," is assumed in each of the amplified objectives.

Run

A period when both feet are off the floor.
Arms move in opposition to legs, elbows bent.
Toe-heel or heel-toe foot contact.

Gallop

Brief periods where both feet are off the floor.
Trailing foot does not cross in front of lead foot at floor contact.

Hop

Carriage of nonsupport leg is slightly bent.
Maintain upright body position, elbows bent.
Arms swing forward and upward.

Skip

A step-hop pattern of alternate feet.
Arms move in opposition to legs and are slightly bent.

Horizontal Jump

Take off and land on two feet.
Arms thrust during take off with full extension of legs.

Slide

Period when both feet are off the floor and remain parallel.
Weight transfer from trail foot to lead foot along a straight line to the side.

Leap

Take off on one foot, balanced landing on the other foot.
Forward reach with arm opposite lead foot.

Overhand Throw

Downward arc of throwing arm to initiate throw.
Hip and trunk rotation.
Weight transfer to foot opposite throwing arm.
Follow through well beyond ball release.

Catch

Hands in front of body, elbows bent.
Extension of arms in preparation for ball contact.
Contact and control ball with hands only.

Stationary Bounce

Contact the ball between thighs and waist.
Push ball with fingers of one hand only.
Maintain a stable stationary position.

Figure 27. A summary of specific qualitative components selected to represent each skill.

Kick

A preliminary forward leap on the non-kicking leg.
Foot placement next to the ball.
A single continuous kicking motion.

Push-Ups

Assume a prone position parallel to floor, toes on floor, hands directly under shoulders.
Keep body parallel to floor while lowering body to 1-3" above floor.
Raise body to starting position.

Two-Hand Strike

Side orientation toward desired direction of travel.
Hip and spine rotation during swing.
Weight transfer onto front foor during swing.

Run/Walk

Run or walk continuously.
Five consecutive minutes.

Sit-Ups

Lie on back, knees bent, feet flat on floor, arms crossed over chest.
Curl up to touch elbows to thighs.
Return to lying position.

Sit and Reach

Assume a sitting position with legs together and knees straight.
Bend and reach forward to feet with one hand on top of the other.
Keep legs straight.
Hold position for 3 seconds.

Figure 27 Continued.

The rationale for this assumption is based on teachers' comments concerning students that learn a skill by learning one qualitative behavior at a time. The final product of their performance sometimes looks mechanical in form because they have not mastered a smooth rhythmical sequence of the components.

A quantitative behavior component was also included in each amplified objective as a standard for mastery. In most objectives, the student was required to perform the specific skill three consecutive trials before mastery was achieved. In certain cases it was more appropriate to use a distance standard such as 50 feet.

Standardizing the Directions for Administration and Scoring of Each Test Item

The most important factor in the development of standard directions was the requirements of Public Law 94-142 concerning nondiscriminatory evaluation practices. Four stimulus attributes were included in the directions for each test item administration in an attempt to maximize any discriminatory practices.

To allow for a range of performance levels that would be expected from normal, educable, and trainable students in the age range of 3 through 12 years, four response alternatives were used in each test item.

A complete test manual with test items, directions for administration, scoring, and interpretation, can be found in Appendix F.

Validity of the Criterion-Referenced Test Items

This study utilized three content experts to investigate (1) content validity, (2) descriptive validity, and (3) criterion-selection validity.

Content Validity. Two aspects of content validity were studied:

1. The degree to which the objectives represent the fundamental motor skill (locomotor and object control skills), and physical fitness skill domains; and
2. The degree to which the objectives represent the following factors (subareas) identified in most motor ability studies of normal, educable mentally retarded and trainable mentally retarded children:
   a. Balance
   b. Flexibility
   c. Limb-eye coordination
   d. Leg power
   e. Upper body strength

The three content experts independently rated all 16 skill objectives relative to the above two aspects of content validity. A 97.6% of agreement was obtained from the three raters across all skills. Table 4.1 provides a summary of ratings (n = 3 raters) for content validity.

Table 4.1
A Summary of Ratings for Content Validity

	DOMAINS						SUBAREAS									
	LOCOMOTOR SKILL		OBJECT CONTROL SKILL		PHYSICAL FITNESS SKILL		BALANCE		FLEXIBILITY		LEG POWER		LIMB-EYE COORDINATION		UPPER BODY STRENGTH	
OBJECTIVES	X[a]	O[b]	X	O	X	O	X	O	X	O	X	O	X	O	X	O
GALLOP	3			3		3		3		3		3		3		3
HOP	3			3		3	3			3	3			3		3
HORIZONTAL JUMP	3			3		3	2	1		3	3			3	1	2
LEAP	3			3		3	3			3	3			3		3
RUN	3			3		3		3		3		3		3		3
SKIP	3			3		3	3			3		3		3		3
SLIDE	3			3		3	1	2		3		3		3		3
BOUNCE		3	3			3		3		3		3	3			3
CATCH		3	3			3		3		3		3	3			3
KICK		3	3			3	3			3	3		3			3
OVERHAND THROW		3	3			3		3		3		3		3	3	
SIDEARM STRIKE		3	3			3		3		3		3	3			3
SIT-UPS		3		3	3			3	3			3		3		3
PUSH-UPS		3		3	3			3		3		3		3	3	
RUN/WALK	3			3	3			3		3		3		3		3
SIT AND REACH		3		3	3			3	3			3		3		3
	8		5		4		5		2		4		4		2	

[a]The Objective is a measure of that domain or subarea.
[b]The Objective is not a measure of that domain or subarea.

<u>Descriptive validity</u>. This form of validity was measured by having three content experts judge the CRT item-objective congruency. The judges were required to rate the ability of the test item to measure its amplified objective. All 16 test items received a +1 rating, meaning that, in the opinion of the content experts, all test items measured the objective that they were designed to measure. Table 4.2 presents a summary of the descriptive validity ratings (n = 3 raters). According to these results, the test items had excellent descriptive validity.

<u>Criterion-Selection Validity</u>. This form of validity was evaluated by having three content experts rate the components selected for each test item on (1) their observability in the physical education class and (2) their consistency with what research describes as a mature pattern. All components used in the 16 test items received at least two "X's" on their observability and consistency (an "X" denoting a positive rating), resulting in acceptable criterion validity. Complete criterion validity ratings can be found in Appendix G.

According to these results, the test has adequate validity.

Table 4.2

A Summary of Descriptive Validity Ratings

Objective	Rater		
	1	2	3
Run	+1	+1	+1
Gallop	+1	+1	+1
Hop	+1	+1	+1
Skip	+1	+1	+1
Horizontal Jump	+1	+1	+1
Slide	+1	+1	+1
Leap	+1	+1	+1
Abdominal Strength	+1	+1	+1
Overhand Throw	+1	+1	+1
Trunk and Leg Flexibility	+1	+1	+1
Catch	+1	+1	+1
Bounce	+1	+1	+1
Arm and Shoulder	+1	+1	+1
Kick	+1	+1	+1
Two-Hand Strike	+1	+1	+1
Cardiorespiratory Endurence	+1	+1	+1

Note: A +1 was used to denote in item that was definitely congruent with its objective.

Reliability of the Test Items

Two aspects of reliability were evaluated in this section of the study. The first measure studied was the internal consistency of the test. This was evaluated by calculating Cronbach's (1951) alpha coefficient for 279 students across all 16 test items. An alpha coefficient of .92 was obtained resulting in acceptance of inter-item homogeneity.

The second aspect of reliability studied was the test-retest stability. Three raters independently assessed an available sample of seven educable mentally impaired students on all 16 test items. This means that a total of 336 data were collected. Exactly two weeks later the same three raters readministered the same 16 test items to the same students. A Pearson product-moment correlation was calculated on the two groups of test scores. A correlation coefficient of .97 was obtained. This high correlation between test and re-test data is not extremely surprising in light of the attempt to validate qualitative behaviors that were observable in the physical education class. These behaviors that were rated "difficult to observe" were deleted from the test items. It is important to note that the student's test scores did not change significantly due to vagueness of test instructions, variations in testing conditions or other irrelevant factors that are common in the field. Most of these conditions can be controlled through the use of videotapes but, to keep the situation as realistic as possible, live performances were analyzed.

Based on the above results, Cohen's kappa statistic was not needed. Mehrens (1980) Ebel (1980) believe that the traditional strategies of using correlation coefficients is fine as long as high coefficients are obtained.

Normative Data

Analysis of the Student Performance Data

Initial analysis of the data consisted of computing a three-way analysis of variance to test for an interaction between sex, age, and student classification. The run test item resulted in a significant (.007) disordinal effect for age by student classification. The gallop test item resulted in a significant (.026) disordinal interaction for age by student classification. The horizontal jump test item resulted in a significant (.003) disordinal interaction for sex by age by student classification and also for age by student classification (.037). The leap test item resulted in a significant (.014) disordinal interaction for sex by student classification. There was no significant interaction for any of the five object control skill test items. The sit and reach test item resulted in a significant disordinal interaction (.002) for sex by age by student classification.

Subsequent fo the initial data analysis, the main effects of sex, age, and student classification were evaluated. Table 4.3 shows that the only two test items that resulted in significant differences across sex were the overhand throw and the two-hand strike. Tables 4.4 through 4.19 show that all 12 fundamental motor skill test items resulted in significant main effects for age (3-12), while none of the physical fitness test items resulted in significant age effects.

Significant main effects resulted for all 16 test items across the three student classifications (normal, educable mentally impaired, and trainable mentally impaired). Tables 4.20 through 4.35 present a summary of the analysis of variance for student classification differences on each test item.

Where significant main effects were realized, the Tukey Multiple Range Test was conducted to pinpoint exactly where the significance was located. Tables 4.36 through 4.63 provide data on the results of the Tukey test for age and student classification differences. In 10 of the test items, neither of the three student

Table 4.3

Mean Standard Deviation and F Ratio
for Sex Differences on Each Test Item

Item	Sex	M	SD	F
Run	M	3.37	.75	.75
	F	3.30	.67	
Gallop	M	3.08	.88	1.59
	F	3.22	.87	
Hop	M	3.10	.96	.01
	F	3.11	1.06	
Skip	M	2.66	1.08	1.58
	F	2.83	1.13	
H. Jump	M	3.24	.85	.02
	F	3.25	.74	
Slide	M	2.92	.89	.18
	F	2.97	.95	
Leap	M	2.31	.98	.01
	F	2.30	1.06	
O. Throw	M	3.19	.75	9.30[a]
	F	2.94	.59	
Catch	M	3.12	.80	4.08
	F	2.93	.79	
Stationary Bounce	M	3.12	.85	1.43
	F	3.00	.80	
Kick	M	3.15	.57	3.95
	F	3.02	.51	
Two-Hand Strike	M	3.27	.65	26.12[a]
	F	2.90	.57	
Sit-ups	M	3.11	.72	.01
	F	3.10	.66	
Sit and Reach	M	3.22	.58	1.20
	F	3.30	.60	
Push-ups	M	2.79	.92	5.03
	F	2.47	.79	
Run/Walk	M	3.69	.76	.07
	F	3.66	.78	

[a]Significant at .01 level.

Table 4.4

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE RUN TEST ITEM

Source	df	MS	F	p
Between Groups	9	3.07	7.33	.001
Within Groups	269	.42		
Total	278			

Table 4.5

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE GALLOP TEST ITEM

Source	df	MS	F	p
Between Groups	9	7.19	13.11	.001
Within Groups	267	.55		
Total	276			

Table 4.6

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE HOP TEST ITEM

Source	df	MS	F	p
Between Groups	9	10.30	14.64	.001
Within Groups	269	.70		
Total	278			

Table 4.7

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE SKIP TEST ITEM

Source	df	MS	F	p
Between Groups	9	9.95	10.71	.001
Within Groups	269	.93		
Total	278			

Table 4.8

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE HORIZONTAL JUMP TEST ITEM

Source	df	MS	F	p
Between Groups	9	3.68	6.88	.001
Within Groups	269	.54		
Total	278			

Table 4.9

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE SLIDE TEST ITEM

Source	df	MS	F	p
Between Groups	9	9.07	16.09	.004
Within Groups	267	.56		
Total	276			

Table 4.10

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE LEAP TEST ITEM

Source	df	MS	F	p
Between Groups	9	10.50	14.67	.01
Within Groups	269	.72		
Total	278			

Table 4.11

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE OVERHAND THROW TEST ITEM

Source	df	MS	F	p
Between Groups	9	2.91	7.55	.01
Within Groups	269	.39		
Total	278			

Table 4.12

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE CATCH TEST ITEM

Source	df	MS	F	p
Between Groups	9	6.16	13.59	.001
Within Groups	268	.45		
Total	277			

Table 4.13

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE STATIONARY BOUNCE TEST ITEM

Source	df	MS	F	p
Between Groups	9	8.17	19.13	.001
Within Groups	268	.43		
Total	277			

Table 4.14

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE KICK TEST ITEM

Source	df	MS	F	p
Between Groups	9	2.16	9.08	.01
Within Groups	267	.24		
Total	276			

Table 4.15

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE TWO-HAND STRIKE TEST ITEM

Source	df	MS	F	p
Between Groups	9	2.77	8.42	.01
Within Groups	267	.33		
Total	276			

Table 4.16

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE SIT-UPS TEST ITEM

Source	df	MS	F	p
Between Groups	4	.64	1.37	.23
Within Groups	142	.47		
Total	146			

Table 4.17

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE SIT AND REACH TEST ITEM

Source	df	MS	F	p
Between Groups	4	.78	2.33	.02
Within Groups	142	.34		
Total	146			

Table 4.18

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE PUSH-UPS TEST ITEM

Source	df	MS	F	p
Between Groups	4	1.25	1.70	.14
Within Groups	142	.73		
Total	146			

Table 4.19

SUMMARY OF ANALYSIS OF VARIANCE FOR AGE
DIFFERENCES ON THE RUN/WALK TEST ITEM

Source	df	MS	F	p
Between Groups	4	1.37	2.46	.04
Within Groups	142	.56		
Total	146			

Table 4.20

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE RUN TEST ITEM

Source	df	MS	F	p
Between Groups	2	13.43	32.67	.001
Within Groups	276	.41		
Total	278			

Table 4.21

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE GALLOP TEST ITEM

Source	df	MS	F	p
Between Groups	2	29.31	52.74	.001
Within Groups	274	.56		
Total	276			

Table 4.22

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE HOP TEST ITEM

Source	df	MS	F	p
Between Groups	2	32.24	40.90	.001
Within Groups	276	.79		
Total	278			

Table 4.23

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE SKIP TEST ITEM

Source	df	MS	F	p
Between Groups	2	57.92	71.50	.001
Within Groups	276	.81		
Total	278			

Table 4.24

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE HORIZONTAL JUMP TEST ITEM

Source	df	MS	F	p
Between Groups	2	15.85	30.12	.001
Within Groups	276	.53		
Total	278			

Table 4.25

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE SLIDE TEST ITEM

Source	df	MS	F	p
Between Groups	2	28.53	44.63	.001
Within Groups	274	.64		
Total	276			

Table 4.26

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE LEAP TEST ITEM

Source	df	MS	F	p
Between Groups	2	32.12	39.77	.001
Within Groups	276	.81		
Total	278			

Table 4.27

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE OVERHAND THROW TEST ITEM

Source	df	MS	F	p
Between Groups	2	8.17	19.88	.001
Within Groups	276	.41		
Total	278			

Table 4.28

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE CATCH TEST ITEM

Source	df	MS	F	p
Between Groups	2	14.72	27.47	.001
Within Groups	275	.54		
Total	277			

Table 4.29

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE STATIONARY BOUNCE TEST ITEM

Source	df	MS	F	p
Between Groups	2	10.69	17.64	.001
Within Groups	275	.61		
Total	277			

Table 4.30

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE KICK TEST ITEM

Source	df	MS	F	p
Between Groups	2	3.14	11.24	.001
Within Groups	274	.28		
Total	276			

Table 4.31

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE TWO-HAND STRIKE TEST ITEM

Source	df	MS	F	p
Between Groups	2	6.30	17.23	.001
Within Groups	274	.37		
Total	276			

Table 4.32

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE SIT-UPS TEST ITEM

Source	df	MS	F	p
Between Groups	2	12.86	42.12	.001
Within Groups	146	.31		
Total	148			

Table 4.33

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE SIT AND REACH TEST ITEM

Source	df	MS	F	p
Between Groups	2	1.96	5.78	.004
Within Groups	146	.34		
Total	148			

Table 4.34

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE PUSH-UPS TEST ITEM

Source	df	MS	F	p
Between Groups	2	15.57	28.49	.001
Within Groups	146	.55		
Total	148			

Table 4.35

SUMMARY OF ANALYSIS OF VARIANCE FOR STUDENT CLASSIFICATION DIFFERENCES ON THE RUN/WALK TEST ITEM

Source	df	MS	F	p
Between Groups	2	10.37	23.25	.001
Within Groups	141	.45		
Total	143			

Table 4.36

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN AGE FOR THE RUN TEST ITEM

Subset	Homogeneous Ages							
1	3	4	5	6				
2	4	5	6	7				
3	5	6	7	8	11	10	9	12

Table 4.37

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN AGE FOR THE GALLOP TEST ITEM

Subset	Homogeneous Ages							
1	3	4						
2	5	6	11	10	7	8	9	12

Table 4.38

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN AGE FOR THE HOP TEST ITEM

Subset	Homogeneous Ages					
1	3	4				
2	4	5	6			
3	5	6	7	11	8	12
4	7	11	8	12	9	10

Table 4.39

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN AGE FOR THE SKIP TEST ITEM

Subset	Homogeneous Ages							
1	3	4						
2	11	5	6	7	8	12	10	9

Table 4.40

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN AGE FOR THE HORIZONTAL JUMP TEST ITEM

Subset	Homogeneous Ages							
1	3	4	6					
2	4	6	11	5	7			
3	6	11	5	7	8	9	12	10

Table 4.41

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN AGE FOR THE SLIDE TEST ITEM

Subset	Homogeneous Ages					
1	3	4				
2	4	5				
3	5	6	7	11	8	
4	6	7	11	8	10	
5	7	11	8	10	9	12

Table 4.42

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN AGE FOR THE LEAP TEST ITEM

Subset	Homogeneous Ages				
1	3	4	5		
2	4	5	6		
3	5	6	7		
4	6	7	8	11	
5	8	11	9	10	12

Table 4.43

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN AGE FOR THE OVERHAND THROW TEST ITEM

Subset	Homogeneous Ages					
1	3	4	5	6		
2	4	5	6	10	7	
3	5	6	10	7	8	11
4	10	7	8	11	9	12

Table 4.44

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN AGE FOR THE CATCH TEST ITEM

Subset	Homogeneous Ages					
1	3	4	5			
2	4	5	6			
3	5	6	7			
4	6	7	8	11		
5	7	8	11	9	10	12

Table 4.45

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN AGE FOR THE STATIONARY BOUNCE TEST ITEM

Subset	Homogeneous Ages				
1	3	4			
2	5	6	7		
3	7	8	11	10	
4	8	11	10	9	12

Table 4.46

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN AGE FOR THE KICK TEST ITEM

Subset	Homogeneous Ages					
1	3	4	5			
2	4	5	6	7		
3	5	6	7	8		
4	6	7	8	9	10	11
5	7	8	9	10	11	12

Table 4.47

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN AGE FOR THE TWO-HAND STRIKE TEST ITEM

Subset	Homogeneous Ages							
1	3	4						
2	4	5	6					
3	5	6	8	7	11	10	9	12

Table 4.48

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATION FOR THE RUN TEST ITEM

Subset	Homogeneous Classifications
1	TMI[a]
2	EMI[b]
3	Normal[c]

[a]Denotes the Trainable Mentally Impaired Classification
[b]Denotes the Educable Mentally Impaired Classification
[c]Denotes the Normal Classification

Table 4.49

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATION FOR THE GALLOP TEST ITEM

Subset	Homogeneous Classifications
1	TMI
2	EMI
3	Normal

Table 4.50

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATION FOR THE HOP TEST ITEM

Subset	Homogeneous Classifications
1	TMI
2	EMI
3	Normal

Table 4.51

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATION FOR THE SKIP TEST ITEM

Subset	Homogeneous Classifications
1	TMI
2	EMI
3	Normal

Table 4.52

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATIONS FOR THE HORIZONTAL JUMP TEST ITEM

Subset	Homogeneous Classifications
1	TMI
2	EMI
3	Normal

Table 4.53

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATIONS FOR THE SLIDE TEST ITEM

Subset	Homogeneous Classifications
1	TMI
2	EMI
3	Normal

Table 4.54

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATIONS FOR THE LEAP TEST ITEM

Subset	Homogeneous Classifications
1	TMI
2	EMI
3	Normal

Table 4.55

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATIONS FOR THE OVERHEAD THROW TEST ITEM

Subset	Homogeneous Classifications	
1	TMI	
2	EMI	Normal

Table 4.56

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATIONS FOR THE CATCH TEST ITEM

Subset	Homogeneous Classifications
1	TMI
2	EMI
3	Normal

Table 4.57

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATIONS
FOR THE STATIONARY BOUNCE TEST ITEM

Subset	Homogeneous Classifications	
1	TMI	
2	EMI	Normal

Table 4.58

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATIONS FOR THE KICK TEST ITEM

Subset	Homogeneous Classifications	
1	TMI	
2	EMI	Normal

Table 4.59

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATIONS FOR THE TWO-HAND STRIKE TEST ITEM

Subset	Homogeneous Classifications	
1	TMI	
2	EMI	Normal

Table 4.60

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATIONS FOR THE SIT-UPS TEST ITEM

Subset	Homogeneous Classifications
1	TMI
2	EMI
3	Normal

Table 4.61

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATIONS FOR THE SIT AND REACH TEST ITEM

Subset	Homogeneous Classifications	
1	EMI	TMI
2	TMI	Normal

Table 4.62

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATIONS FOR THE PUSH-UPS TEST ITEM

Subset	Homogeneous Classifications
1	TMI
2	EMI
3	Normal

Table 4.63

SUMMARY OF TUKEY MULTIPLE RANGE TEST
FOR DIFFERENCES IN STUDENT CLASSIFICATIONS FOR THE RUN/WALK TEST ITEM

Subset	Homogeneous Classifications	
1	TMI	
2	EMI	Normal

classifications resulted in homogeneous student performances. Six of the test items (overhand throw, stationary bounce, kick, two-hand strike, sit and reach, and run/walk) had homogeneous student performance in some combination relative to student classification. Those test items that resulted in homogeneous performance relative to student classification were pooled for the purpose of student profile development.

The next phase of student performance data analysis was the development of cumulative percentiles by age and student classification for each student performance level (criterion, rudimentary, assisted, and other). Tables 4.64 through 4.79 present the cumulative percentiles for each test item. It is important to remember that the physical fitness test items were administered only to students 8 years old and above. Table 4.80 presents a summary of ages at which the criterion level of performance first appeared for each student classification on each test item. Table 4.81 presents a summary of ages at which the criterion level of performance was mastered by 25%, 50%, and 75% or more of the students in each of the student classification samples.

The final phase of student performance data analysis resulted in the formulation of student profiles for ease

Table 4.64

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Run Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	3	11	89	100	100
	4	50	100	100	100
	5	47	100	100	100
	6	63	100	100	100
	7	56	100	100	100
	8	44	100	100	100
	9	94	100	100	100
	10	86	100	100	100
	11	100	100	100	100
	12	83	100	100	100
Educable Mentally Impaired	3	0	67	100	100
	4	0	100	100	100
	5	0	83	100	100
	6	0	86	100	100
	7	41	100	100	100
	8	40	100	100	100
	9	71	100	100	100
	10	67	100	100	100
	11	57	86	86	100
	12	100	100	100	100
Trainable Mentally Impaired	3	0	0	33	100
	4	0	29	58	100
	5	0	50	83	100
	6	0	83	83	100
	7	14	86	100	100
	8	25	100	100	100
	9	25	88	100	100
	10	0	83	100	100
	11	33	100	100	100
	12	33	100	100	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.65

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Gallop Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	3	0	56	89	100
	4	0	75	100	100
	5	53	100	100	100
	6	69	94	100	100
	7	89	100	100	100
	8	88	94	100	100
	9	88	100	100	100
	10	64	100	100	100
	11	80	100	100	100
	12	100	100	100	100
Educable	3	0	17	77	100
	4	0	33	100	100
	5	0	83	100	100
	6	0	86	100	100
	7	44	100	100	100
	8	20	100	100	100
	9	52	100	100	100
	10	60	80	100	100
	11	14	100	100	100
	12	80	80	100	100
Trainable	3	0	0	0	100
	4	0	0	14	100
	5	0	0	33	100
	6	0	50	67	100
	7	0	72	86	100
	8	0	88	100	100
	9	0	100	100	100
	10	0	67	100	100
	11	33	66	89	100
	12	17	100	100	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.66

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Hop Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	3	11	34	67	100
	4	25	63	100	100
	5	53	100	100	100
	6	44	94	94	100
	7	78	100	100	100
	8	88	100	100	100
	9	88	100	100	100
	10	93	100	100	100
	11	80	100	100	100
	12	83	100	100	100
Educable	3	0	0	50	100
	4	0	17	84	100
	5	0	33	83	100
	6	14	57	86	100
	7	37	100	100	100
	8	33	100	100	100
	9	62	100	100	100
	10	83	100	100	100
	11	72	86	100	100
	12	80	100	100	100
Trainable	3	0	0	0	100
	4	0	0	14	100
	5	0	0	33	100
	6	0	0	67	100
	7	0	29	86	100
	8	38	63	100	100
	9	25	63	100	100
	10	33	83	83	100
	11	33	66	77	100
	12	17	83	100	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.67

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Skip Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	3	0	22	44	100
	4	13	38	88	100
	5	41	94	94	100
	6	38	94	100	100
	7	33	77	100	100
	8	38	100	100	100
	9	82	100	100	100
	10	71	100	100	100
	11	40	100	100	100
	12	67	100	100	100
Educable	3	0	0	0	100
	4	0	0	50	100
	5	17	33	83	100
	6	42	71	100	100
	7	29	94	94	100
	8	33	80	100	100
	9	48	96	100	100
	10	33	100	100	100
	11	43	86	86	100
	12	33	67	100	100
Trainable	3	0	0	0	100
	4	0	0	13	100
	5	0	0	17	100
	6	0	17	17	100
	7	0	14	14	100
	8	0	13	63	100
	9	13	38	100	100
	10	0	17	67	100
	11	11	22	44	100
	12	25	38	88	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.68

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Horizontal Jump Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	3	0	67	100	100
	4	38	100	100	100
	5	53	100	100	100
	6	44	100	100	100
	7	67	100	100	100
	8	50	100	100	100
	9	77	100	100	100
	10	86	100	100	100
	11	40	100	100	100
	12	83	100	100	100
Educable	3	0	17	100	100
	4	17	83	100	100
	5	17	100	100	100
	6	42	71	71	100
	7	35	94	94	100
	8	40	87	100	100
	9	52	100	100	100
	10	67	100	100	100
	11	57	86	86	100
	12	80	100	100	100
Trainable	3	0	0	0	100
	4	0	14	43	100
	5	0	33	83	100
	6	17	34	67	100
	7	0	86	100	100
	8	38	100	100	100
	9	13	100	100	100
	10	33	83	100	100
	11	22	78	89	100
	12	17	100	100	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.69

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Slide Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	3	0	33	78	100
	4	25	75	100	100
	5	35	88	100	100
	6	13	100	100	100
	7	44	88	100	100
	8	38	100	100	100
	9	71	100	100	100
	10	64	100	100	100
	11	100	100	100	100
	12	100	100	100	100
Educable	3	0	0	0	100
	4	0	0	83	100
	5	0	17	100	100
	6	0	71	100	100
	7	25	100	100	100
	8	27	94	100	100
	9	48	100	100	100
	10	50	100	100	100
	11	29	86	100	100
	12	100	100	100	100
Trainable	3	0	0	33	100
	4	0	0	0	100
	5	0	0	17	100
	6	0	33	67	100
	7	0	33	67	100
	8	0	63	100	100
	9	0	100	100	100
	10	0	50	100	100
	11	0	67	78	100
	12	0	83	100	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.70

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Leap Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	3	0	0	33	100
	4	0	0	88	100
	5	0	29	76	100
	6	13	69	94	100
	7	11	67	89	100
	8	13	88	100	100
	9	47	100	100	100
	10	50	93	100	100
	11	80	100	100	100
	12	67	100	100	100
Educable	3	0	0	0	100
	4	0	0	0	100
	5	0	0	50	100
	6	0	14	43	100
	7	6	42	71	100
	8	13	33	87	100
	9	14	62	100	100
	10	17	100	100	100
	11	0	57	86	100
	12	20	100	100	100
Trainable	3	0	0	0	100
	4	0	0	0	100
	5	0	0	0	100
	6	0	0	0	100
	7	0	14	14	100
	8	0	25	63	100
	9	0	25	75	100
	10	0	17	67	100
	11	0	11	78	100
	12	0	33	83	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentaty level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.71

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Overhand Throw Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	3	12	56	100	100
	4	25	88	88	100
	5	6	88	100	100
	6	19	94	100	100
	7	56	100	100	100
	8	31	94	100	100
	9	71	100	100	100
	10	21	100	100	100
	11	40	100	100	100
	12	50	100	100	100
Educable	3	0	50	83	100
	4	0	83	100	100
	5	0	67	100	100
	6	0	71	100	100
	7	35	94	100	100
	8	33	93	100	100
	9	38	95	100	100
	10	33	100	100	100
	11	29	100	100	100
	12	100	100	100	100
Trainable	3	0	0	67	100
	4	0	14	100	100
	5	0	33	100	100
	6	0	17	100	100
	7	0	57	86	100
	8	13	88	100	100
	9	0	75	100	100
	10	17	50	100	100
	11	33	89	100	100
	12	17	100	100	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.72

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Catch Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	3	0	33	100	100
	4	0	63	100	100
	5	12	82	100	100
	6	33	80	100	100
	7	67	100	100	100
	8	63	94	100	100
	9	76	94	100	100
	10	71	100	100	100
	11	100	100	100	100
	12	83	100	100	100
Educable	3	0	0	83	100
	4	0	0	100	100
	5	0	50	100	100
	6	0	43	100	100
	7	30	71	100	100
	8	27	100	100	100
	9	43	91	100	100
	10	50	100	100	100
	11	43	100	100	100
	12	80	100	100	100
Trainable	3	0	0	67	100
	4	0	0	100	100
	5	0	0	100	100
	6	0	17	100	100
	7	0	29	86	100
	8	0	100	100	100
	9	0	71	100	100
	10	0	67	100	100
	11	33	55	100	100
	12	17	67	100	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.73

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Stationary Bounce Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	3	0	14	100	100
	4	0	37	100	100
	5	12	77	100	100
	6	25	75	100	100
	7	56	89	100	100
	8	50	94	100	100
	9	76	100	100	100
	10	79	100	100	100
	11	80	100	100	100
	12	100	100	100	100
Educable	3	0	0	83	100
	4	0	0	100	100
	5	20	80	100	100
	6	0	57	100	100
	7	29	76	100	100
	8	33	100	100	100
	9	90	100	100	100
	10	33	100	100	100
	11	43	100	100	100
	12	100	100	100	100
Trainable	3	0	0	33	100
	4	0	0	86	100
	5	0	0	100	100
	6	0	17	100	100
	7	0	43	100	100
	8	12	75	100	100
	9	25	75	100	100
	10	17	83	83	100
	11	33	78	100	100
	12	33	83	100	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.74

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Kick Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	3	0	56	100	100
	4	0	100	100	100
	5	6	94	100	100
	6	6	94	100	100
	7	22	100	100	100
	8	27	100	100	100
	9	47	100	100	100
	10	64	100	100	100
	11	75	100	100	100
	12	50	100	100	100
Educable	3	0	33	100	100
	4	0	67	100	100
	5	0	100	100	100
	6	0	100	100	100
	7	18	100	100	100
	8	33	93	100	100
	9	14	100	100	100
	10	17	100	100	100
	11	29	100	100	100
	12	80	100	100	100
Trainable	3	0	33	100	100
	4	0	57	100	100
	5	0	50	83	100
	6	0	67	100	100
	7	0	57	100	100
	8	13	100	100	100
	9	13	88	100	100
	10	0	83	100	100
	11	22	100	100	100
	12	17	100	100	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level or performance

Table 4.75

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Two-Hand Strike Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	3	0	44	100	100
	4	25	88	100	100
	5	13	93	100	100
	6	25	100	100	100
	7	44	100	100	100
	8	31	94	100	100
	9	53	100	100	100
	10	29	100	100	100
	11	20	100	100	100
	12	50	100	100	100
Educable	3	0	0	100	100
	4	0	33	100	100
	5	17	100	100	100
	6	14	71	100	100
	7	35	94	100	100
	8	33	93	100	100
	9	33	100	100	100
	10	50	100	100	100
	11	57	100	100	100
	12	80	100	100	100
Trainable	3	0	33	100	100
	4	0	14	100	100
	5	0	33	100	100
	6	0	33	100	100
	7	14	71	100	100
	8	0	75	100	100
	9	0	75	100	100
	10	17	83	100	100
	11	13	88	100	100
	12	0	100	100	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.76

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Sit-Ups Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	8	38	100	100	100
	9	53	100	100	100
	10	72	93	100	100
	11	80	100	100	100
	12	67	100	100	100
Educable	8	15	69	100	100
	9	5	84	100	100
	10	20	100	100	100
	11	29	86	100	100
	12	80	100	100	100
Trainable	8	0	38	100	100
	9	10	60	100	100
	10	18	76	100	100
	11	26	91	100	100
	12	51	98	100	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.77

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Sit and Reach Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	8	31	94	100	100
	9	47	100	100	100
	10	50	100	100	100
	11	40	100	100	100
	12	83	100	100	100
Educable	8	20	93	100	100
	9	5	95	100	100
	10	17	83	100	100
	11	43	86	86	100
	12	60	100	100	100
Trainable	8	25	88	100	100
	9	38	100	100	100
	10	33	67	100	100
	11	33	100	100	100
	12	33	100	100	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.78

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Push-Ups Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	8	19	75	100	100
	9	41	88	100	100
	10	14	71	100	100
	11	20	100	100	100
	12	50	83	100	100
Educable	8	0	23	85	100
	9	10	74	90	100
	10	0	67	100	100
	11	14	57	86	100
	12	0	100	100	100
Trainable	8	0	38	75	100
	9	0	25	88	100
	10	0	17	50	100
	11	0	11	56	100
	12	17	33	83	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.79

Cumulative Percentiles by Age and Student Classification for Each Student Performance Level on the Run/Walk Test Item

Classification	Age	C[a]	R[b]	A[c]	O[d]
Normal	8	92	100	100	100
	9	100	100	100	100
	10	100	100	100	100
	11	100	100	100	100
	12	100	100	100	100
Educable	8	85	85	100	100
	9	90	95	100	100
	10	100	100	100	100
	11	86	86	86	100
	12	100	100	100	100
Trainable	8	38	50	100	100
	9	100	100	100	100
	10	50	67	100	100
	11	56	56	100	100
	12	17	17	67	100

[a]Denotes the criterion level of performance

[b]Denotes the rudimentary level of performance

[c]Denotes the assisted level of performance

[d]Denotes the other level of performance

Table 4.80

Age at Which the Criterion Level of Performance First Appeared for Each Student Classification

Test Item	Classification	Age	Test Item	Classification	Age
Run	N[a]	3	Catch	N	5
	E[b]	7		E	7
	T[c]	7		T	11
Gallop	N	5	Stationary Bounce	N	5
	E	7		E	5
	T	11		T	8
Hop	N	3	Kick	N	5
	E	6		E	7
	T	8		T	8
Skip	N	4	Two-Hand Strike	N	4
	E	5		E	5
	T	9		T	7
Horizontal Jump	N	4	Sit-ups[e]	N	8
	E	4		E	8
	T	6		T	9
Slide	N	4	Sit and Reach[e]	N	9
	E	7		E	8
	T	-[d]		T	8
Leap	N	6	Push-ups[e]	N	8
	E	7		E	9
	T	-		T	12
Overhand Throw	N	3	Run/Walk[e]	N	8
	E	7		E	8
	T	8		T	8

[a] Denotes the normal student classification
[b] Denotes the educable mentally impaired student classification
[c] Denotes the trainable mentally impaired student classification
[d] Denotes that the criterion level of performance did not appear in the age range 3-12
[e] This test item was administered to students 8 years old and above

Table 4.81

Age at Which the Criterion Level of Performance was Mastered by 25%, 50% and 75% or More of the Students in Each of the Student Classification Samples

		Age					Age		
Test Item	Classification	25%	50%	75%	Test Item	Classification	25%	50%	75%
Run	N[a]	4	4	9	Catch	N	6	7	9
	E[b]	7	9	12		E	7	10	12
	T[c]	8	-[d]	-		T	11	-	-
Gallop	N	5	5	7	Stationary	N	6	7	9
	E	7	9	12	Bounce	E	7	9	9
	T	11	-	-		T	9	-	-
Hop	N	4	5	7	Kick	N	8	10	11
	E	7	9	10		E	8	12	12
	T	8	-	-		T	-	-	-
Skip	N	5	9	9	Two-Hand	N	4	9	-
	E	6	-	-	Strike	E	7	10	12
	T	12	-	-		T	-	-	-
Horizontal	N	4	5	9	Sit-ups[e]	N	8	9	11
Jump	E	6	9	12		E	11	12	12
	T	8	-	-		T	11	12	-
Slide	N	4	9	11	Sit and Reach[e]	N	8	10	12
	E	7	10	12		E	11	12	-
	T	-	-	-		T	8	-	-
Leap	N	9	10	11	Push-ups[e]	N	9	12	-
	E	-	-	-		E	-	-	-
	T	-	-	-		T	-	-	-
Overhand	N	4	7	-	Run/Walk[e]	N	8	8	8
Throw	E	7	12	12		E	8	8	8
	T	11	-	-		T	8	9	9

[a] Denotes the normal student classification
[b] Denotes the educable mentally impaired student classification
[c] Denotes the trainable mentally impaired student classification
[d] Denotes that the criterion level was not mastered by this percent of the students at any age between 3-12
[e] This test item was administered to students 8 years old and above

Table 4.82

The Mean and Standard Deviation for the Locomotor Skill Test Items by Age and Student Classification

Skill	Classification	3 M	3 SD	4 M	4 SD	5 M	5 SD	6 M	6 SD	7 M	7 SD	8 M	8 SD	9 M	9 SD	10 M	10 SD	11 M	11 SD	12 M	12 SD
Run	Normal	3.0	.50	3.5	.54	3.5	.51	3.6	.50	3.6	.53	3.4	.51	3.9	.24	3.9	.36	3.9	.25	3.8	.41
	Educable	2.7	.52	3.0	.25	2.8	.41	2.7	.76	3.4	.50	3.4	.50	3.7	.46	3.7	.52	3.3	1.11	3.9	.25
	Trainable	1.3	.58	1.9	.89	2.3	.82	2.7	.82	3.0	.58	3.3	.46	3.1	.64	2.8	.41	3.3	.50	3.3	.52
Gallop	Normal	2.4	.73	3.8	.46	3.5	.51	3.6	.62	3.9	.33	3.8	.54	3.9	.33	3.6	.50	3.8	.45	3.9	.33
	Educable	1.8	.75	2.3	.52	2.8	.41	2.9	.38	3.4	.51	3.2	.41	3.5	.51	3.4	.89	3.1	.38	3.6	.89
	Trainable	1.1	.33	1.1	.38	1.3	.52	2.2	.98	2.6	.79	2.9	.35	3.0	.50	2.7	.52	2.9	1.05	3.2	.41
Hop	Normal	2.1	1.05	2.9	.84	3.5	.51	3.3	.79	3.8	.44	3.9	.34	3.9	.33	3.9	.27	3.8	.45	3.8	.41
	Educable	1.5	.55	2.0	.63	2.2	.75	2.6	.98	3.4	.50	3.3	.49	3.6	.50	3.8	.41	3.6	.79	3.8	.45
	Trainable	1.1	.31	1.1	.38	1.3	.52	1.7	.52	2.1	.69	3.0	.93	2.9	.84	3.0	1.10	2.8	1.20	3.0	.63
Skip	Normal	1.7	.87	2.4	.92	3.3	.77	3.3	.60	3.1	.78	3.4	.50	3.8	.39	3.7	.47	3.4	.55	3.7	.52
	Educable	1.1	.33	1.5	.55	2.3	1.03	2.3	1.11	3.2	.73	3.1	.74	3.4	.60	3.3	.52	3.1	1.07	3.0	1.00
	Trainable	1.1	.33	1.1	.35	1.2	.41	1.3	.82	1.3	.76	1.8	.71	2.5	.76	1.8	.75	1.8	1.09	2.5	1.07
H. Jump	Normal	2.7	.50	3.4	.52	3.5	.51	3.4	.51	3.7	.50	3.5	.52	3.8	.44	3.9	.36	3.4	.55	3.8	.41
	Educable	2.2	.41	3.0	.60	3.2	.41	2.9	1.35	3.2	.75	3.3	.70	3.5	.51	3.7	.52	3.3	1.11	3.8	.45
	Trainable	1.1	.35	1.6	.79	2.2	.75	2.2	1.17	2.9	.38	3.4	.52	3.1	.35	3.2	.75	2.9	.93	3.2	.41
Slide	Normal	2.1	.78	3.0	.76	3.2	.66	3.1	.34	3.3	.71	3.4	.50	3.7	.47	3.6	.50	3.9	.33	4.0	0
	Educable	1.1	.33	1.8	.41	2.2	.41	2.7	.49	3.3	.45	3.2	.56	3.5	.51	3.5	.55	3.1	.69	4.0	0
	Trainable	1.3	.58	1.3	.25	1.2	.41	2.0	.89	2.0	.89	2.6	.52	3.0	.40	2.5	.55	2.4	.88	2.8	.41
Leap	Normal	1.3	.50	1.9	.35	2.1	.75	2.8	.75	2.7	.87	3.0	.52	3.5	.51	3.4	.65	3.8	.45	3.7	.52
	Educable	1.0	.00	1.0	.00	1.5	.55	1.6	.79	2.2	.95	2.3	.90	2.8	.70	3.2	.41	3.4	.79	3.2	.45
	Trainable	1.0	.00	1.0	.00	1.0	.00	1.0	.00	1.3	.76	1.9	.84	2.0	.76	1.8	.75	1.9	.60	2.2	.75
Locomotor Sub-Total	Normal	2.2	.57	2.8	.44	3.2	.43	3.3	.37	3.4	.46	3.5	.21	3.8	.16	3.7	.27	3.7	.27	3.8	.27
	Educable	1.7	.31	2.1	.21	2.4	.37	2.5	.66	3.1	.44	3.1	.43	3.4	.33	3.5	.17	3.1	.76	3.7	.34
	Trainable	1.1	.08	1.2	.26	1.5	.31	1.9	.50	2.2	.44	2.7	.39	2.8	.30	2.5	.54	2.4	.81	2.8	.36

of teacher interpretation. The mean and standard deviation for each test item, subtotal score (locomotor skills, object control skills, and physical fitness skills), and total test score was calculated by age and student classification. Table 4.82 provides the mean and standard deviation for all seven locomotor skill test items and subtotal by age and student classification. Table 4.83 presents the mean and standard deviation for each of the five object control skill test items along with the object control skill subtotal and total fundamental motor skill test score (combines both locomotor and object control skills). Table 4.84 presents the mean and standard deviation for each of the four physical fitness test items by student classification and age along with the physical fitness subtotal and total test score (this combines all 16 test items for ages 8-12). This phase of analysis required that a numerical value be assigned to each performance level (criterion = 4, rudimentary = 3, assisted = 2, and other = 1).

In reviewing Tables 4.82, 4.83, and 4.84, a definite trend was noticed. All locomotor skill subtotals resulted in the normal student classification scoring the highest, followed by educables and then trainables for all ages. This same trend existed for the object control

Table 4.83

The Mean and Standard Deviation for the Object Control Skill Test Items and Total Fundamental Motor Skill Test Score by Age and Student Classification

Skill	Classification	Age 3 M	3 SD	4 M	4 SD	5 M	5 SD	6 M	6 SD	7 M	7 SD	8 M	8 SD	9 M	9 SD	10 M	10 SD	11 M	11 SD	12 M	12 SD
Overhand Throw	Normal	2.7	.71	3.0	.93	2.9	.43	2.9	.25	3.6	.53	3.3	.58	3.7	.47	3.2	.43	3.4	.55	3.5	.55
	Educable	2.3	.82	2.8	.41	2.7	.52	2.7	.49	3.3	.59	3.3	.59	3.3	.58	3.3	.52	3.3	.49	3.9	.25
	Trainable	1.7	.58	2.1	.38	2.3	.52	2.2	.41	2.4	.79	3.0	.54	2.8	.46	2.7	.82	3.2	.67	3.2	.41
Catch	Normal	2.3	.50	2.6	.52	2.9	.56	3.1	.74	3.7	.50	3.6	.63	3.7	.59	3.7	.47	3.9	.34	3.8	.41
	Educable	1.8	.41	2.0	.50	2.5	.55	2.4	.53	3.0	.79	3.3	.46	3.3	.66	3.5	.55	3.4	.53	3.8	.45
	Trainable	1.7	.58	3.0	.33	2.0	.51	2.2	.41	2.1	.69	3.0	.50	2.7	.49	2.7	.52	2.9	.93	2.8	.75
Stationary Bounce	Normal	2.1	.38	2.4	.52	2.9	.60	3.0	.73	3.4	.73	3.4	.63	3.8	.44	3.8	.43	3.8	.45	3.9	.45
	Educable	1.8	.41	2.0	.27	3.0	.71	2.6	.54	3.1	.75	3.3	.49	3.6	.67	3.3	.52	3.4	.54	3.9	.54
	Trainable	1.3	.58	1.9	.38	2.0	.31	2.17	.41	2.4	.54	2.9	.64	3.0	.76	2.8	.98	3.1	.78	3.2	.75
Kick	Normal	2.6	.53	3.0	.40	3.0	.35	3.0	.37	2.8	.44	3.3	.46	3.5	.51	3.6	.50	3.8	.50	3.5	.55
	Educable	2.3	.52	2.7	.52	3.0	.57	3.0	.51	3.2	.39	3.3	.59	3.1	.36	3.2	.41	3.4	.54	3.8	.45
	Trainable	2.3	.58	2.6	.54	2.3	.82	2.7	.52	2.6	.54	3.1	.35	3.0	.54	2.8	.41	3.2	.44	3.2	.41
Two-Hand Strike	Normal	2.4	.53	3.1	.64	3.1	.46	3.3	.45	3.4	.53	3.3	.58	3.5	.52	3.3	.47	3.2	.45	3.5	.55
	Educable	2.0	.30	2.3	.52	3.2	.41	2.9	.69	3.3	.59	3.3	.59	3.3	.48	3.5	.55	3.6	.53	3.8	.45
	Trainable	2.3	.58	2.1	.38	2.3	.52	2.3	.52	2.9	.69	2.8	.46	2.8	.46	3.0	.63	3.0	.53	3.0	.51
Object Control Sub-total	Normal	2.5	.44	2.8	.28	3.0	.36	3.1	.36	3.5	.34	3.4	.46	3.6	.36	3.5	.24	3.6	.24	3.7	.32
	Educable	2.1	.38	2.4	.24	2.9	.24	2.7	.30	3.2	.48	3.3	.50	3.4	.42	3.4	.26	3.3	.36	3.9	.18
	Trainable	1.9	.30	2.1	.16	2.2	.28	2.3	.30	2.5	.48	3.0	.26	2.9	.40	2.8	.50	3.1	.48	3.1	.38
Fundamental Motor Skill Test Score	Normal	2.3	.49	2.8	.26	3.1	.25	3.2	.22	3.4	.27	3.4	.13	3.7	.09	3.6	.18	3.7	.16	3.8	.16
	Educable	1.7	.18	2.2	.13	2.6	.22	2.6	.38	3.1	.26	3.2	.25	3.4	.19	3.4	.10	3.2	.44	3.8	.20
	Trainable	1.4	.05	1.6	.15	1.8	.18	2.0	.29	2.3	.26	2.8	.23	2.8	.16	2.7	.32	2.7	.48	2.9	.21

Table 4.84

The Mean and Standard Deviation for the Physical Fitness Skill Test Items and the Total Test Score by Age and Student Classification

Skill	Classification	Age									
		8		9		10		11		12	
		M	SD	M	SD	M	SD	M	SD	M	SD
Sit-ups	Normal	3.4	.50	3.5	.51	3.6	.63	3.8	.45	3.7	.52
	Educable	2.9	.69	2.9	.46	3.2	.45	3.1	.69	3.8	.45
	Trainable	2.4	.52	2.4	.52	2.7	.52	2.4	.53	2.8	.41
Sit & Reach	Normal	3.3	.58	3.5	.51	3.5	.52	2.4	.53	2.8	.41
	Educable	3.1	.52	3.0	.32	3.0	.63	3.1	1.10	3.6	.55
	Trainable	3.1	.64	3.4	.52	3.0	.89	3.3	.50	3.3	.52
Push-ups	Normal	2.9	.68	3.3	.69	2.9	.67	3.2	.45	3.3	.82
	Educable	2.1	.64	2.7	.81	2.7	.52	2.6	.98	3.0	.51
	Trainable	2.1	.83	2.1	.64	1.7	.82	1.7	.71	2.3	1.00
Run/Walk	Normal	3.9	.28	3.9	.38	3.9	.24	3.9	.25	3.9	.22
	Educable	3.7	.75	3.8	.50	3.9	.34	3.6	1.10	3.9	.21
	Trainable	2.9	.99	3.9	.34	3.2	.98	3.1	1.10	2.0	1.10
Physical Fitness (Sub-total)	Normal	3.3	.38	3.6	.25	3.5	.35	3.6	.23	3.7	.40
	Educable	2.9	.45	3.2	.38	3.2	.30	3.0	.83	3.6	.23
	Trainable	2.6	.40	3.0	.25	2.6	.63	2.7	.53	2.6	.58
Total Test Score	Normal	3.4	.24	3.7	.18	3.6	.19	3.7	.19	3.8	.28
	Educable	3.1	.34	3.3	.27	3.4	.19	3.2	.54	3.7	.21
	Trainable	2.8	.31	2.9	.26	2.6	.51	2.7	.58	2.9	.44

subtotal and total fundamental motor skill score, with only one slight exception. The 12-year-old educables scored slightly above normal students in the object control skill subtotal. This may have been due in part to several of the normal 12-year-olds' attitudes toward participating in "kids" activities. It would be normal for most 12-year-olds to be more interested in sports skills than in fundamental skills. Physical fitness subtotals along with the total test score resulted in the same trend.

Subsequent to the calculation of means and standard deviations for all test items, subscores and total test score, student profiles were developed by age and student classification (Appendix F). The four possible performance levels for each item (criterion = C, rudimentary = R, assisted = A, and other = O) were plotted on the student profile according to where that performance fell relative to the normative sample. If a 3-year-old normal student performed a rudimentary run, it would be an average performance, but if the same student performed the run at the criterion level, it would result in a performance 2 standard deviations above the mean relative to that child's peer group. It was decided that 4 standard deviations above or below the mean was adequate for describing a student's performance on each test item. It was also decided that 3 standard deviations above or below the mean for subtotal scores and total test scores was adequate description of student performance. The subtotals and total test score were reported in numerical form on each student profile.

On several test items all of the students in a particular student classification group performed at the same level. For example, all trainable mentally impaired 3-year-olds performed at the other level on the leap test item. This resulted in no standard deviation and was reflected as such on the profile. This was also common on several of the fundamental motor skills for normal 12-year-olds. It was felt that this represented the actual performance level expected by 12-year-olds rather than lack of sample size, although given a large sample, it would be most likely that a small standard deviation would develop.

Guidelines for Student Profile Interpretation

The final subsection of this chapter deals with the presentation of several possible guidelines to aid teachers and parents in interpreting the student profiles and making appropriate deciions relative to the functions of this test.

The first function of the test can be achieved by observing which skills the student performs poorly. If a specific student performs below the mean on a test item when compared to the performance of his/her peer group, that skill could be listed as a unique need of that student. By plotting the student's performance across all 16 test items on the student profile, the teacher or parent can quickly see which skills should be listed as weaknesses. If the student performs below the mean on several of the test items within a given skill area (locomotor, object control, or physical fitness), or if the student's subtotal test score is below the mean for that peer group, then the entire skill area can be listed as a general weakness. It will be more precise to pinpoint each skill that is in need of improvement rather than the general skill area. General skill area weaknesses can be used to develop program goal statement for each student.

It would be useful and appropriate to include this test in the initial screening inventory that all kindergarten children are administered prior to enrollment in public school. This way, the teacher and parent can identify specific or general motor skill needs prior to the first day of class.

The second test function can be attained by establishing local criteria for eligibility of special education services in the physical education domain. This function may be best served by observing the subtotal and total test score for each student. For example, if a student's performance falls at 2 standard deviations or more below the mean when compared to normal students, a school system may decide that the student should be eligible for special education in physical education. It is the reaponsibility of the local school district to set the criteria, whether it's 1, 2, or 3 standard deviations below the mean. Most qualified physical education teachers that individualize their instruction should be able to meet the needs of students that perform a 1 standard deviation below the mean without specially designed instruction, given that all other learning and social variables are within a normal or average range.

These local criteria should be applied to the locomotor skill, object control skill, and physical fitness skill subtorals along with the total test score for each student. By plotting the student's performance levels on the student profile for each of the 16 test items and converting the performance levels to a numerical scale (4-1 for criterion level through the other level respectively) and adding up the subtotals and total test score, a teacher should be able to make more precise decisions concerning each student's eligibility. If a student meets the local eligibility criteria, specially designed instruction should be developed and placed in his/her IEP.

Placement decisions can be made by comparing a student's performance with the performance of other students in a particular class or setting. If any student, regardless of classification, performs around the mean when compared to a given group, then it appears that the proper placement would be with that group or class.

A school district may want to establish criteria for alternative placements. The following criteria should be viewed as possible examples:

1. Regular class placement - Students that perform around the mean or above on the test items under consideration.
2. Remedial or adapted class placement - Students that exhibit several minor skill weaknesses. This might be for short periods of time until several skill easknesses are remediated or in addition to the regular class placement. This is appropriate for students that perform at 1 or 2 standard deviations below the mean on a few skills but at the mean or above on most of them. If a student is placed in this setting, specially designed instruction should be developed to remediate skill weaknesses.
3. Special class placement - Students that perform well below the mean on most or all of the skills. This placement is appropriate for students with gross deficiencies in physical education. The placement may be on a full- or part-time basis depending on the student's performance. If a student is placed in this class, specially designed instruction should be mandatory.

If a regular class placement is being considered for educable or trainable mentally impaired students, then it is necessary to compare their performance on each test item with that of the normal students. This can be accomplished by plotting the student's scores on the normal student classification portion of the student profile. A different colored pencil can be used for this purpose rather than a new student profile sheet.

Instructional programming decisions can be made if a local school district's physical education curriculum matches the objectives measured by this test. A precise instructional prescription can be developed based on the identification of skills that the student has not mastered. If a student has mastered the mature run, hop, and horizontal jump, it would be more appropriate to prescribe instruction on the skills that have not been mastered, rather than duplicating instruction on already learned skills. This last function does not require the use of the student profiles. Simply by looking at the raw test scores on each item, a teacher can determine the student's present level of performance and select those skills that have not been mastered at the criterion level. This same strategy can be employed for making instructional grouping decisions within a class. By grouping those students together that have the same instructional needs, a teacher can plan more appropriate learning activities.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The objectives of this study were: 1) select and standardize criterion-referenced test items in the physical education domain based on the definition in Public Law 94-142; and 2) develop a set of norms for intellectually normal, educable mentally impaired, and trainable mentally impaired students in the age range of 36 months to 155 months.

Public Law 94-142 mandates assessment for three purposes: to determine eligibility for special education services, instructional planning (including the IEP development), and evaluating the effectiveness of the instructional plan. The mandate requires that state and local educational agencies shall ensure that assessment and evaluation procedures be nondiscriminatory. Assessment must be viewed clearly and simply as the process of collecting data for the purpose of making nondiscriminatory decisions about students using the guidelines provided by PL 94-142.

Review of current assessment instruments available to physical educators as to their conformance with the requirements of PL 94-142 shows tremendous need for revision and development of valid and reliable instrumentation in the physical education domain.

A CRT approach should facilitate nondiscriminatory decision-making because the process becomes one of 1) identifying basic skills that all students are expected to master, 2) assessing all students to determine which of these basic skills are present, and 3) designing appropriate instruction so the remaining skills can be learned.

The CRT instrument validated in this study was designed to assess part of the physical education domain, not the entire area specified by PL 94-142. Using I CAN and the Michigan Performance Objectives, locomotor skills, object control skills, and physical fitness skills were identified as representative skill areas for selection of objectives from physical education program goals. The CRT items were constructed to assess the following four levels of performance: criterion (a mature or quality pattern), rudimentary, assisted, and other.

A review of the literature on CRT construction provided eight guidelines that were followed during the CRT standardization process:

1) Determine the function(s) or purpose of the test.
2) Prepare or select objectives to be measured by the test items.
3) Develop the test items.
4) Standardize the directions for administration and scoring of each test item.
5) Evaluate the validity of the test.
6) Evaluate the reliability of the test.
7) Collect normative data.
8) Develop guidelines for test score interpretation.

An effort was made during the standardization process to include test items designed to measure the common factor structures of the motor domain established in studies by Rarick and Dobbins (1972 and 1975) and Rarick and McQuillan (1977).

The criterion level of performance for each test item was established by content experts. The most common qualitative behaviors for each test item were rated on their observability and consistency with research findings on quality performances.

Content validity, descriptive validity and criterion-selection validity were established along with the internal consistency of the test items and the test-retest reliability.

Student performance data were collected on a sample of 279 students in the age range of 3 through 12. The student performance data were analyzed to determine significant interaction effects between sex, age, and student classification. Main effects were also evaluated. All test items that resulted in significant interaction were of a disordinal nature.

The overhand throw and the two-hand strike resulted in significant differences by sex. All 12 fundamental motor skills resulted in significant differences across age (3-12). Significant student performance differences existed in all 16 test items across the three student classifications (normal, educable, and trainable). Where significant performance differences were detected, the Tukey Multiple Range Test was conducted to pinpoint the exact location of the difference to aid in student profile development. Where no difference was detected, the performance data were pooled and are reflected on the student profiles.

Cumulative percentiles by age and student classification were computed for each student performance level (criterion, rudimentary, assisted, and other). A Table was constructed to provide a summary of ages at which the criterion level of performance first appeared for each student classification on each test item. Another Table was designed to present a summary of ages at which the criterion level of performance was mastered by 25%, 50% and 75% or more of the students in each of the student classification samples.

The mean and standard deviation for each test item, subtotal score (locomotor, object control, and physical fitness skills), and total test score were calculated by age and student classification. These data were used to develop student profiles by age and student classification for ease of interpreting the test results. The four possible performance levels for each test item were plotted on the student profile according to where that performance fell relative to the normative sample.

Conclusions

Within the limitations of this study, the following conclusions were drawn:

1. The CRT standardized in this study has acceptable content validity, descriptive validity, and criterion selection validity.
2. The CRT standardized in this study has excellent internal consistency and test-retest reliability.
3. Males perform significantly better than females on the overhand throw and two-hand strike across all ages and student classifications in the total sample.
4. A significant difference in student performance existed on all 12 fundamental motor skill test items across age for the total sample.
5. A significant difference in student performance existed on all 16 test items across the three student classifications.

Recommendations

The following suggestions are recommended for future research concerned with the development and standardization of criterion-referenced tests in physical education:

1. Add the functional performance level to each skill test item. This would result in five performance levels. The functional level would become the criterion performance while the present criterion level would be called the mature level.

2. Develop and standardize CRT items to measure the other content areas within the PL 94-142 definition of physical education.

3. Collect normative data on other student classification groups such as the hearing impaired and visually impaired. Determine if there is a significant difference between these student classifications.

4. Develop a group test score sheet that is capable of facilitating continuous assessment and evaluation on the same student. This can be achieved by allotting space on the test sheet where several dates and scores can be placed.

5. Evaluate the CRT's sensitivity to appropriately planned instruction.

6. Using a regression strategy, determine which test items account for the most variance within each subtotal and total test score. This might help to determine which test items are the most predictive of the locomotor and object control subtotals, the fundamental motor skill subtotal, and the physical fitness skill subtotal.

7. Determine the observability of the qualitative components placed in the criterion level of performance by collecting assessment data across several content experts. Those components that can be reliably assessed by a majority of the experts should be placed in the criterion level.

8. Establish the degree to which teachers can be trained to use the CRT instrument.

9. Evaluate the CRT's ability to aid parents, teachers, and administrators in making special education eligibility decisions in the physical education content area.

10. Evaluate the CRT's ability to aid in making placement and instructional programming decisions.

APPENDICES

APPENDIX A

PUBLIC LAW 94-142 SUGGESTIONS CONCERNING ASSESSMENT AND EVALUATION PROCEDURES WITH APPLICATIONS TO PHYSICAL EDUCATION

PUBLIC LAW 94-142 SUGGESTIONS CONCERNING ASSESSMENT AND EVALUATION PROCEDURES WITH APPLICATIONS TO PHYSICAL EDUCATION

PL 94-142 suggests that assessment and evaluation procedures:	Application to physical education
Be tailored to assess specific program areas of educational need for instruction in physical education and not merely those designed to provide a single general test quotient.	Use tests designed to assess a student's physical performance in program areas for instruction for this purpose and not as a means of determining basic motor development or general ability levels.
	Make generalizations cautiously; because a student does well or not on a specific test item should not automatically be interpreted as strength or weakness in more than that particular skill or pattern.
	Avoid a single general physical or motor quotient in the same way and for the same reasons a single general intelligence quotient is prohibited.
	Base decisions about physical and motor needs of the student on appropriate assessment data about physical and motor levels of functioning on the specific objectives of the program not on categorical information about handicapping conditions.

Be selected and administered so as to best ensure that when a test is given to a student with impaired sensory, manual, or speaking skills, test results accurately reflect the student's achievement level or whatever other factors the test purports to measure rather than reflecting the student's impairment.	Use test items that measure an individual's present level of performance in physical and motor fitness, fundamental motor skills or patterns or skills in aquatics, dance, individual and group games and sports, and lifetime sports.
Be provided and administered in the student's native tongue or other mode of communication unless it is clearly not feasible to do so.	Give necessary instruction for assessment in physical education program areas in the student's native language. Use accurate demonstrations of the skill being assessed. Include other modes of communication where necessary, i.e., sign language, braille, etc.
Be administered by trained personnel in conformance with instructions provided by the producer of the test.	Use physical education instructions or others specially trained to assess the physical education content area. Select or develop physical education tests that have clear standardized procedures and directions for administration and scoring.
Be provided with student data in all areas related to the suspected disability including, where appropriate, health, vision, social and emotional status, general intelligence, academic performance, communication, and physical and medical status.	Assess the physical education skills of each student suspected of having a disability. Develop special education eligibility guidelines in physical education. Screen students for gross deficiencies in physical or motor development.

Not be a single procedure as a single criterion for determining an appropriate educational program for a child.	Use a variety of sources and procedures to determine specific physical and motor needs of each student, such as formal and informal tests, qualitative and quantitative measures, objective and subjective data, and observational and anecdotal input.
Be conducted by a multidisciplinary team or group of persons including at least one teacher or other specialist with knowledge of the area of suspected disability.	

APPENDIX B

A REVIEW OF THE QUALITATIVE COMPONENTS MOST COMMONLY USED TO DESCRIBE THE FUNDAMENTAL MOTOR SKILLS SELECTED FOR INCLUSION IN THE CRITERION-REFERENCED TEST

A REVIEW OF THE QUALITATIVE COMPONENTS MOST COMMONLY USED TO DESCRIBE THE FUNDAMENTAL MOTOR SKILLS SELECTED FOR INCLUSION IN THE CRITERION-REFERENCED TEST

Qualitative Components of a Mature Run

Literature Source	Components
Cratty (1979)	- Reciprocal arm action
Espenschade and Eckert (1967)	- Short period where body is propelled off the ground
I CAN (Wessel, 1976)	- Brief periods of nonsupport - 90° leg flexion of the non-support leg - Foot placement approximately two inches either side of a one-inch line - Heel-toe (moderate speed) or toe-heel-toe (fast speed) foot placement - Arms in opposition to legs, elbows bent - Smooth integration of the above components
Minimal Performance Objectives in Physical Education in the State of Michigan (Department of Education, 1979 Draft)	- Knee of nonsupporting leg bent at least 90° - Consistent periods of non-support - Foot placement near or on line - Heel-toe and/or toe-heel-toe foot contact - Arms in direct opposition to legs, with elbows bent - Smooth integration of the above
Seefeldt, Reuschlein, and Vogel (1972)	- Heel-toe-foot contact at modest speed - Arm opposition to leg action - Knee of nonsupport leg may flex until it is nearly in contact with buttocks

Wickstrom (1977)	- Period of no support from either leg - Flexion of the knee of the swinging leg being brought closer to the buttocks - Arms in synchronized opposition to the leg action - Toe-heel-foot contact

<u>Qualitative Components of a Mature Gallop</u>

Literature Source	Components
Espenschade and Eckert (1967)	- A coordinated lift of the arms to assist with the balancing
Fuller (1973)	- Maintain a steady rhythmical pattern
I CAN (Wessel, 1976)	- A step forward with the lead foot is followed by a step with the rear foot to a position slightly behind the heel of the lead foot - Brief periods of nonsupport as the rear foot approaches the lead foot, during which time the weight is shifted - Slight flexion of the rear knee during nonsupport phase so that rear foot does not drag on the ground - Arms flexed at sides at about waist level, lifting in a coordinated movement in front of the body during weight transfer - Smooth integration of the above
Minimal Performance Objectives in Physical Education in the State of Michigan (1979)	- A step forward with the lead foot followed by a step with the rear foot to a position slightly behind the heel of the lead foot

	- A period of nonsupport as the rear foot approaches lead foot during which time the weight is shifted to trail foot - A lifting of flexed arms at waist level coordinated with the shift of weight - A smooth integration of the above
Nester (1977)	- Able to use either foot as lead leg - A coordinated lift of the arms
Sapp (1980)	- A smooth, rhythmical pattern - Trail leg may cross in front of or move adjacent to the lead leg during the airborne phase but is placed adjacent to or behind the lead leg at contact
Sinclair (1971)	- Rhythmical and steady action

Qualitative Components of a Mature Hop

Literature Source	Components
Espenschade and Eckert (1967)	- Take off on one foot and land on same foot
Haubenstricker and Seefeldt (1975)	- Knee of nonsupport leg is flexed at 90^{o} or less - Pendular leg swing to aid in force production - Arms carried close to the sides of the body with elbow flexion at 90^{o}
I CAN (Wessel, 1976)	- Carriage of nonsupport leg near the mid-line of body and slightly flexed at the knee so that nonsupport foot is not more than six inches from the floor

	- Lift of both arms in front of body coordinated with the push-off phase of the hopping action; elbows bent approximately 90° - Smooth integration of the above
Loovis (1975)	- Arms swing forward and upward synchronously
Minimal Performace Objectives in Physical Education in the State of Michigan (1979)	- Upright trunk carriage over the support foot - Carriage of nonsupport leg near the midline and slightly flexed at the knee - Lift up both arms in front of body coordinated with the take-off phase of the hop, elbows bent about 90° - Smooth integration of the above
Nester (1977)	- Upright trunk carriage over support leg - Carrying nonsupport leg near the midline of body - Arms lift in coordination with the push-off phase of hop - Foot of nonsupport leg is carried in back of body

Qualitative Components of a Mature Leap

Literature Source	Components
Broer (1973)	- Take off on one foot and land on opposite foot with smooth transfer of weight - Forward reach
Espenschade and Eckert (1967)	- Take off on one foot and land on the alternate foot
Godfrey and Kephart (1969)	- Take off on one foot and land on opposite foot

I CAN (Wessel, 1976)	- Take off on one foot, land on the other, with a gliding motion in the air and a longer period of nonsupport than in running - An in-flight forward trunk lean of 80° or less - Forward reach with arm opposite the lead foot - Land on the lead foot without losing balance - Smooth integration of the above components
Latchaw (1954)	- Take off on one foot and land on opposite foot - Forward reach in a forceful manner
Minimal Performance Objectives in Physical Education in the State of Michigan (1979)	- Take off on one foot, land on the other - An in-flight forward trunk lean of 80° or less from side view - Forward reach with arm opposite the lead foot - Land on the lead foot without losing balance - Smooth integration of the above
Milne (1972)	- Take off on one foot and land on opposite foot
Nester (1977)	- Push off with one foot and land on opposite foot - Forward and upward reach of arm opposite lead leg - More than 75° forward body lean
Schurr (1967)	- Take off on one foot and land on opposite foot - Forward reach

Qualitative Components of a Mature Horizontal Jump

Literature Source	Components
Cratty (1979)	- Two-foot takeoff and two-foot landing
Espenschade and Eckert (1967)	- Two-foot takeoff and two-foot landing
I CAN (Wessel, 1976)	- Preparatory movement includes 90° ($\pm$ 20°) flexion of both knees with arms extended behind the body - Forceful thrust of both arms and full extension of the legs at takeoff - Take off and land on both feet - Takeoff angle at 45°($\pm$ 5°) - Feet make contact with floor ahead of body mass - Thighs near parallel to the floor at touch-down; simultaneous forward arm action during landing - Smooth integration of the above
Minimal Performance Objectives in Physical Education in the State of Michigan (1979)	- Preparatory movement includes 90°($\pm$ 20°) flexion of both knees with arms extended behind the body - A forceful forward-upward thrust of both arms and full extension of the legs at takeoff. - Takeoff angle at 45°($\pm$ 5°) - Simultaneous foot contact at landing well ahead of body's center of mass - Thighs near parallel to the floor at touch-down - Arms extended forward during landing - Smooth integration of the above

Nester (1977)	- A 90° forward and upward thrust of both arms - Take off with both legs simultaneously - Approximate full extension of both legs in forward and upward direction - Both feet landing simultaneously - Simultaneous vigorous forward thrust of arms as feet make contact
Seefeldt (1976)	- Arms extend vigorously forward and upward upon takeoff, reaching full extension above the head - Full extension of hips and knees - Arms are brought downward and legs are thrust forward until the thigh is parallel to the surface
Wickstrom (1977)	- Crouching and swinging the arms backward and upward - Arms swing forward and upward and body extends - Lower legs flex - Hips flex, arms and trunk move forward and downward - Lower legs extend just prior to landing - Knees bend at impact

Qualitative Components of a Mature Skip

Literature Source	Components
Espenschade and Eckert (1967)	- A step-hop pattern
Godfrey and Kephart (1969)	- Step-hop action - Arm and leg opposition

I CAN (Wessel, 1976)	- Repeat the step and hop on alternate feet - Arms move in opposition to legs at about waist level - A smooth integration of the above components
Minimal Performance Objectives in Physical Education in the State of Michigan (1979)	- Repeat the step and hop on alternate feet - Arms move in a lifting action in opposition to legs - A period of nonsupport with each step-hop - Smooth integration of the above
Nester (1977)	- A rhythmical repetition of the step-hop on alternate feet along a straight line - A smooth low flexion on non-support leg near surface - Arms alternately moving in opposition to leg at about waist level - A smooth-flowing transfer of body weight
Seefeldt and Haubenstricker (1974)	- A step-hop pattern - Rhythmical transfer of weight - Reduced arm action during transfer of weight phase - Foot of supporting leg carried near surface during hopping phase
Sinclair (1973)	- Arm and leg opposition

Qualitative Components of a Mature Slide

Literature Source	Components
I CAN (Wessel, 1976)	- Trunk maintained in an upright position - Weight transfer from the following foot to lead foot along a straight line - Body turned sideways to desired direction of travel

	- Slide to the right and to the left - Smooth integration of the above components
Latchaw (1969)	- A long sideways step of the lead foot - Smooth-flowing weight transfer along a straight line
Minimal Performance Objectives in Physical Education in the State of Michigan (1979)	- Trunk maintained in an upright position - A step sideways followed by a slide of the training leg to a landing position within six inches of the original lead foot position - A period of nonsupport as the trail foot is brought forward - Weight transfer from the following foot to the lead foot - Body turned sideways to the desired direction of travel - Smooth integration of the above
Nester (1977)	- A step sideways with the lead foot - Body faces forward - A slide of following foot next to lead foot - A short nonsupport period - A smooth-flowing weight transfer from following leg to lead foot - Can slide in a straight line
Schurr (1967)	- A long sideways step of the lead foot - Smooth-flowing weight transfer along a straight line
Sinclair (1971)	- A long sideways step of the lead foot - Arms aid in balancing - Sideways straight line motion

Qualitative Components of a Mature Bounce

Literature Source	Components
Espenschade and Eckert (1967)	- One hand contacts ball - Ball contact on upward portion of the bounce
I CAN (Wessel, 1976)	- Contact ball at hip height - Push ball with fingers of either hand - Flex wrist and extend elbow to impart force to ball - Ball contacts floor in front of the foot on the side of the bounding arm - Smooth integration of the above components
Minimal Performance Objectives in Physical Education in the State of Michigan (1979)	- Contact ball at hip height - Contact ball with fingers of either hand - Ball contacts floor in front of (or slightly outside of) the foot on the side of the bouncing arm - Smooth integration of the above

Qualitative Components of a Mature Catch

Literature Source	Components
Cratty (1975)	- Elbows bend as ball makes contact with hands
Espenschade and Eckert (1967)	- Elbows are to the sides of body - Hands are cupped with thumbs or little fingers together
I CAN (Wessel, 1976)	- Hands in front of body, elbows flexed near sides - Extension of the arms in preparation for ball contact - Contact ball with hands only - Elbows bend as arms absorb the force of the ball - Smooth integration of the above components

Minimal Performance Objectives in Physical Education in the State of Michigan (1979)	- Preparatory positioning with hands in front of body, elbows flexed and near the sides - Extension of arms in preparation for ball contact - Contact ball with hands only - Elbows bend as arms absorb the force of the ball (hands retract at least six inches) - Smooth integration of the above
Seefeldt (1976)	- Preparation phase where elbows are flexed and arms are ahead of frontal plane - Ball is caught and controlled by hands only
Wickstrom (1977)	- Move hands into position for catching - Arms raise in front of body - Hands are cupped with fingers oriented towards the ball - Hands grasp and control ball - Hands give upon contact with ball

Qualitative Components of a Mature Kick

Literature Source	Components
Cratty (1975)	- Running approach to ball - A continuous single motion of approach and kick
Espenschade and Eckert (1967)	- Arm-foot opposition - Full leg backswing with a concomitant forward body lean - Follow-through of kicking leg

I CAN (Wessel, 1976)	- Step forward on the non-kicking leg with foot placement next to ball - Hip extension and knee flexion (at least 120^{o}) during preliminary kicking motion - Contact center of ball with toes or instep - Forward swing of arm opposite kicking leg - Follow-through of kicking foot in an upward motion - Smooth integration of the above components
Minimal Performance Objectives in Physical Education in the State of Michigan (1979)	- A preliminary forward step on nonkicking leg with foot landing next to the ball - Hip extension and knee flexion (of at least 120^{o}) during preliminary kicking motion - Contact ball with toes or instep - Forward swing of the arm opposite kicking leg - Follow-through to a foot position well beyond and above the point of contact - Smooth integration of the above
Seefeldt and Haubenstricker (1975)	- The distance just prior to the kick is covered by a leap - The knee of the kicking leg is slightly flexed just prior to kicking - The trunk is inclined backward prior to and during contact - The momentum of the kick is dissipated by hopping on the support leg

Wickstrom (1977)	- A preliminary forward step on the support leg - A forward swing of the kicking leg with simultaneous flexion at the hip and knee - Vigorous extension by the lower part of kicking leg - A forward swing of opposite arm in reaction to the action of the kicking leg

<u>Qualitative Components of a Mature Overhand Throw</u>

Literature Source	Components
Cratty (1975)	- Step forward with foot opposite to the throwing arm - Hip-trunk and shoulder rotation - Shift weight to the forward foot
Espenschade and Eckert (1967)	- Trunk rotation - Weight transfer to foot opposite throwing arm
I CAN (Wessel, 1976)	- Almost complete extension of the throwing arm to initiate windup - Side orientation toward direction of throw - Weight transfer to foot opposite throwing arm - Hip and spine rotation in preparation for and during the throwing action - Follow-through well beyond ball release and toward the desired direction of travel - Smooth integration of the above components

Minimal Performance Objectives in Physical Education in the State of Michigan (1979)	- Side orientation with weight on rear leg to initiate the throw - Near complete extension of the throwing arm to initiate the throw - Weight transfer to the foot opposite the throwing arm - Marked hip and spine rotation during throw - A follow-through well beyond the ball release and in line with target - Smooth integration of the above
Seefeldt and Haubenstricker (1976)	- A downward arc of the throwing arm initiates the windup - Rotation of the hip and spine - The leg opposite throwing arm strides forward - Derotation of the hips, spine and shoulder - Extension of contralateral leg at the knee - Follow-through toward side opposite throwing arm
Wickstrom (1977)	- The body pivots to the side of throwing arm with weight on same foot and throwing arm swings backward and upward - The opposite foot strides forward in the intended direction of throw - Hips, spine and shoulder rotation - The upper arm is rotated medially and then the forearm is extended - Ball release - Follow-through diagonally downward across body toward the stable forward foot

Qualitative Components of a Mature Two-Hand Strike

Literature Source	Components
I CAN (Wessel, 1976)	- Dominant hand gripping bat above nondominant hand - Side orientation toward direction of travel - Bat is held behind dominant shoulder prior to strike - Hip and spine rotation during swing and follow-through - Weight transfer from back foot to front foot during swing - Follow through well beyond point of contact - Smooth integration of the above components
Minimal Performance Objectives in Physical Education in the State of Michigan (1979)	- Dominant hand gripping ball (palm up) above nondominant hand (palm down) - Side orientation (non-dominant side toward direction of travel) - Bat held behind dominant shoulder prior to strike - Hip and spine rotation during swing and follow-through - Weight transfer from back foot to front foot during swing - Follow through well beyond point of contact - Smooth integration of the above
Seefeldt and Haubenstricker (1976)	- Transfer of weight is in a contralateral pattern - Shift of weight to forward foot occurs while bat is still moving backward - Bat is kept near body at the initiation of forward movement - Weight is on forward foot at ball contact

Wickstrom (1977)	- Body weight is shifted in direction of intended hit while shoulders and arms are coiled in the opposite direction - Hips and spine are rotated in rapid succession in the same direction as the weight shift - Arms swing around and forward

APPENDIX C

THE TEST ITEMS COMMONLY USED TO ASSESS THE SELECTED PHYSICAL FITNESS PARAMETERS TO BE USED IN THE CRITERION-REFERENCED TEST

THE TEST ITEMS COMMONLY USED TO ASSESS THE SELECTED PHYSICAL FITNESS PARAMETERS TO BE USED IN THE CRITERION-REFERENCED TEST

Literature pertaining to the physical fitness skills selected for inclusion in this study has been reviewed in an attempt to identify various test items that are appropriate for assessing the following four common physical fitness parameters. The four parameters appear to be representative of the physical fitness domain. Various tests will be identified along with the test item corresponding to each parameter.

Measures of Cardiorespiratory Endurance

Test Source	Test Item
AAHPER Special Fitness Test for Mildly Mentally Retarded Persons (AAHPER, 1976)	300 yd. run-walk for time
AAHPER Youth Fitness Test (Hunsicker & Reiff, 1976)	600 yd. run-walk for time
Fait Physical Fitness Battery for Mentally Retarded Children (Fait, 1972)	300 yd. run-walk
Fitness and Work Capacity Testing (Sharkey, 1977)	1.5-mile run for time
I CAN (Wessel, 1976)	Jog/walk continuously for 5-15 minutes depending on age

Test Source	Test Item
Minimal Performance Objectives in Physical Education in the State of Michigan (1979)	- Run-walk continuously until designated distance is covered - Maintain a steady pace throughout the run - At least one mile distance
Motor Fitness Test for the Moderately Mentally Retarded (Johnson and Londeree, 1976)	- 300 yd. run-walk for time - 300 yd. run for time
Measures of Abdominal Strength	
AAHPER Special Fitness Test for Mildly Mentally Retarded Persons (AAHPER, 1976)	Sit-ups
AAHPER Youth Fitness Test (Hunsicker and Reiff, 1976)	Flexed leg sit-ups
Fitness and Work Capacity Testing (Sharkey, 1977)	Bent let sit-ups in 30 seconds
I CAN (Wessel, 1976)	Perform continuous bent leg sit-ups for specified minimal performance criteria (from 5-31 depending on age and sex)
Kraus-Weber Minimum Muscular Fitness Tests (Kraus and Hirschland, 1954)	One bent leg sit-up
Minimal Performance Objectives in Physical Education in the State of Michigan (1979)	A Quality Bent Leg Sit-up - Student lies on floor in supine position with knees bent approximately 90^{o}, feet flat on floor and together, hands clasped behind head - Initiates curl-up by tucking chin and lifting trunk - Completes curl-up by touching the elbows to the knees

Test Source	Test Item
	- Lowers upper body in a controlled movement (shoulder blades return to surface) - Five consecutive times
Motor Fitness Test for the Moderately Mentally Retarded (Johnson and Londeree, 1976)	Bent leg sit-ups in 30 seconds
Physical Fitness for the Mentally Retarded (Hayden, 1964)	Sit-ups in 30 seconds

Measures of Trunk and Leg Flexibility

Frostig Movement Skills Test Battery (Orpet, 1972)	Sitting, bending, reaching
I CAN (Wessel, 1976)	Perform a sit and reach for 3 seconds
Minimal Performance Objectives in Physical Education in the State of Michigan (1979)	A Quality Sit and Reach - Sit on floor and place legs such that the heels are 5-7 inches apart and feet in contact with a vertical surface - Knees maintain contact with the floor (legs straight) - Place one hand on top of the other, lean forward, leaning as far past the feet as possible - Maintain the flexed position for at least 3 seconds
Motor Fitness Test for the Moderately Mentally Retarded (Johnson and Londeree, 1976)	Sitting bob and reach

Test Source	Test Item
Physical Fitness for the Mentally Retarded (Hayden, 1964)	Standing floor touch

Measures of Arm and Shoulder Strength

Test Source	Test Item
AAHPER Special Fitness Test for Mildly Mentally Retarded Persons (AAHPER, 1976)	Flexed-arm hang for time
AAHPER Youth Fitness Test (Hunsicker and Reiff, 1976)	Flexed-arm hang (females), pull-ups (males)
Fait Physical Fitness Battery for Mentally Retarded Children (Fait, 1972)	Bent-arm hang for time
Fitness and Work Capacity Testing (Sharkey, 1977)	Push-ups in 60 seconds
I CAN (Wessel, 1976)	To perform a flexed-arm hang
Minimal Performance Objectives in Physical Education in the State of Michigan (1979)	A Quality Push-up - Assume a prone position parallel to floor, hands directly under shoulder, toes on floor (K-3, knees on floor) - Body lowered until elbows flexed to 90^{o} or less - At least 5 seconds - Three consecutive push-ups
Motor Fitness Test for the Moderately Mentally Retarded (Johnson and Londeree, 1976)	Flexed-arm hang for time
Physical Fitness for the Mentally Retarded (Hayden, 1964)	Straight-arm hang for time

APPENDIX D

LIST OF AMPLIFIED OBJECTIVES

LIST OF AMPLIFIED OBJECTIVES

1. Given a demonstration, verbal request, 50 feet of clear space, and three trials, the student can perform a quality run three consecutive times in the following manner:
   - A period when both feet are off the floor
   - Toe-heel or heel-toe foot contact (not flat-footed)
   - Arms move in opposition to legs, elbows bent
   - A smooth pattern for 50 feet with the above components.
2. Given a demonstration, verbal request, 30 feet of clear space, and three trials, the student can perform a quality gallop three consecutive strides leading with each foot in the following manner:
   - Brief period when both feet are off the floor
   - Training foot does not cross in front of lead foot at floor contact
   - Arms swing forward and upward.
3. Given a demonstration, a verbal request, a minimum of 10 feet of clear space, and three trials, the student can perform a quality hop three consecutive times forward on each foot in the following manner:

- Carriage of nonsupport leg is slightly bent
- Maintain upright body position, elbows bent
- Arms swing forward and upward.

4. Given a demonstration, a verbal request, 30 feet of clear space, and three trials, the student can perform a quality skip for three consecutive skipping cycles in the following manner:
   - Repeat the step and hop on alternate feet in a rhythmical pattern
   - Arms move in opposition to legs and are slightly bent.
5. Given a demonstration, a verbal request, 10 feet of clear space, and three trials, the student can perform a quality horizontal jump three consecutive times in the following manner:
   - Two-foot takeoff and a two-foot landing
   - Arm thrust during takeoff with full extension of legs
   - Jump 2/3 of standing height or more.
6. Given a demonstration, a verbal request, 30 feet of clear space, and three trials, the student can perform a quality slide three consecutive times to each side in the following manner:

- Period where both feet are off the ground and remain parallel
- Weight transfer from trail foot to lead foot along a straight line to the side.

7. Given a demonstration, a verbal request, 50 feet of clear space, and three trials, the student can perform a quality leap three consecutive times in the following manner:
   - Takeoff on one foot and a balanced landing on the other foot
   - Forward reach with arm opposite lead foot.
8. Given a demonstration, a verbal request, an 8-10-inch playground ball, and three trials, the student can perform a quality stationary ball bounce for three consecutive bounces in the following manner:
   - Contact the ball between thighs and waist
   - Push ball with fingers of one hand only
   - Maintain a stable, stationary position.
9. Given a demonstration, a verbal request, a 6-inch playground ball tossed underhand to chest height from a distance of 15 feet, and three trials, the student can perform a quality catch three consecutive times in the following manner:
   - Hands in front of the body, elbows bent
   - Extension of arms in preparation for ball contact
   - Contact and control ball with hands only.

10. Given a demonstration, a verbal request, an 8-10-inch playground ball, and three trials, the student can perform a quality kick three consecutive times in the following manner:
    - A preliminary forward leap on the nonkicking leg with foot placement next to the ball
    - A continuous kicking motion
    - Contact ball so it travels forward at least 30 feet.
11. Given a demonstration, a verbal request, a 3-4-inch ball, and three trials, the student can perform a quality overhand throw three consecutive times in the following manner:
    - Downward arc of throwing arm to initiate overhand throw
    - Hip/trunk rotation
    - Weight transfer to foot opposite the throwing arm
    - Follow through well beyond ball release
    - Ball travels forward 30 feet or more.
12. Given a demonstration, a verbal request, a plastic bat, a light-weight, 6-inch ball suspended at waist height, and three trials, the student can perform a quality two-hand, side-arm strike three consecutive times in the following manner:

- Side orientation toward desired direction of travel
- Hip and spine rotation during swing
- Transfer weight onto front foot during swing.

13. Given a demonstration, a verbal request, and three trials, the student can perform a quality bent-leg sit-up three consecutive times in the following manner:
    - Lie on back, knees bent, feet flat on floor and arms crossed over chest
    - Curl up to touch elbows to thighs
    - Return to lying position.
14. Given a demonstration, a verbal request, a flat surface, and three trials, the student can perform a quality push-up three consecutive times in the following manner:
    - Assume prone position (belly down) parallel to floor, toes on floor, hands directly under shoulders
    - Keep body parallel to floor while lowering body until it is 1-3 inches above floor
    - Raise body to starting position.

15. Given a demonstration, a verbal request, a flat vertical surface, and three trials, the student can perform a quality sit and reach three consecutive times in the following manner:
    - Assume a sitting position with legs together and knees straight
    - Bend and reach forward to feet with one hand on top of the other
    - Keep legs straight
    - Hold for three seconds.
16. Given a demonstration, a verbal request and a minimum of 30-by-50 feet of clear space, the student will perform a continuous run/walk for five consecutive minutes.

APPENDIX E

CONTENT EXPERT RATINGS OF THE MOST COMMON QUALITATIVE COMPONENTS OF EACH SKILL SELECTED FOR THE CRT EXPRESSED IN CUMULATIVE FREQUENCY

SKILL	QUALITATIVE COMPONENTS	OBSERV-ABILITY X[a]	OBSERV-ABILITY O[b]	CONSISTENT W/RESEARCH X[a]	CONSISTENT W/RESEARCH O[b]
RUN	.A period when both feet are off the floor	3	0	3	0
	.Arms move in opposition to legs, elbows bent	3	0	3	0
	.90° leg flexion of non support leg	1	2	2	1
	.Toe-heel or heel toe foot contact	3	0	3	0
GALLOP	.Brief periods of non-support	3	0	3	0
	.A coordinated lift of the arms	1	2	3	0
	.Trailing foot does not cross in front of lead foot at floor contact	3	0	3	0
	.A steady rhythmical pattern	3	0	3	0
HOP	.Carriage of nonsupport leg is slightly flexed	3	0	3	0
	.Synchronized arm swing forward and upward and elbows bent slightly	3	0	3	0
	.Upright trunk carriage over the support leg	3	0	3	0
SKIP	.A step-hop pattern on alternate feet	3	0	3	0
	.Arms move in opposition to legs and slightly bent	3	0	3	0
HORIZONTAL JUMP	.Two foot take off and two foot landing	3	0	3	0
	.Forceful trust of both arms and full extention of legs at take off	2	1	3	0
	.Thighs near parallel to floor at touch down and arms move forward during landing	0	3	2	1
	.Arms extend vigorously forward and upward upon take-off	3	0	3	0

[a]X denotes a yes response
[b]O denotes a no response

SKILL	QUALITATIVE COMPONENTS	OBSERV-ABILITY X	OBSERV-ABILITY 0	CONSISTENT W/RESEARCH X	CONSISTENT W/RESEARCH 0
SLIDE	.Weight transfer from the trailing foot to the lead foot along a straight line to the side	3	0	3	0
	.Period where both feet are off ground and remain parallel	3	0	3	0
LEAP	.Take off on one foot and a balanced landing on the other foot	3	0	3	0
	.Forward reach with arm opposite lead foot	3	0	3	0
	.An in-flight forward trunk lean of 80^{o} or less	0	3	2	1
OVERHAND THROW	.Step forward with foot opposite the throwing arm	3	0	3	0
	.Hip, trunk, and shoulder rotation	2	1	3	0
	.A downward arc of the throwing arm to initate overhand throw	3	0	3	0
	.Shift weight to forward foot	3	0	3	0
	.Side orientation toward direction of throw	3	0	1	2
	.Follow through well beyond ball release	3	0	3	0
CATCH	.Hands in front of body, elbows flexed near sides	3	0	3	0
	.Extension of the arms in preparation for ball contact	2	1	2	1
	.Contact the ball with hands only	3	0	3	0
	.Elbows bend as ball makes contact with hands	1	2	2	1
STATIONARY BOUNCE	.Contact ball at about hip height	3	0	3	0
	.Push ball with fingers of either hand	3	0	3	0
	.Ball contacts floor in front of the foot on the side of the bouncing arm	3	0	0	3

SKILL	QUALITATIVE COMPONENTS	OBSERV-ABILITY X	OBSERV-ABILITY 0	CONSISTENT W/RESEARCH X	CONSISTENT W/RESEARCH 0
KICK	.Arm-foot opposition	3	0	3	0
	.Step forward on the non-kicking leg with foot placement next to the ball	2	0	1	2
	.Hip extension and knee flexion	0	3	2	1
	.Leap just prior to the kick	3	0	3	0
	.A single continuous motion	3	0	3	0
TWO-HAND STRIKE	.Side orientation toward direction of travel	3	0	3	0
	.Hip and spine rotation	3	0	3	0
	.Weight transfer from back foot to front foot during swing	3	0	3	0
SIT-UPS	.Lie on back, knees bent, feet flat on floor, arms crossed over chest	3	0	3	0
	.Curl up to touch elbows to thighs	3	0	3	0
	.Return to lying position	3	0	3	0
SIT AND REACH	.Bend and reach forward to feet with one hand on top of the other	3	0	3	0
	.Keep legs straight	3	0	3	0
	.Hold position	3	0	2	1
PUSH-UPS	.Assume a prone position parallel to floor, toes on floor, hands directly under shoulders	3	0	3	0
	.Keeping body parallel to floor while lowering to 1-3" above floor	3	0	3	0
	.Raise body to starting position	3	0	3	0
RUN/WALK	.Run or walk continuously for five consecutive minutes	3	0	2	1

APPENDIX F

A STANDARDIZED CRITERION-REFERENCED TEST IN FUNDAMENTAL MOTOR SKILLS AND PHYSICAL FITNESS

TEST MANUAL

FIELD TEST EDITION

Field Service Unit in Physical Education
and Recreation for the Handicapped
Michigan State University

A STANDARDIZED CRITERION-REFERENCED
TEST IN FUNDAMENTAL MOTOR SKILLS AND
PHYSICAL FITNESS

TEST MANUAL

Name ______________ Student Classification ______________

Sex ______________ Test Date ______________

Age ________ Test Administrator ______________

OBJECTIVE	EQUIPMENT	DIRECTIONS	CRITERIA
RUN C R A O	50 feet of clear space; colored tape; chalk or other marking devices	Mark off a starting line. Mark off a finishing line 50 feet away. Instruct student to "run fast" from one line to the other.	- A period when both feet are off the floor - Toe-heel or heel-toe foot contact (not flat footed) - Arms move in opposition to legs, elbows bent - Smooth (not jerky) pattern for 50 feet
GALLOP C R A O	A minimum of 30 feet of clear space	Ask student to gallop leading with one foot and then the other.	- Brief period where both feet are off the floor - Trailing foot does not cross in front of lead foot at floor contact - Arms swing forward and upward - 3 consecutive gallop strides leading with each foot

SCORING KEY

C - Performs to criterion level as specified in description.
R - Performs skill, but lacks some quantitative or qualitative aspects.
A - Requires some physical assistance.
O - No response or inappropriate.

OBJECTIVE	EQUIPMENT	DIRECTIONS	CRITERIA
HOP C R A O	A minimum of 10 feet of clear space	Ask student to hop 3 times, first on one foot and then on the other.	- Carriage of non-support leg is slightly bent - Maintain upright body position, elbows bent - Arms swing forward - 3 consecutive hops forward on each foot
SKIP C R A O	A minimum of 30 feet of clear space	Ask student to skip. Emphasize the step-hop.	- Repeat the step and hop on alternate feet in a rhythmical (not jerky) pattern - Arms move in opposition to legs and are slightly bent - 3 consecutive skipping cycles (a step-hop)
HORIZONTAL JUMP C R A O	10 feet of clear space; tape or other marking devices	Mark off a starting line on floor, mat or carpet. Have student start behind the line. Tell student to "jump far."	- Two-foot takeoff and a two-foot landing - Arm thrust during takeoff with full extension of legs - 3 consecutive jumps of 2/3 of standing height or more
SLIDE C R A O	A minimum of 30 feet of clear space; colored tape or other marking device	Mark off a straight line. Ask student to do a slide 3 times to each side, staying on the line.	- Period where both feet are off ground and remain parallel - Weight transfer from trail foot to lead foot along a straight line to the side - 3 consecutive slides to each side
LEAP C R A O	A minimum of 50 feet of clear space	Ask student to leap. Tell him/her to take large steps leaping from one foot to the other.	- Takeoff on one foot and a balanced landing on the other foot - Forward reach with arm opposite lead foot - 3 consecutive leaps

OBJECTIVE	EQUIPMENT	DIRECTIONS	CRITERIA
SIT-UPS (Abdominal strength) C R A O	Mat or carpet; stopwatch	Instruct student that on the signal "go," s/he should do 3 sit-ups. Student's ankles should be held by another person. Delete this item for students under 8 years of age.	- Lie on back, knees bent, feet flat on floor, arms crossed over chest - Curl up to touch elbows to thighs - Return to lying position - 3 consecutive sit-ups
OVERHAND THROW C R A O	3-4-inch ball; target 5x5 feet; wall; 40 feet of clear space; tape for marking purposes	Place a target on the wall. Mark off a starting line 30 feet from wall. Ask student to throw ball "hard" at the target.	- Downward arc of throwing arm to initiate overhand throw - Hip/spine rotation - Weight transfer to foot opposite throwing arm - Follow-through well beyond ball release - Ball travels 30 feet or more - 3 consecutive throws
SIT AND REACH (Trunk/Leg flexibility) C R A O	Bench or flat vertical surface (wall)	Place a bench on its side. Sitting with legs straight and feet flat against bench, ask student to touch his/her toes without bending the knees.	- Assume a sitting position with legs together and knees straight - Bend and reach forward to feet with one hand on top of the other keeping legs straight - Hold for 3 seconds - 3 consecutive reaches
CATCH C R A O	6-8-inch playground ball; 20 feet of clear space; tape or other marking device	Mark off 2 lines 15 feet apart. Student stands on one line and the tosser on the other. Toss the ball underhand directly to student with a slight arc and tell him/her to "catch it with your hands." Only count those tosses that are between student's shoulders and waist.	- Hands in front of body, elbows bent near sides - Extension of arms in preparation for ball contact - Contact and control ball with hands only - 3 consecutive catches

OBJECTIVE	EQUIPMENT	DIRECTIONS	CRITERIA
STATIONARY BOUNCE C R A O	8-10-inch playground ball; hard, flat surface (floor, pavement)	Ask student to bounce ball as many times as he/she can, using one hand.	- Contact the ball between thighs and waist - Push ball with fingers of one hand only - Maintain a stable, stationary position - 3 consecutive bounces
PUSH-UPS (Arm/ shoulder strength) C R A O	A flat surface	Ask student to do 3 push-ups. Delete this item for students under 8 years of age.	- Assume prone position parallel to floor, toes on floor, hands directly under shoulders - Keep body parallel to floor while lowering body to 1-3 inches above floor - Raise body to starting position - 3 consecutive push-ups
KICK C R A O	8-10-inch playground ball; 30 feet of clear space; tape or other marking device	Mark off a line 30 feet away from a wall. Place the ball on the line and ask student to kick ball toward the wall.	- A preliminary forward leap on the non-kicking leg with foot placement next to the ball - Continuous kicking motion - Contact ball so it travels forward at least 30 feet - 3 consecutive kicks
TWO-HAND SIDEARM STRIKE C R A O	4-6-inch light-weight ball; batting tee; plastic bat; traffic cone or ball suspended on rope	Place the ball on the batting tee. Give student the bat and ask him/ her to "hit the ball far."	- Side orientation toward desired direction of travel - Hip and spine rotation during swing - Transfer weight onto front foot during swing - 3 consecutive strikes

OBJECTIVE	EQUIPMENT	DIRECTIONS	CRITERIA
WALK/RUN (Cardio-respiratory endurance) C R A O	A minimum of 30x50 feet of clear space; 4 chairs or cones; stopwatch or clock	Mark off a clear oval space by placing chairs or cones at 4 corners of a gym. If available, use a round track. Ask student to run around the outside of the chairs. Tell student to run slowly and as long as he/she can. If student tires, he/ she may walk. Delete this item for students under 8 years of age.	- Run/walk continuously - 5 consecutive minutes

Functions of the Test

1. Screening for the identification of students with specific needs in the fundamental motor and physical fitness skill areas.
2. Aid teachers, administrators and parents in making special education eligibility decisions in the physical education content area.
3. Aid teachers, administrators and parents in making placement and instructional programming decisions to meet the unique needs of the student in physical education.

Appropriate Student Population

This test was standardized on students within the age range of 3-12 years. The stratified sample consisted of 145 male (52%) and 134 female (48%) students in the regular education setting (normal, learning disabled and emotionally impaired), educable mentally impaired and trainable mentally impaired students. The sample was representative of whites (82%), blacks (15%), and other minority groups (3%). The standardization sample was reflective of the white collar and blue collar labor force. Students in urban, suburban, and rural schools in Michigan were sampled.

Directions for Administration and Scoring

General Directions

1. Before testing, read the entire test to acquaint yourself with each item, its equivalent requirements, directions, and criteria.
2. Assemble all materials prior to testing. Mark off the appropriate distances.

Materials	Distances to Mark Off
Colored tape/chalk/ marking devices	50 feet for running
Mat or carpet	30 feet for overhand throw and kick
Stopwatch	15 feet for catch
3-4-inch ball	30x50 feet of clear space for run/walk
Target 5x5 feet	
Bench or flat vertical surface	
6-8-inch playground ball	
9-10-inch playground ball	
4-6-inch light-weight ball	
Plastic bat	
Batting tee/traffic cone/ suspended ball	
4 chairs or cones	

3. The test administrator may want to set up stations to facilitate moving students from one item to the next. Start at any point in the test battery; however,

continue according to the specified order. The test was designed to alternate strenuous with less strenuous items.

4. The average administration time to assess one student on all 16 items is 25 minutes. This will obviously vary according to the age and performance level of the student.

Specific Directions

1. Precede the assessment by an accurate demonstration and verbal request.
2. Allow two practice trials where no feedback or instruction is provided to assure that the student understands what to do.
3. Provide one additional demonstration where the student does not know what to do on the first trial.
4. Provide instructions in the student's native language or mode of communication (e.g., sign language, bliss symbols).

Scoring

The scoring grid would appear as such:

OBJECTIVE	EQUIPMENT	DIRECTIONS	CRITERIA
STATIONARY BOUNCE C R A O	One 8-10-inch playground ball	Instruct student to release ball, and then bounce (dribble) it in place.	-Contact ball between thigh and waist -Push ball with fingers of one hand only -Maintain a stable, stationary position -Bounce ball 3 consecutive trials

There are four letters that can be circled to describe the student's response. Circle the letter that best describes the individual's responses.

C - Student completes the item according to all stated criteria. Any quantitative criteria stating "consecutive trials" requires performance of all qualitative criteria the stated number of times.

R - Student responds according to some of the criteria but not all of the stated criteria (lacks quantitative or qualitative aspects).

A - Student needs some form of physical assistance to respond, such as manipulating the student, guiding a student's hand or tapping of student's limb. Through physical assistance, the student can perform a minimum of one qualitative criterion.

O - The student does not respond, responds inappropriately, resists assistance, or cannot perform a minimum of one qualitative criterion with physical assistance.

Sample Item: Stationary Bounce

The materials needed are an 8-10-inch playground ball and a hard surface (floor, pavement) for the student to bounce the ball on. Ask the student to bounce (dribble) the ball. Demonstrate the skill according to the stated criteria. Observe student to make sure he/she performs the skill the specified number of repetitions or the specified distance. Then, circle the letter on the score sheet that best describes the student's performance.

C - Criterion Level: The student performs the skill in accordance with all the stated criteria:

- contact ball between thigh and waist
- push ball with fingers of one hand only
- maintain stable, stationary position
- bounce ball 3 consecutive times

R - Rudimentary Level: The student performs the skill without assistance but not in accordance with all the stated criteria. There must be at least one observable qualitative or quantitative aspect. Examples for "bounce":

- slaps ball with hand; or
- uses two hands to bounce; or
- travels while bouncing; or
- doesn't bounce 3 consecutive times

A - Assistance Level: The student performs the skill at the rudimentary level with some physical prompting. Example for "bounce":

- the teacher holds student's hand(s) and guides him/her through the bouncing action.

O - Other Response Level: The student does not achieve a minimum of one criterion even with physical assistance, refused to respond, or responds inappropriately. Examples for "bounce":

- after student is given a demonstration, he/she throws the ball away
- student releases ball and chases after it as it bounces, attempting to hit it, but without success. Student resists assistance of instructor.

Suggestions to Aid in the Administration and Scoring of the Test Based on Field Test Experiences

RUN

1. Listen to the feet as student runs. If they are slapping loudly on the floor, this is a sign that toe-heel, heel-toe contact is not being made.
2. When assisting the student, hold his/her hand and run slightly in front of him/her, pulling on the arm. To score "A", be sure there is a period when both feet are off the floor; otherwise it is just a walk.

GALLOP

1. Tell student to put his/her favorite foot in front and gallop. Then, tell him/her to put the other foot in front. Be sure to observe student with each foot as the lead foot.
2. If student needs assistance, hold his/her hand and verbally cue as you gallop together.

HOP

1. The hop should be controlled. The student should be balanced on his/her landing.
2. If student needs assistance, allow him/her to hold onto something for support (person, wall, chair, etc.).

SKIP

1. Student must demonstrate s/he can perform the step-hop pattern on each foot in order to receive a "C" or an "R".
2. Skip should look smooth and rhythmical and not choppy.
3. If student needs assistance, hold his/her hand and verbally cue as you slowly "step-hop" together.

HORIZONTAL JUMP

1. When assisting student, stand facing him/her, hold both hands and pull forward as you verbally cue him/her to jump.

SLIDE

1. Have student do the slide along a line near a wall. Cue the student to watch the wall. Be sure feet are parallel to each other.
2. Watch to see that there is a moment when both feet are off the ground.
3. If student needs assistance, stand behind him/her and touch each leg as you verbally cue student to pick it up and slide it along a line.

LEAP

1. Verbally cue student to jump from one foot to the other. Tell him/her to pretend there are big puddles on the ground and to take "big jumping steps" to get over the puddles.

2. Students will often "run" instead of leap. Watch to make sure the legs are straight (not bent at the knee) for each stride.
3. If student needs assistance, stand beside him/her holding his/her hand. Touch one leg and verbally cue to jump forward onto it. Repeat with the other leg.

SIT-UPS

If the student needs assistance, place your arm across the lower back and exert pressure upward to initiate sitting up.

OVERHAND THROW

1. When beginning the throw, the ball should be extended slightly in front of the student, then dropped downward in an arc (a wind-up), then drawn up and back over the shoulder.
2. In watching for the hip and spine rotation, student's hips should turn first, followed by rotation of the upper body as ball is released.

SIT AND REACH

1. If you do not have a bench to use, you can have student sit with feet against a wall.
2. When assisting student, hold his/her knees down and help him/her reach forward to touch toes.

CATCH

1. If student traps ball against the body with his/her arms, score "R".
2. If student needs assistance, stand behind him/her. Have someone toss the ball. Take student's hands and assist him/her to catch ball with hands or trap the ball against the body.

BOUNCE

1. If student bounces ball with two hands, ask him/her to use only one hand. If student still uses two hands, score "R".
2. Student should not be moving his/her feet while bouncing.
3. If student needs assistance, stand behind him, drop ball for him/her, take student's hand and bounce ball together.

PUSH-UPS

1. Be sure hands are pointed forward, student on toes.
2. The lowering action should be caused by the elbows bending. The raising action will be caused by the elbows extending.
3. If student needs assistance, hold him/her at the waist with both hands. Verbally cue to "bend the elbows" as you assist to lower and "extend the elbows" as you assist to raise.

KICK

1. The forward leap onto the non-kicking foot is sometimes difficult to observe. A good indication that this has occurred is if a) the ball is airborne (vs. rolling on ground, and b) student hops forward onto the non-kicking foot following the kick.
2. When demonstrating, emphasize and verbally cue the leap forward onto the non-kicking foot.
3. If student needs assistance, hold one hand and use your other hand to guide the kicking leg to contact the ball.
4. The teacher or aide should position themselves close to the ball in case student steps on ball and falls backwards.

STRIKE

1. Weight transfer is observable by student picking up forward foot and stepping onto it as s/he swings.
2. If student needs assistance, stand behind him/her, holding your hands on top of the student's. Swing together to hit the ball.

WALK-RUN

1. Score "C" if student walks/runs for five consecutive minutes.
2. Score "R" if student stops once but resumes walking/ running after verbal prompt.

3. Score "A" if student stops one or more times but resumes running/walking after verbal and physical prompting.
4. Score "O" if student stops and sits down, does not respond to verbal and physical prompting, or refuses to continue.

Directions for Using Student Profiles

1. Locate the appropriate student profile (in the next section of this manual) according to the age of the student (3-12 years).
2. Transfer student's test results onto the student profile.
   - 2.1 Determine which student classification is appropriate (normal, educable mentally impaired, or trainable mentally impaired). Determine the student's gender for the overhand throw and two-hand strike (M-F).
   - 2.2 Using the student's test score sheet, transfer the item scores onto the student profile by circling the appropriate letter for each item.
   - 2.3 Convert each of the item scores into a numerical form using the following scale: C = 4, R = 3, A = 2, O = 1.
   - 2.4 Calculate the locomotor subtotal by adding up the first seven test items. Circle the subtotal score on the profile closest to that of the student.

2.5 Calculate the object control subtotal by adding up test items 8-12. Circle the score on the student profile closest to that of the student.

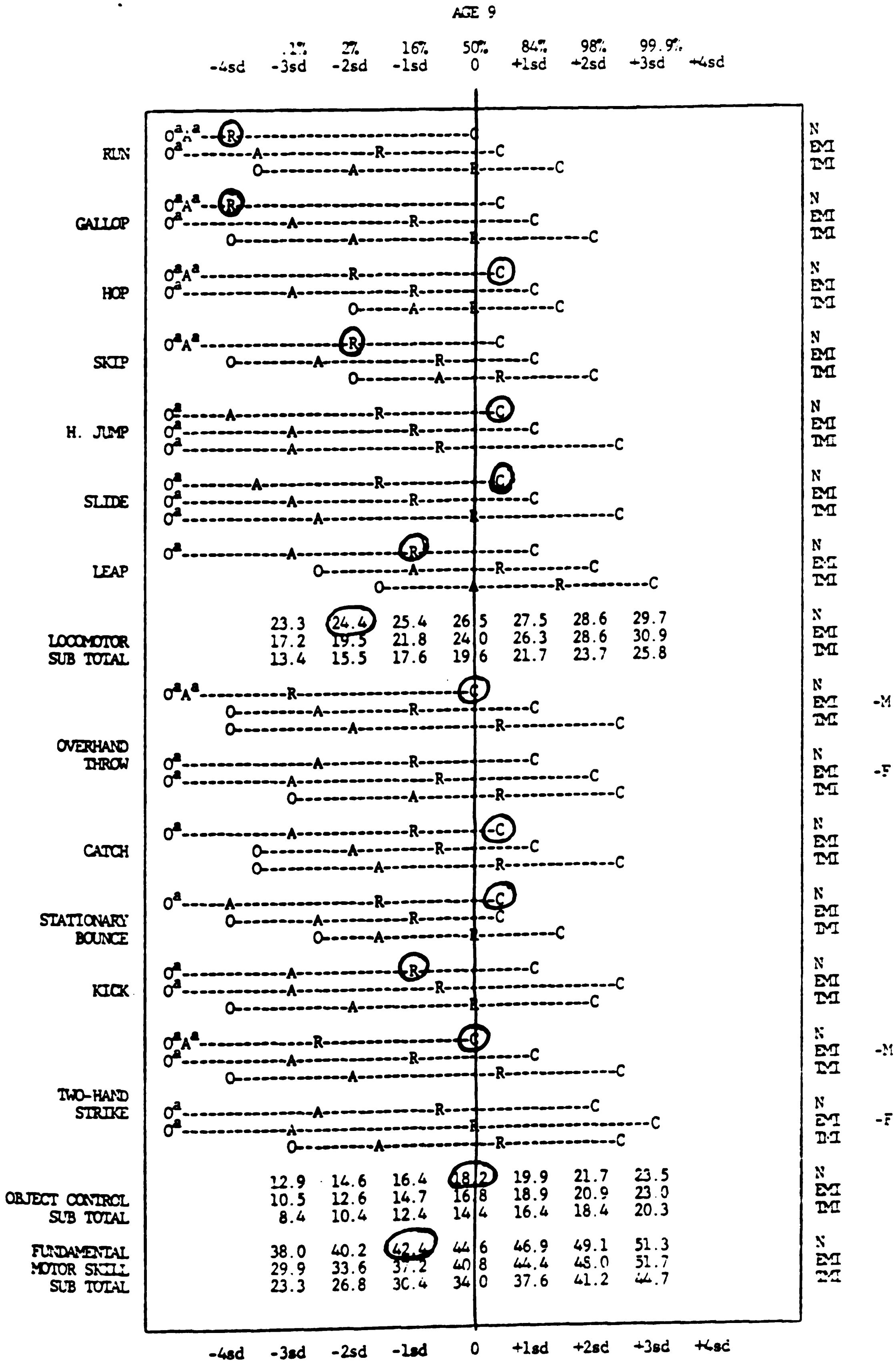
2.6 Calculate the fundamental motor skill subtotal by adding the student's locomotor subtotal and object control subtotal. Circle the score on the profile closest to that of the student.

2.7 For students 8-12 years of age, calculate the physical fitness subtotal by adding up test items 13-16. Circle the score on the profile closest to that of the student.

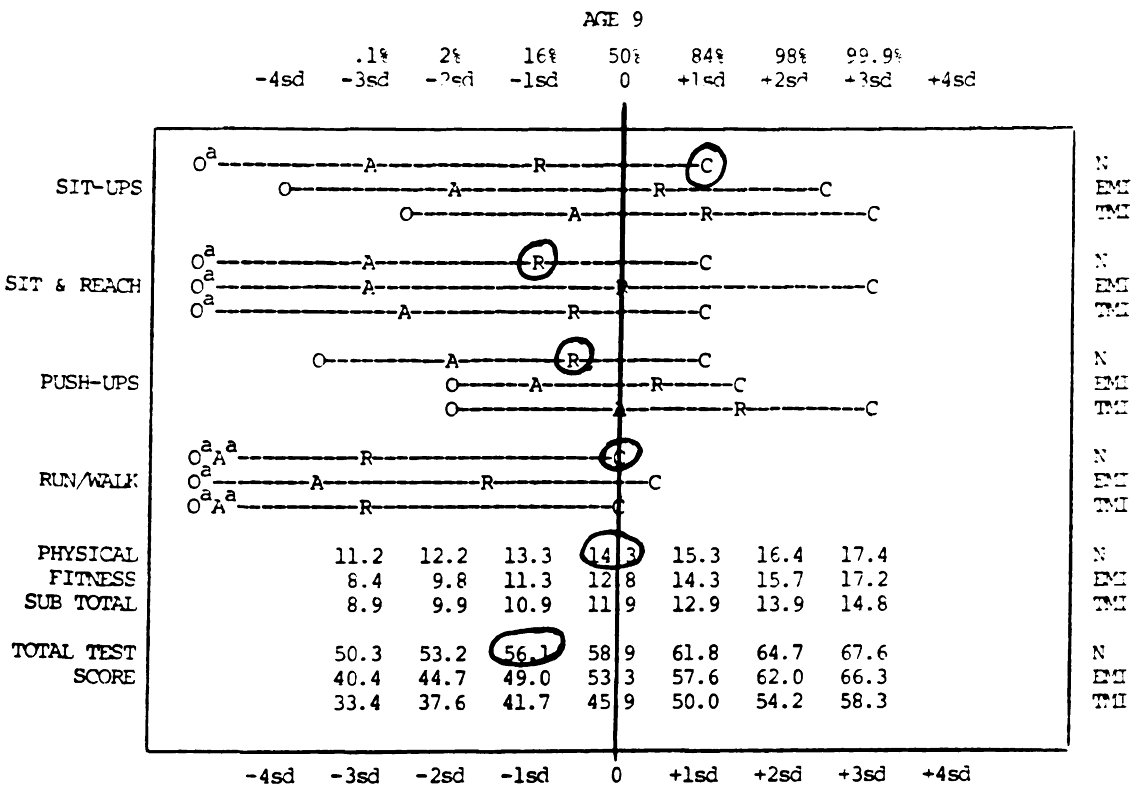
2.8 For students 8-12 years of age, calculate the total test score by adding up the three subtotal scores. Circle the score on the profile closest to that of the student.

3. For ease of viewing the student profile and making decisions about the student's performance, it is recommended that a light line be drawn between the zero at the top of the profile and the zero at the bottom. This line will represent the average performance in which to make normative interpretations.

The completed profile should look something like the following example of a 9-year-old normal male student.

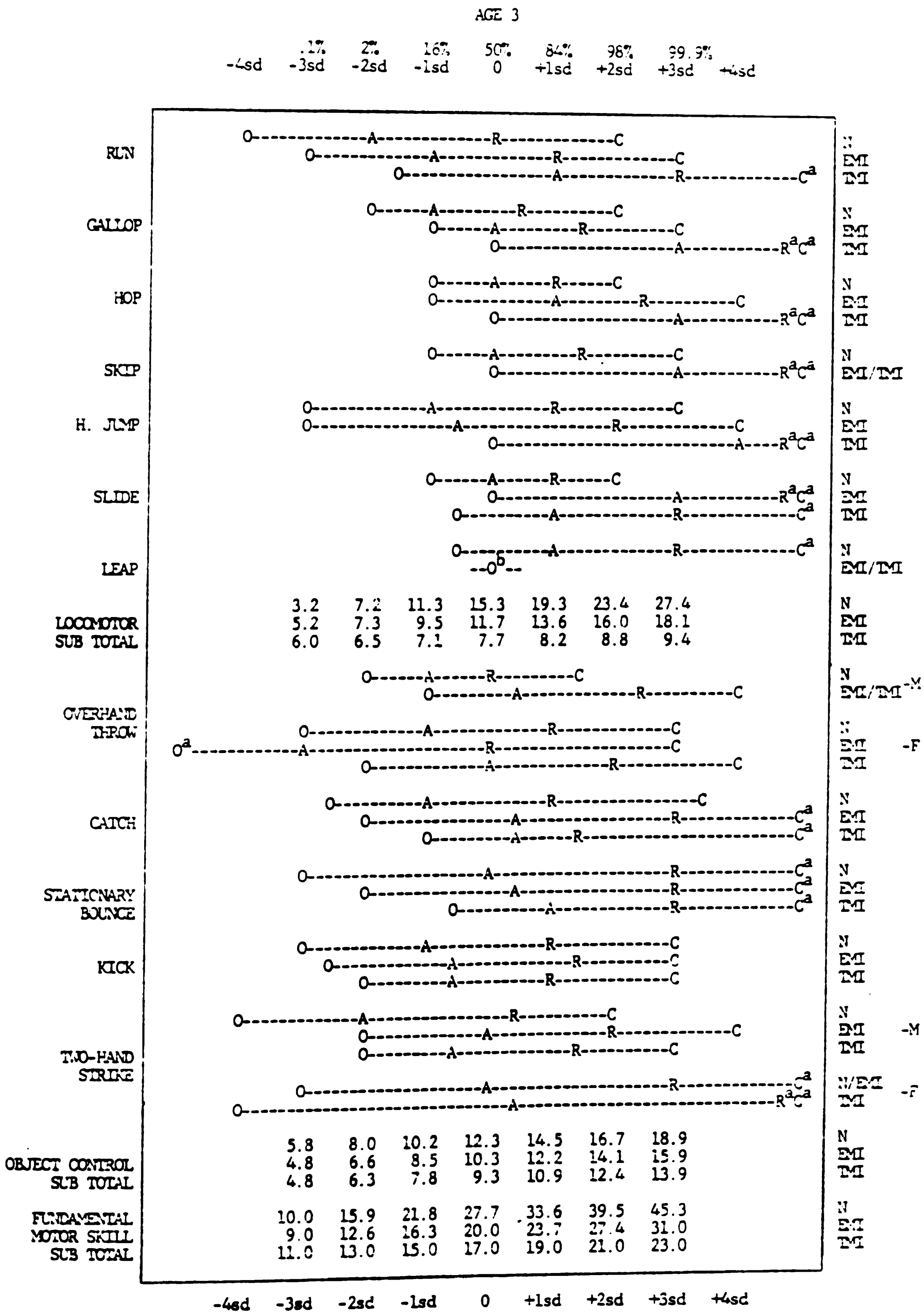


[a]This performance level falls above or below the mean by more than 4 standard deviations

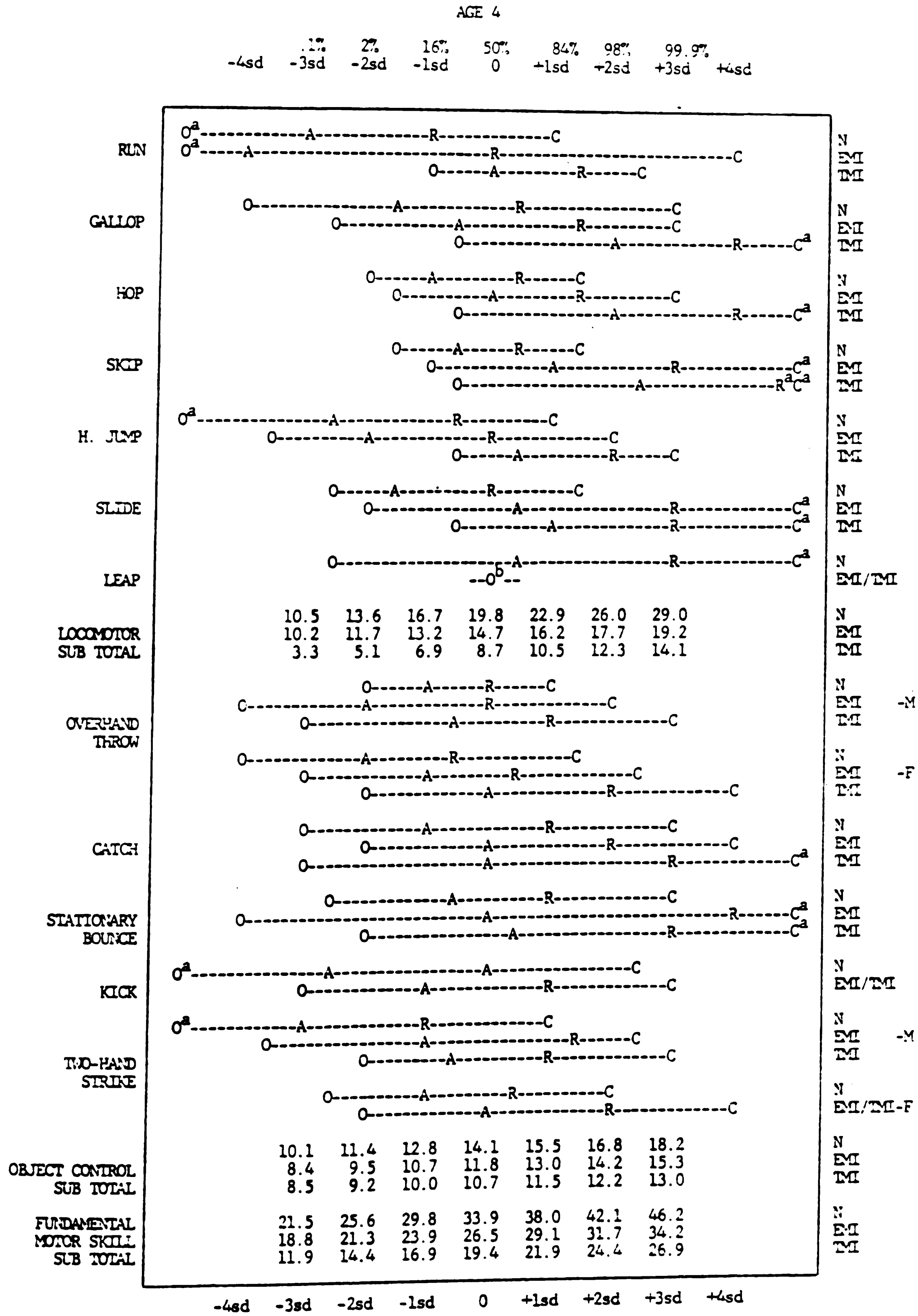


[a]This performance level falls above or below the mean by more than 4 standard deviations

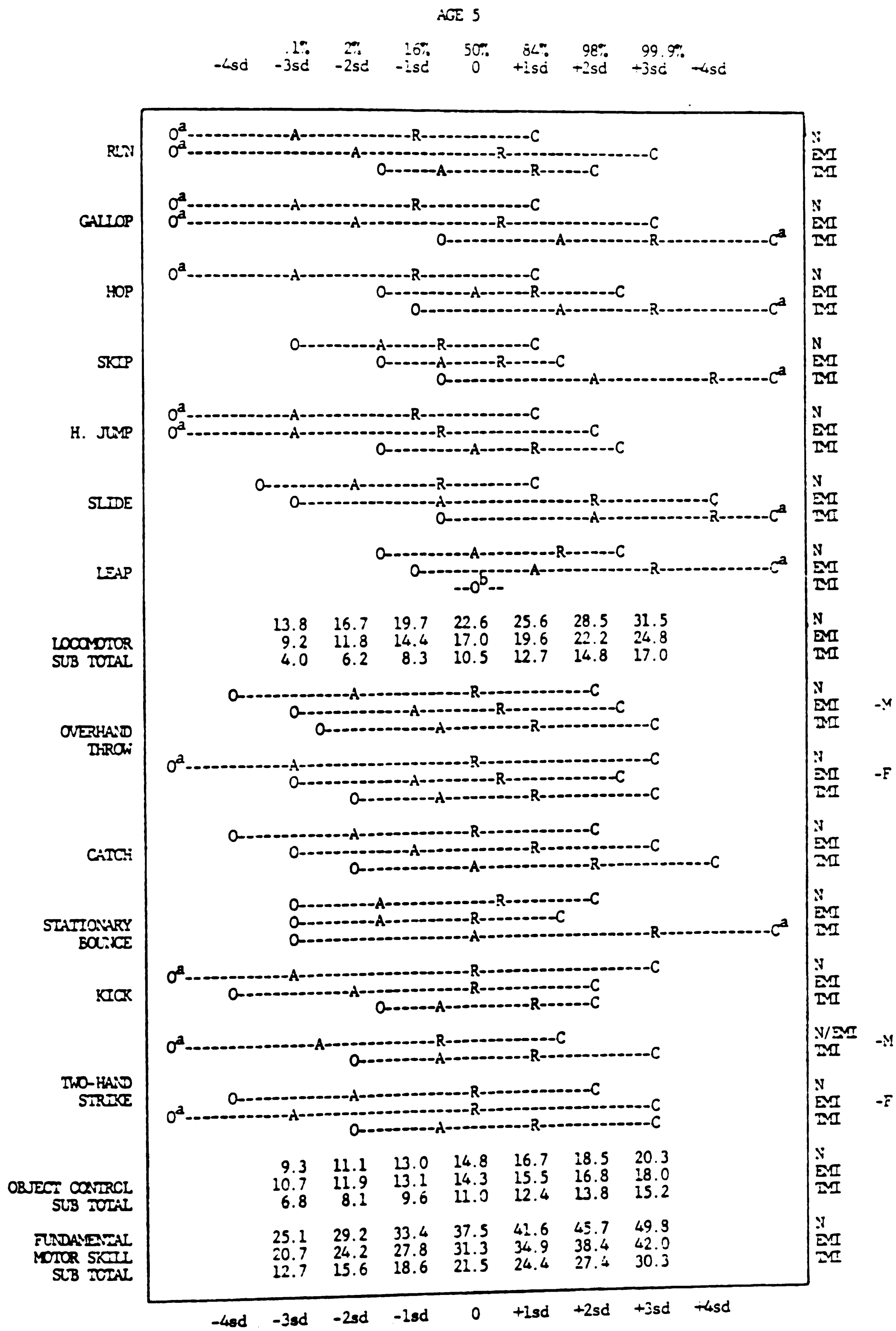
STUDENT PROFILES



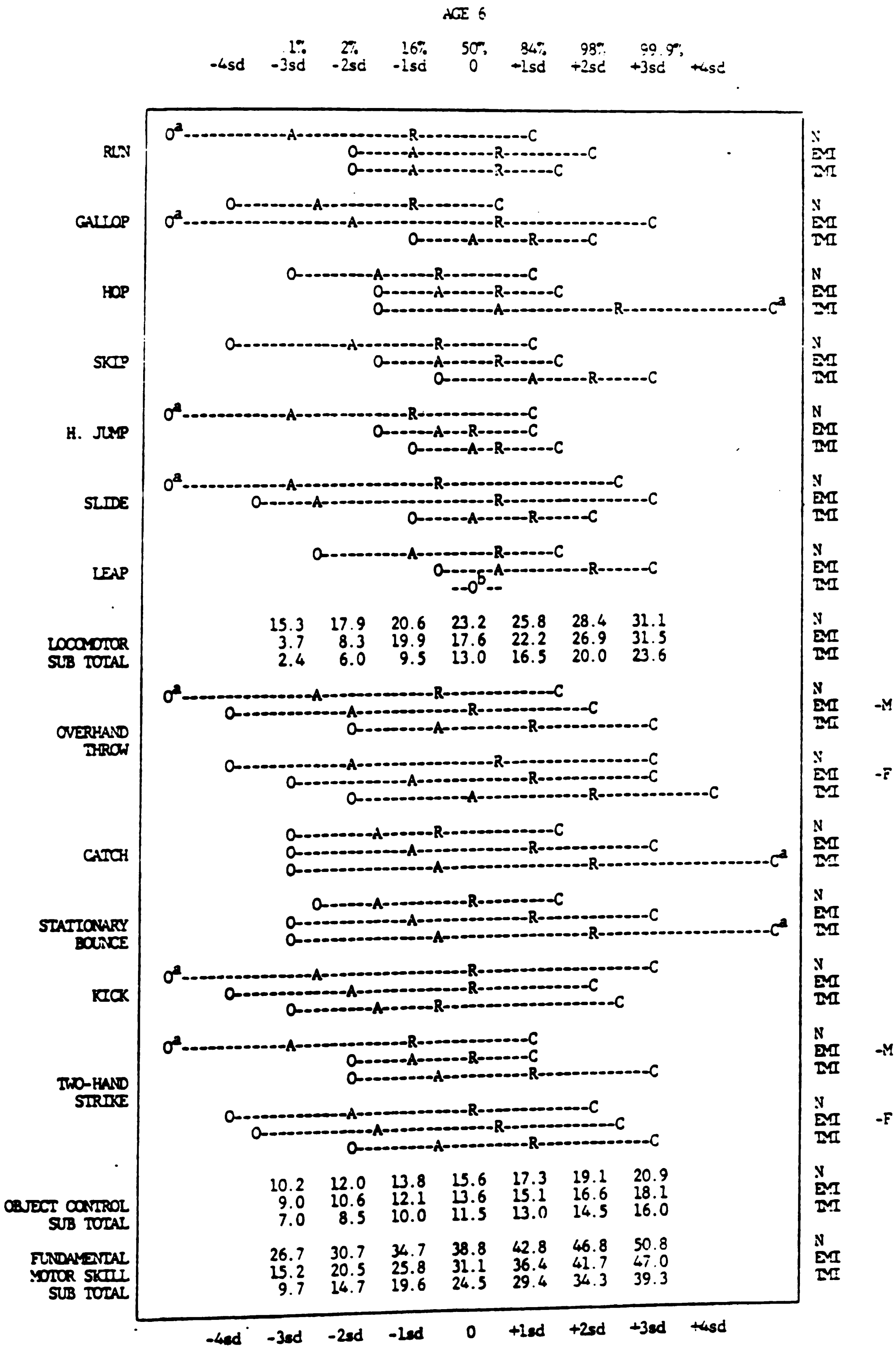
[a] This performance level falls above or below the mean by more than 4 standard deviations
[b] All students in this classification scored at this level of performance



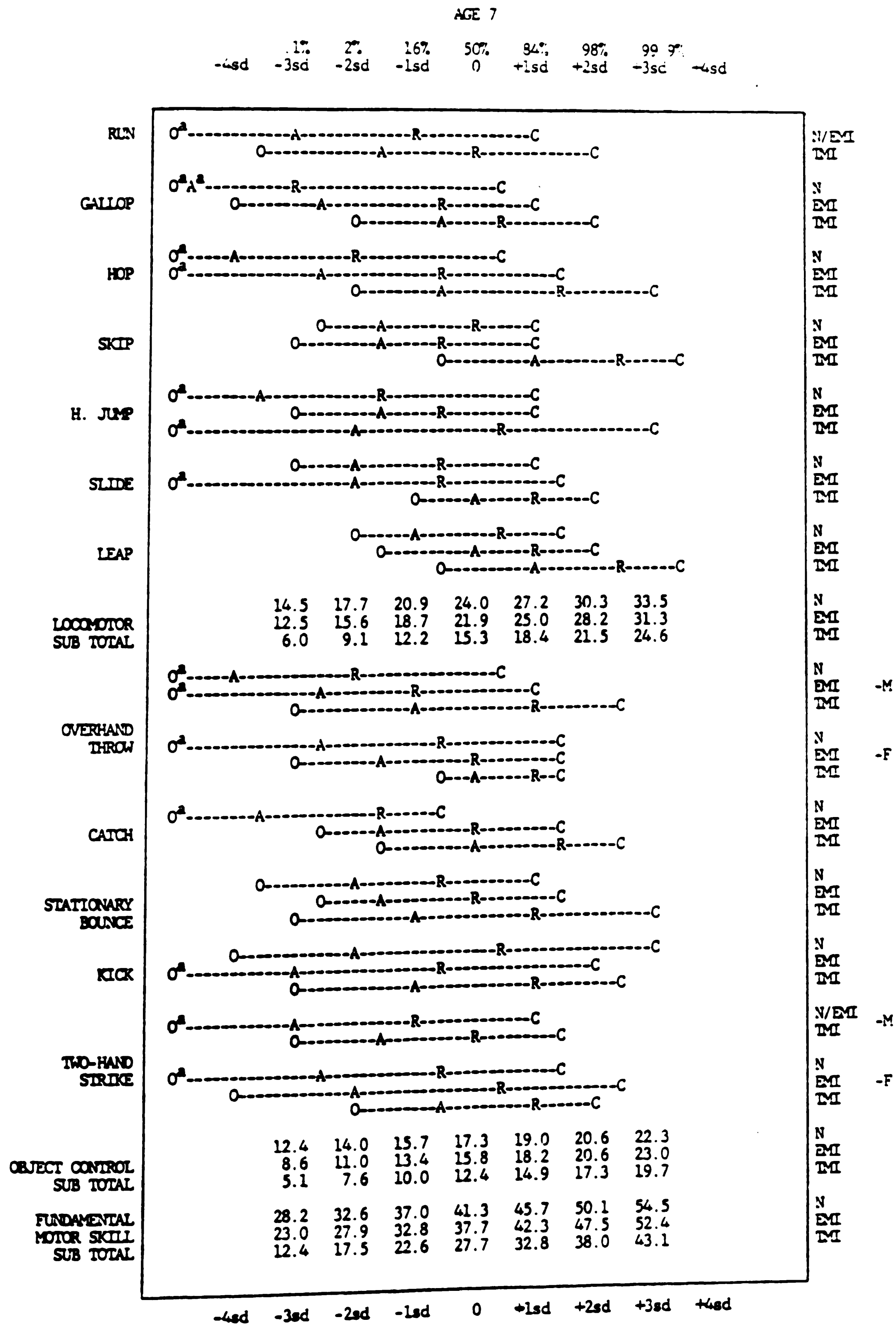
[a] This performance level falls above or below the mean by more than 4 standard deviations
[b] All students in this classification scored at this level of performance



[a]This performance level falls above or below the mean by more than 4 standard deviations
[b]All students in this classification scored at this level of performance

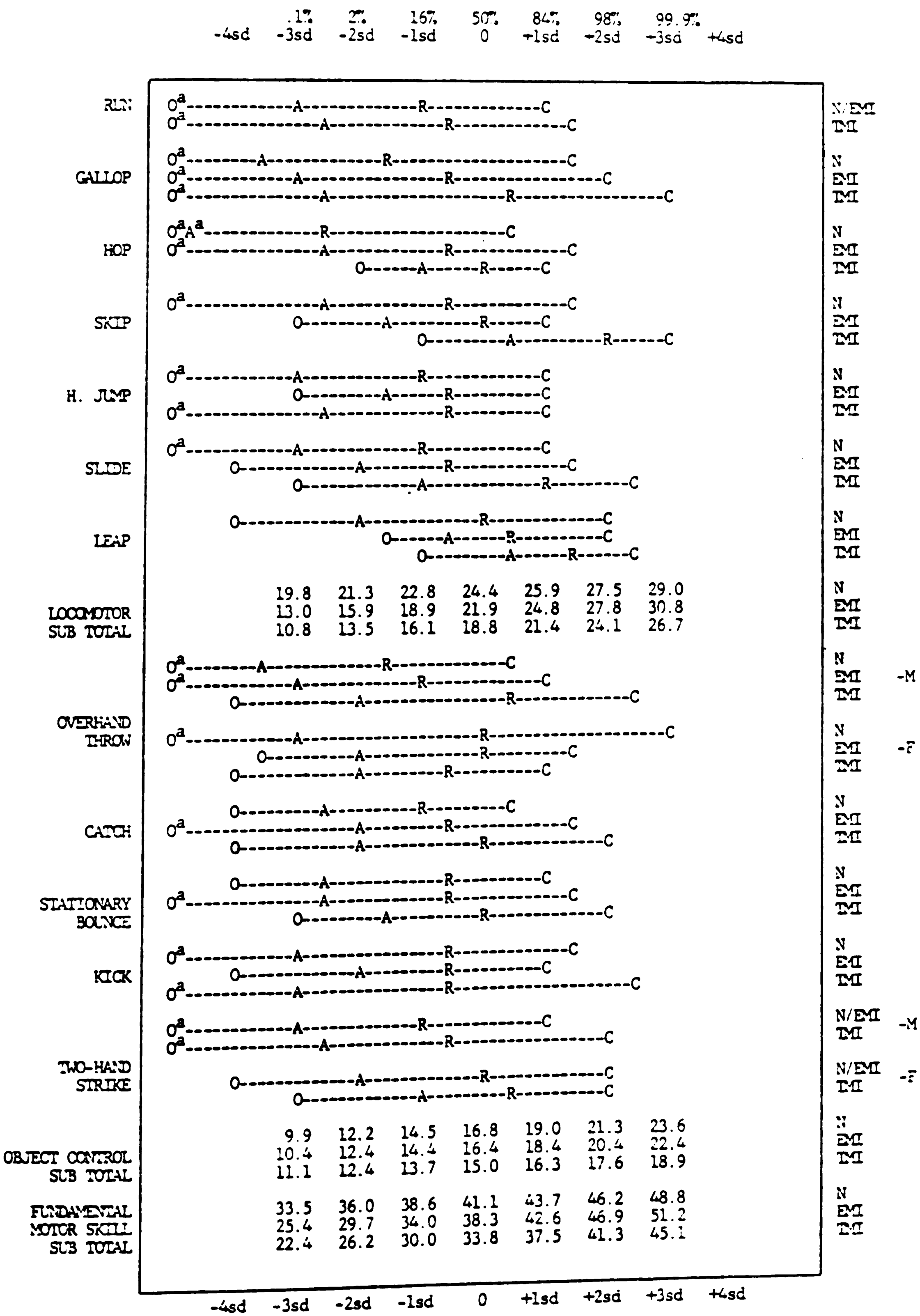


[a]This performance level falls above or below the mean by more than 4 standard deviations
[b]All students in this classification scored at this level of performance



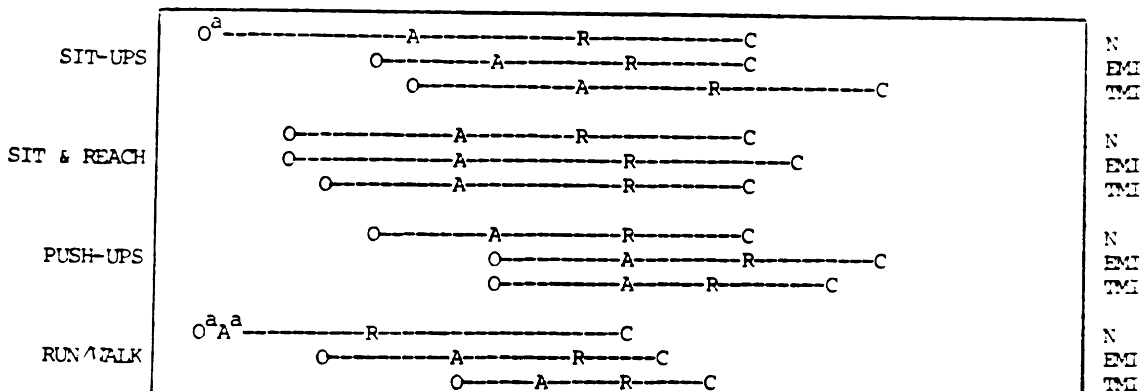
[a]This performance level fails above or below the mean by more than 4 standard deviations

AGE 8



[a]This performance level falls above or below the mean by more than 4 standard deviations

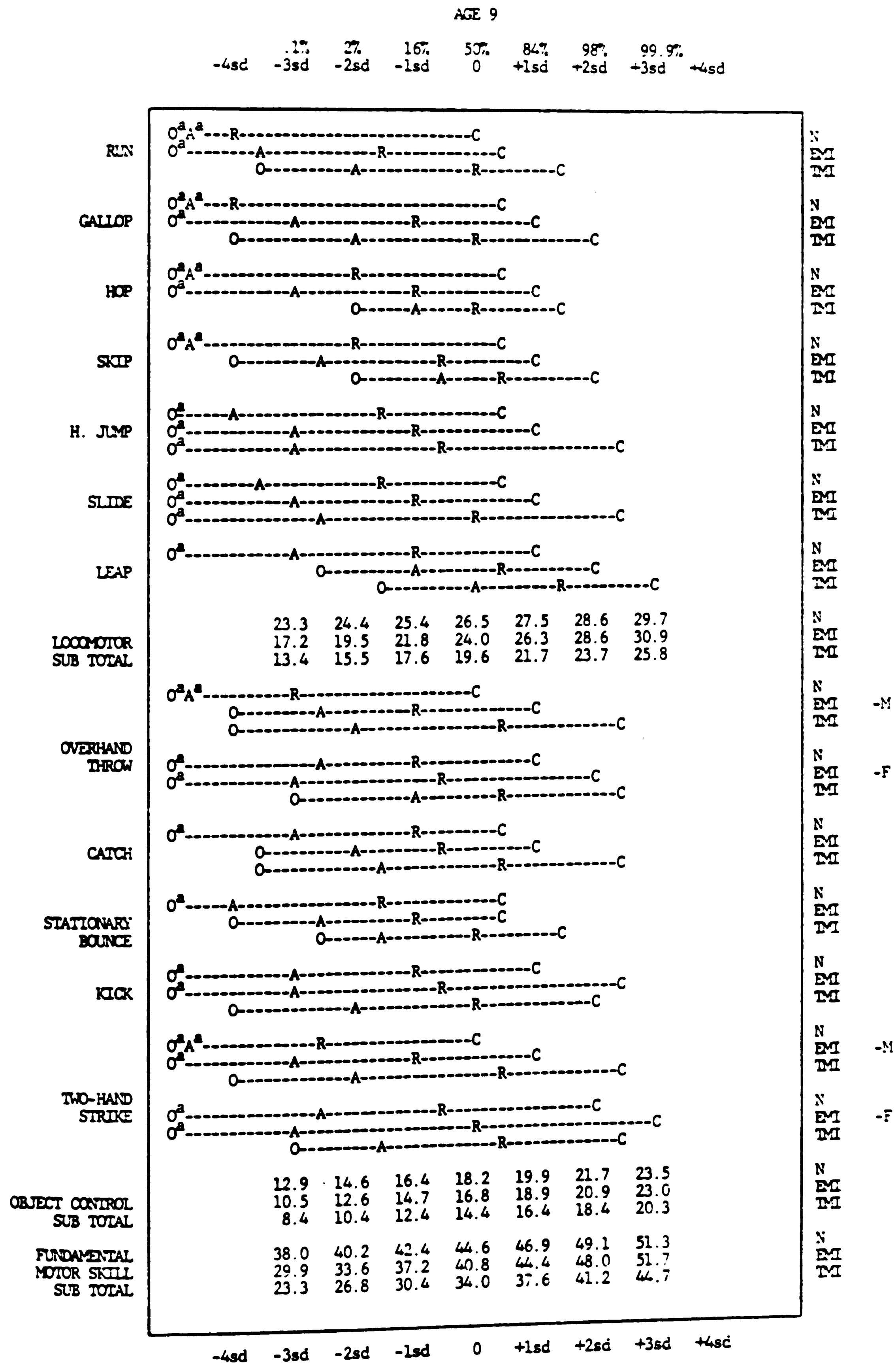
AGE 8



		-4sd	.1% -3sd	2% -2sd	16% -1sd	50% 0	84% +1sd	98% +2sd	99.9% +3sd	+4sd
PHYSICAL FITNESS SUB TOTAL	N		8.8	10.3	11.8	13.3	14.8	16.3	17.8	
	EMI		6.4	8.2	9.9	11.7	13.4	15.2	16.9	
	TMI		5.7	7.3	8.9	10.5	12.1	13.7	15.3	
TOTAL TEST SCORE	N		42.9	46.8	50.6	54.4	58.3	62.1	66.0	
	EMI		33.3	38.7	44.2	49.6	55.1	60.5	66.0	
	TMI		29.4	34.4	39.3	44.3	49.2	54.1	59.1	

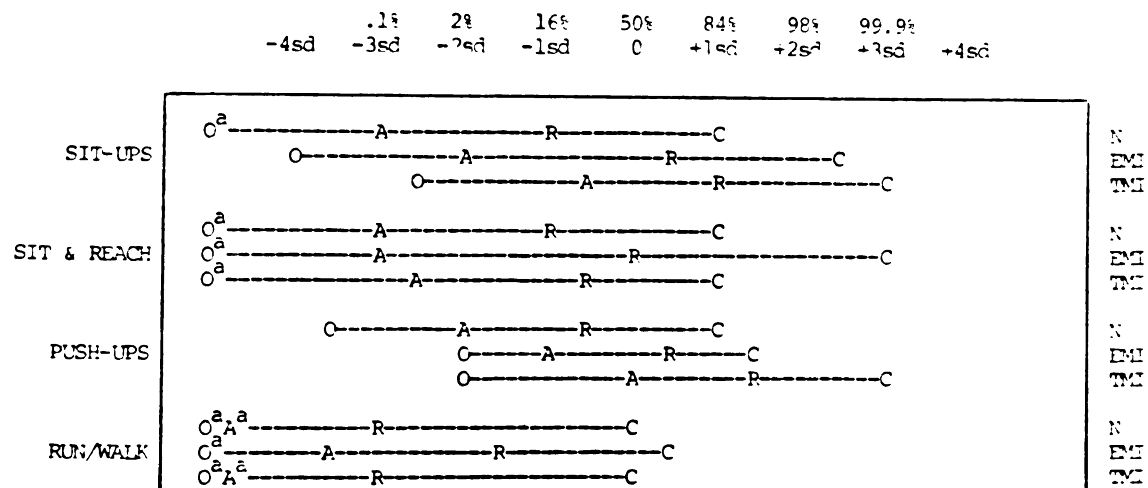
-4sd -3sd -2sd -1sd 0 +1sd +2sd +3sd +4sd

[a] This performance level falls above or below the mean by more than 4 standard deviations



[a]This performance level falls above or below the mean by more than 4 standard deviations

AGE 9

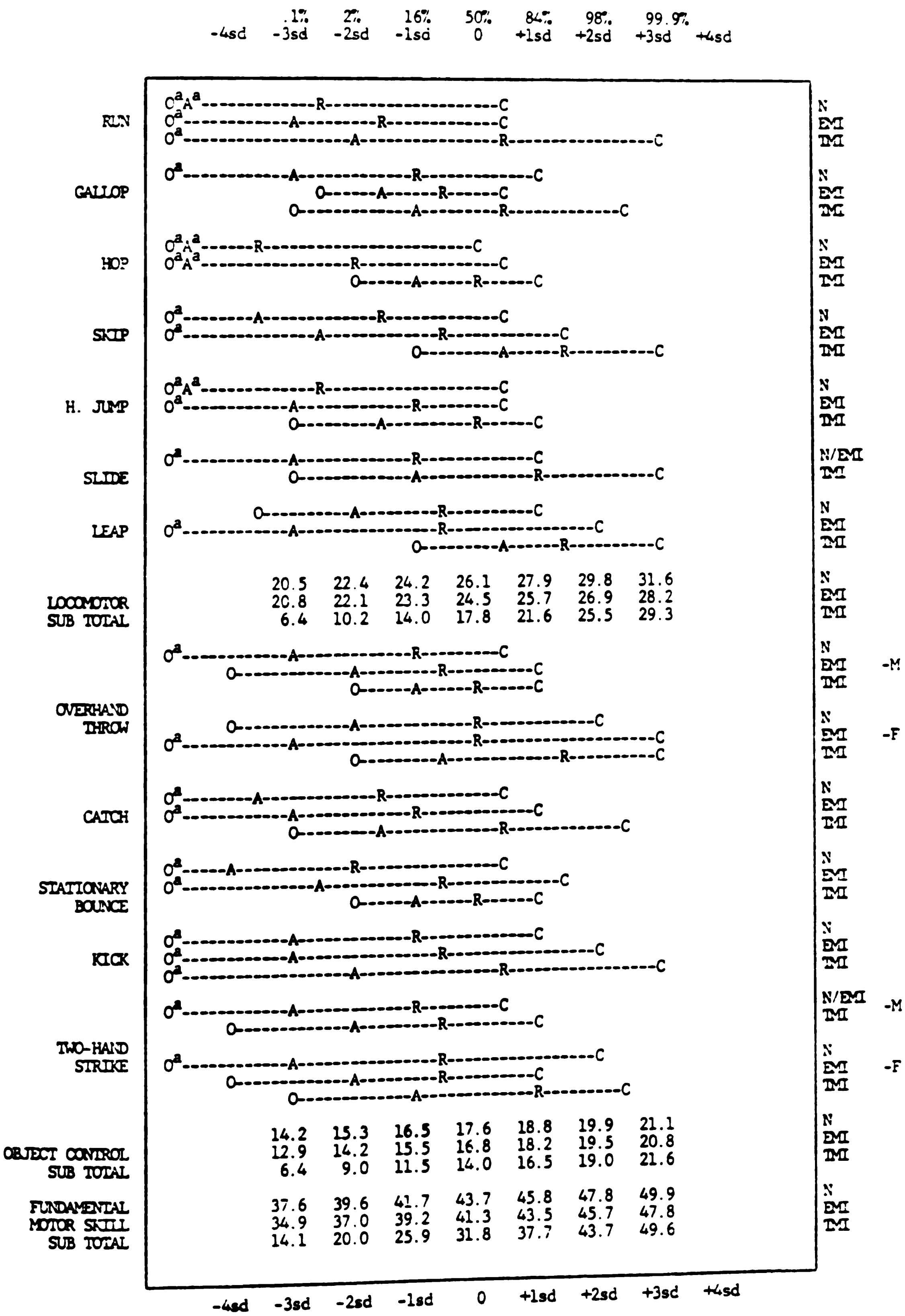


	-3sd	-2sd	-1sd	0	+1sd	+2sd	+3sd	
PHYSICAL FITNESS SUB TOTAL	11.2	12.2	13.3	14.3	15.3	16.4	17.4	N
	8.4	9.8	11.3	12.8	14.3	15.7	17.2	EMI
	8.9	9.9	10.9	11.9	12.9	13.9	14.8	TMI
TOTAL TEST SCORE	50.3	53.2	56.1	58.9	61.8	64.7	67.6	N
	40.4	44.7	49.0	53.3	57.6	62.0	66.3	EMI
	33.4	37.6	41.7	45.9	50.0	54.2	58.3	TMI

-4sd -3sd -2sd -1sd 0 +1sd +2sd +3sd +4sd

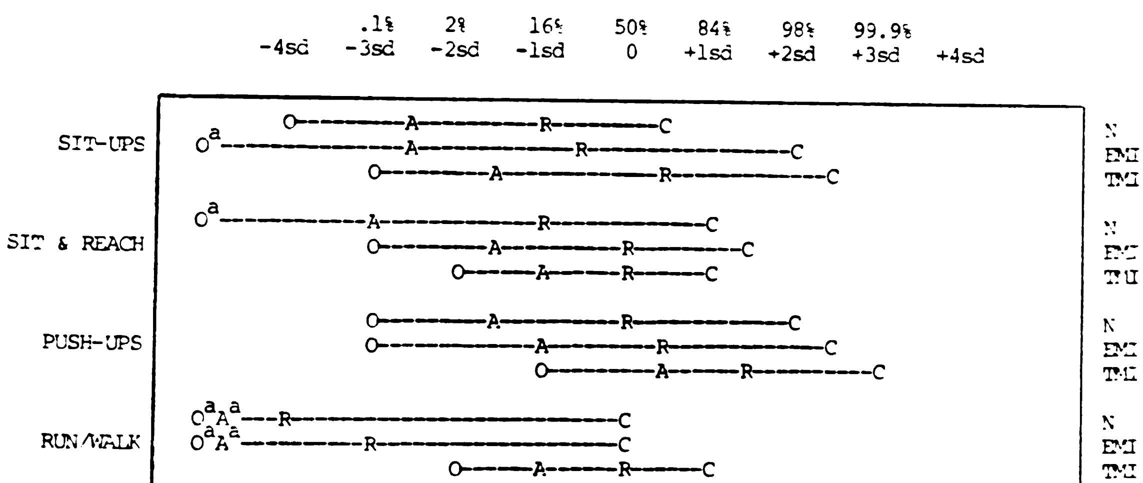
[a]This performance level falls above or below the mean by more than 4 standard deviations

AGE 10



[a]This performance level falls above or below the mean by more than 4 standard deviations

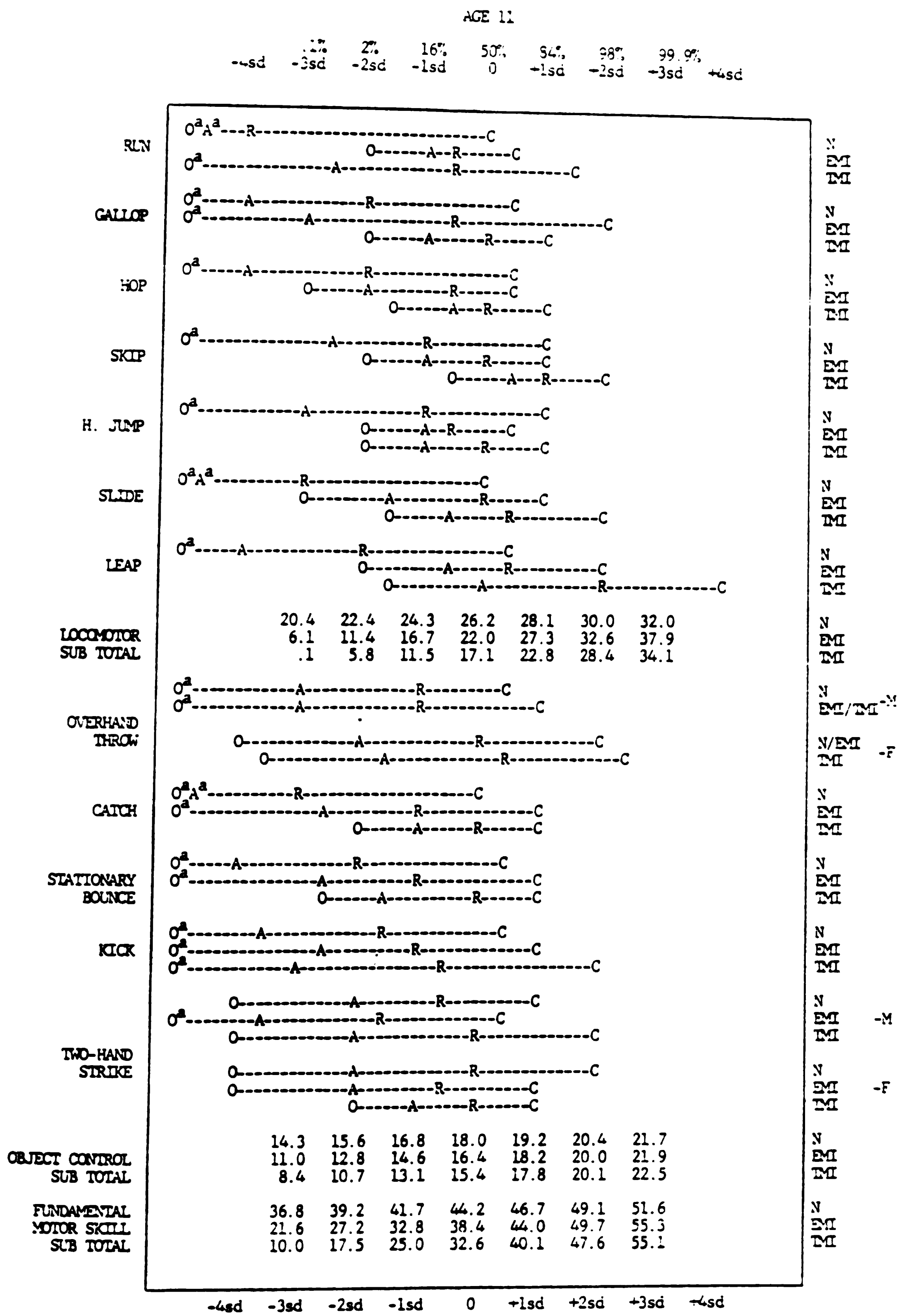
AGE 10



		-3sd	-2sd	-1sd	0	+1sd	+2sd	+3sd	
PHYSICAL FITNESS SUB TOTAL		10.0	11.3	12.7	14.0	15.4	16.7	18.1	N
		9.4	10.5	11.7	12.8	14.0	15.2	16.3	EMI
		3.0	5.5	8.0	10.5	13.0	15.5	18.0	TMI
TOTAL TEST SCORE		49.1	52.0	54.9	57.7	60.7	63.6	66.6	N
		45.0	48.0	51.1	54.2	57.2	60.3	63.3	EMI
		17.8	26.0	34.2	42.3	50.5	58.7	66.9	TMI

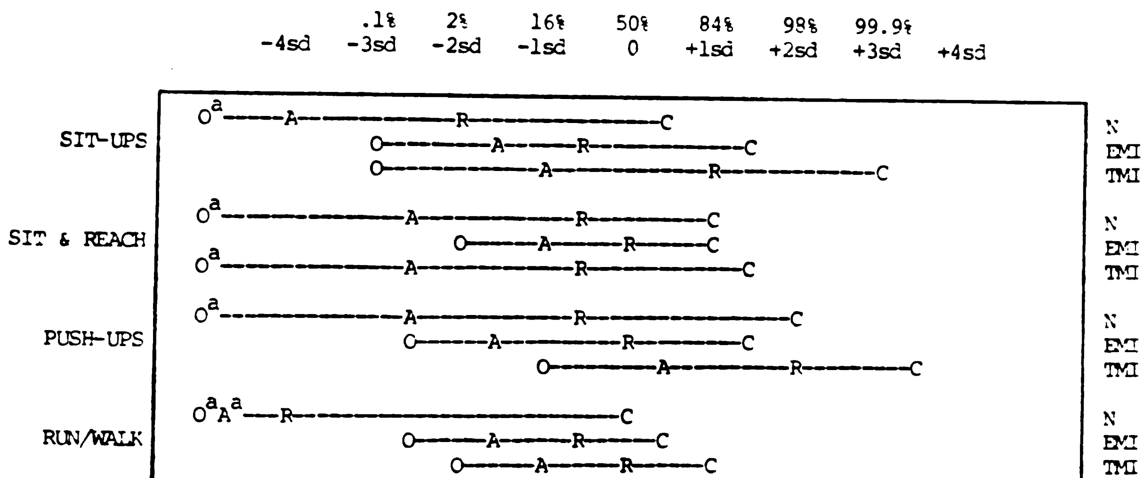
-4sd -3sd -2sd -1sd 0 +1sd +2sd +3sd +4sd

[a]This performance level falls above or below the mean by more than 4 standard deviations



[a]This performance level falls above or below the mean by more than 4 standard deviations

AGE 11

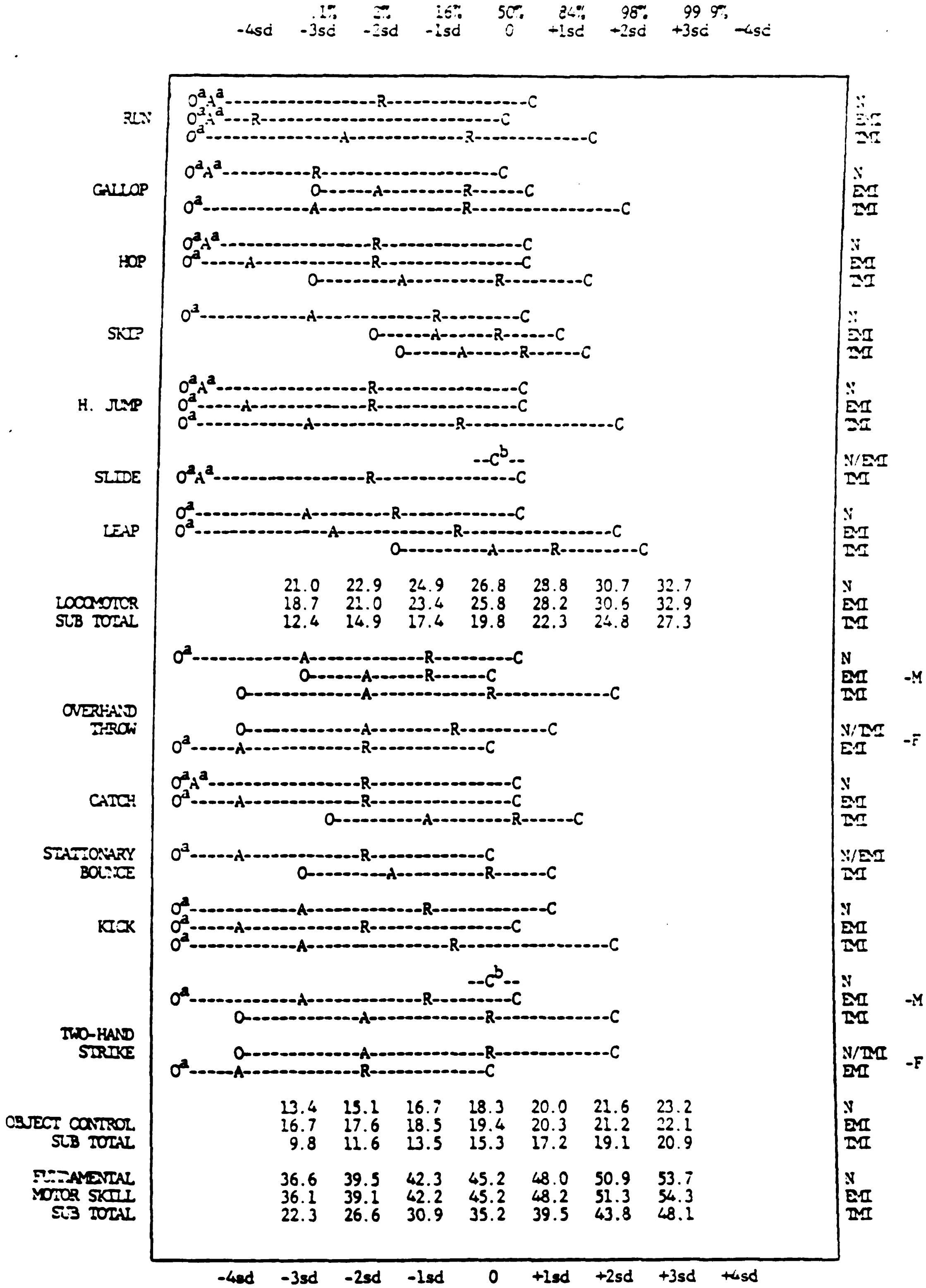


		-3sd	-2sd	-1sd	0	+1sd	+2sd	+3sd	
PHYSICAL FITNESS SUB TOTAL		11.7	12.6	13.5	14.4	15.3	16.2	17.1	N
		2.3	5.6	8.9	12.1	15.4	18.7	22.0	EMI
		4.2	6.3	8.4	10.6	12.7	14.8	16.9	TMI
TOTAL TEST SCORE		49.7	52.7	55.6	58.6	61.6	64.5	67.5	N
		24.4	33.1	41.8	50.6	59.3	68.0	76.8	EMI
		15.2	24.5	33.8	43.1	52.4	61.7	71.0	TMI

-4sd -3sd -2sd -1sd 0 +1sd +2sd +3sd +4sd

[a]This performance level falls above or below the mean by more than 4 standard deviations

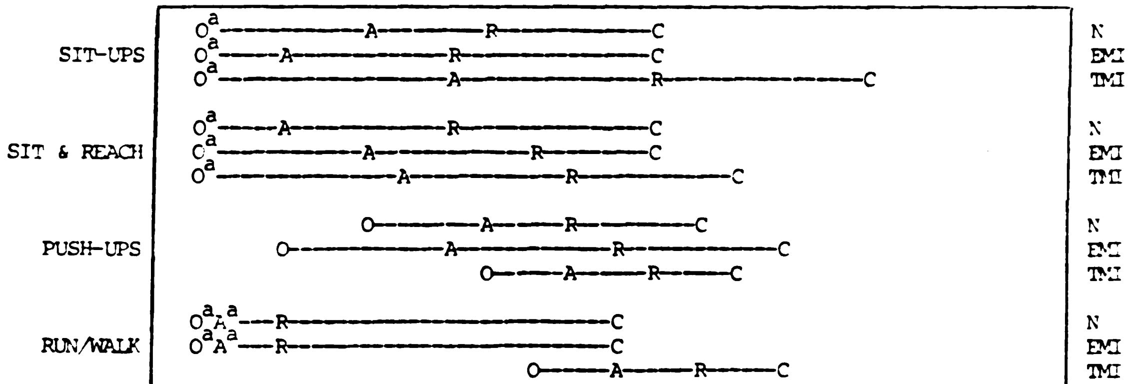
AGE 12



[a]This performance level falls above or below the mean by more than 4 standard deviations
[b]All students in this classification scored at this level of performance

AGE 12

	-4sd	-3sd	-2sd	-1sd	0	+1sd	+2sd	+3sd	+4sd	
		.1%	2%	16%	50%	84%	98%	99.9%		



	-3sd	-2sd	-1sd	0	+1sd	+2sd	+3sd	
PHYSICAL	10.0	11.6	13.2	14.8	16.4	18.0	19.6	N
FITNESS	11.7	12.6	13.5	14.4	15.3	16.2	17.1	EMI
SUB TOTAL	3.8	6.0	8.3	10.5	12.8	15.0	17.3	TMI
TOTAL TEST	46.7	51.2	55.6	60.0	64.4	68.8	73.3	N
SCORE	49.5	52.9	56.2	59.6	63.0	66.3	69.7	EMI
	25.1	32.2	39.3	46.3	53.4	60.5	67.6	TMI

-4sd -3sd -2sd -1sd 0 +1sd +2sd +3sd +4sd

[a]This performance level falls above or below the mean by more than 4 standard deviations

Guidelines for Interpreting Student Profiles

1. Test item scores to the left of the line represent student weakness (or unique needs). Test item scores falling on the line or to the right represent student strengths.
2. Subtotal test scores to the left of the line represent skill area weaknesses. Subtotal test scores falling on the line or to the right represent skill area strengths.
3. Special education eligibility decisions can be made by establishing local criteria. An example of appropriate local special education criteria might be 2 standard deviations or more below the mean when compared to normal students' motor skills. It is recommended that teachers and parents use the subtotal and total test scores to make such eligibility decisions. If a student performs 2 or more standard deviations below the mean, specially designed instruction may be required.
4. Physical education placement decisions can be made by comparing a student's performance with the performance of other students in a particular class or setting. If any student, regardless of classification (assuming the student has acceptable learning characteristics), performs around the mean when compared to a

given group, then it appears that the proper placement would be with that group or class. Placement decisions can be made for units of instruction (e.g., locomotor skills, object control skills, or physical fitness skills). There is no reason why a student that performs poorly in physical fitness but around the mean for other units of instruction should be placed in a special or remedial class for all units. A student should be given as many opportunities as possible to participate in the regular education program.

5. Instructional programming decisions can be made if a local school district's physical education curriculum matches the objectives measured by this test. A precise instructional prescription can be developed based on the identification of skills that the student has not mastered. Duplication of instruction on skills that the student has already learned is not appropriate. Simply by looking at the test score sheet, a teacher can determine the student's present level of performance and select those skills that have not been mastered at the criterion level. By grouping those students together that have the same instructional needs, a teacher can plan more appropriate learning activities.

Reliability of the Test

An alpha coefficient of .92 was obtained on 140 students across all 16 test items, resulting in acceptance of inter-item homogeneity. A test-retest reliability coefficient of .97 was obtained when the test was administered on two separate occasions by three independent administrators.

Validity of the Test

A 97.6% agreement was obtained on content validity when reviewed by three independent content experts.

A perfect +1 item-objective congruency rating was obtained on all 16 test items when reviewed by three independent content experts.

Criterion-selection validity was obtained when three content experts rated the components selected for each test item on (1) their observability in the physical education class, and (2) their consistency with what research describes as a mature pattern.

APPENDIX G

CONTENT EXPERT RATINGS OF CRITERION VALIDITY

OBJECTIVE	COMPONENTS	CONSISTENT W/RESEARCH X[a]	CONSISTENT W/RESEARCH O[b]	OBSERVABLE X	OBSERVABLE O
1. RUN	a. A period when both feet are off the floor	3		3	
	b. Toe-heel or heel-toe foot contact	3		3	
	c. Arms move in opposition to legs, elbows bent	3		3	
	d. Smooth pattern for 50 feet	3		3	
2. GALLOP	a. Brief period where both feet are off the floor	3		3	
	b. Trailing foot does not cross in front of lead foot at floor contact	3		3	
	c. Arms swing forward and upward	3		3	
	d. 3 consecutive gallop strides	3		3	
3. HOP	a. Carriage of nonsupport leg is slightly bent	3		3	
	b. Maintain upright body position, elbows bent	3		3	
	c. Arms swing forward and upward	3		3	
	d. 3 consecutive hops on each foot	3		3	
4. SKIP	a. A step-hop pattern on alternate feet	3		3	
	b. Arms move in opposition to legs and are slightly bent	3		3	
	c. 3 consecutive skipping cycles (step-hop)	3		3	

[a]Denotes a positive rating
[b]Denotes a negative rating

OBJECTIVE	COMPONENTS	CONSISTENT W/RESEARCH		OBSERVABLE	
		X[a]	O[b]	X	O
5. HORIZONTAL JUMP	a. Two feet take off and a two foot landing	3		3	
	b. Arm thrust during take-off with full extension of legs	3		3	
	c. 3 consecutive jumps of 2/3 of standing height or more	2	1	3	
6. SLIDE	a. Weight transfer from trailing foot to lead foot along a straight line to the side	3		3	
	b. Period where both feet are off the ground and remain parallel	3		3	
	c. 3 consecutive slides to each side	3		3	
7. LEAP	a. Take off on one foot and a balanced landing on the other foot	3		3	
	b. Forward reach with arm opposite lead foot	3		3	
	c. 3 consecutive leaps	3		3	
8. OVERHAND THROW	a. A downward arc of the throwing arm to initiate overhand throw	3		3	
	b. Hip and spine rotation	3		3	
	c. Step forward with foot opposite the throwing arm	3		3	
	d. Follow through well beyond release	3		3	
	e. Ball travels 30 feet or more	2	1	3	
	f. 3 consecutive throws	3		3	
9. CATCH	a. Hands in front of body, elbows flexed near sides	3		3	
	b. Extension of arms in preparation for ball contact	3		3	
	c. Contact ball with hands only	3		3	

OBJECTIVE	COMPONENTS	CONSISTENT W/RESEARCH X^a	O^b	OBSERV-ABLE X	O
	d. 3 consecutive catches	3		3	
10. STATION-ARY BOUNCE	a. Contact ball between thighs and waist	2	1	3	
	b. Push ball with fingers of one hand only	3		3	
	c. Maintain a stable, stationary, position	3		3	
	d. 3 consecutive bounces	3		3	
11. KICK	a. A preliminary forward leap on the nonkicking leg with foot placement next to the ball	3		3	
	b. A continuous kicking motion	3		3	
	c. Contact ball so it travels forward at least 30 feet	3		3	
	d. 3 consecutive kicks	3		3	
12. TWO-HAND STRIKE	a. Side orientation toward desired direction of travel	3		3	
	b. Hip and spine rotation	3		3	
	c. Weight transfer from back foot to front foot during swing	3		3	
	d. 3 consecutive strikes	3		3	
13. SIT-UPS	a. Lie on back, knees bent, feet flat on floor, arms crossed over chest	3		3	
	b. Curl up to touch elbows to thighs	3		3	
	c. Return to lying position	3		3	
	d. 3 consecutive sit-ups	3		3	
14. SIT & REACH	a. Assume a sitting position with legs together, knees flat	3		3	
	b. Bend and reach forward to feet with one hand on top of the other, keeping knees straight	3		3	

OBJECTIVE	COMPONENTS	CONSISTENT W/RESEARCH X[a]	O[b]	OBSERVABLE X	O
	c. Hold position for 3 seconds	3		3	
	d. 3 consecutive reaches	3		3	
15. PUSH-UPS	a. Assume a prone position parallel to floor, toes on floor, hands directly under shoulders	3		3	
	b. Keep body parallel to floor while lowering body to 1-3 inches above floor	2	1	3	
	c. Raise body to starting position	3		3	
	d. 3 consecutive push-ups	3		3	
16. RUN/WALK	a. Run or walk continuously	3		3	
	b. 5 consecutive minutes	2	1	3	

REFERENCE NOTES

Reference Notes

Ebel, R. L. Personal communication. Michigan State University, 1980.

Mehrens, W. Personal Communication. Michigan State University, 1980.

Vogel, P. G. Personal communication. Michigan State University, 1980.

Yelon, S. L. Personal communication. Michigan State University, 1980.

BIBLIOGRAPHY

BIBLIOGRAPHY

American Alliance for Health, Physical Education, and Recreation. Joint Committee position paper on physical fitness, 1977.

American Alliance for Health, Physical Education, and Recreation. Special fitness test manual for mildly retarded persons. Washington, D.C., 1976.

Anderson, R. C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 42, 145-170.

Bagnato, S. J., Laub, K. W., & Kurtz, P. D. A descriptive manual and comprehensive process of linking screening, assessment, and programming for the handicapped preschooler. University Park, Pennsylvania: The Pennsylvania State University, 1978 (ERIC Document Reproduction Service No. ED 171 734).

Bailey, D. B. A comparison of non-biased screening procedures to identify high-risk kindergarten children. Unpublished doctoral dissertation, University of Washington, 1979.

Bailey, D. B., & Harbin, G. L. Nondiscriminatory evaluation. Exceptional Children, 1980, 46, 590-596.

Bormuth, J. R. On the theory of achievement test items. Chicago: University of Chicago Press, 1970.

Broer, M. Efficiency of human movement (3rd ed.). Philadelphia: W. B. Saunders Company, 1973.

Carpenter, A. Tests of motor educability for the first three grades. Child Development, 1940, 1, 37-61.

Carpenter, A. The differential measurement of speed in primary school children. Child Development, 1941, 12, 64-71.

Carpenter, A. The measurement of general motor capacity and general motor ability in the first three grades. Research Quarterly, 1942, 13, 444-465.

Clark, H. H., & Shelley, M. R. Maturity structure strength motor ability and intelligence test profiles of outstanding elementary and junior high school athletes. Physical Educator, 1961, 18, 132.

Cohens, J. A. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960, 20, 37-46.

Council for Exceptional Children. Policy regarding nondiscriminatory evaluation. Exceptional Children, 43(6), March 1977.

Cox, R. C., & Graham, G. T. The development of a sequentially scaled achievement test. Journal of Educational Measurement, 1966, 3, 147-150.

Cratty, B. J. Remedial motor activity for children. Philadelphia: Lea and Febiger, 1976.

Cratty, B. J. Perceptual and motor development in infants and children. Englewood Cliffs, New Jersey: Prentice-Hall, 1979.

Cronbach, L. B. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

Cumbee, F. Z. A factorial analysis of motor coordination. Research Quarterly, 1957, 25, 412-420.

Dauer, V. P. Dynamic physical education for elementary school children (4th ed.). Minneapolis: Burgess Publishing Company, 1971

Devault, M. V., Kriewall, T. E., Buchanan, A. E., & Quilling, M. R. Teacher's manual: Computer management for individualized instruction in mathematics and reading. Madison, Wisconsin: Research and Development Center for Cognitive Learning, University of Wisconsin, 1969.

Duffy, J. B., & Fedner, M. L. Educational diagnosis with instructional use. Exceptional Children, 1978, 44, 246-251.

Ebel, R. L. Essentials of educational measurement. Englewood Cliffs, New Jersey: Prentice-Hall, 1979.

Espenschade, A., & Eckert, H. Motor development. Columbus: Charles E. Merrill Publishing Company, 1967.

Fait, H. F. Special physical education: Adapted, corrective, developmental. Philadelphia: W. B. Saunders Co., 1972.

Federal Register, Volume 42, No. 163, Tuesday, August 23, 1977, 42480-42496.

Flanagan, J. C. Functional education for the seventies. Phi Delta Kappan, 1967, 49, 27-32.

Flanagan, J. C. Program for learning in accordance with needs. Psychology in the Schools, 1969, 6, 133-136.

Fuller, D. A. Evaluating motor performance of trainable mentally handicapped boys and girls, ages eight through twenty-one. Unpublished doctoral dissertation, Michigan State University, 1973.

Gallagher, J. D. Motor learning characteristics of low-skilled college men. Research Quarterly, 1970, 41, 59-67.

Gibbons, M. What is individualized instruction? Interchange, 1970, 1, 28-52.

Glaser, R. Instructional technology and the measurement of learning outcomes. American Psychologist, 1963, 18, 519-521.

Glaser, R. Adapting the elementary school curriculum to individual performance. In Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, N.J.: Educational Testing Service, 1968.

Glaser, R. Evaluation of instruction and changing educational models. In M. C. Wittrock & D. E. Wiley (Eds.), The evaluation of instruction. New York: Holt, Rinehart and Winston, 1970.

Glaser, R., & Nitko, A. J. Measurement in learning and instruction. In R. L. Thorndike (Ed.), Educational Measurement. Washington: American Council on Education, 1971.

Glassow, R., & Krause, P. Motor performance of girls age 6 to 14 years. Research Quarterly, 1942, 13, 444-465.

Godfrey, B., & Kephart, N. Movement patterns and motor education. New York: Appleton-Century-Crofts, 1969.

Govatos, L. Relationships and age differences in growth measures and motor skill. Child Development, 1959, 30, 333-340.

Gronlund, N. E. Individualizing classroom instruction. New York: Macmillan Publishing Co., 1974

Hambleton, R. K., & Eignor, D. R. Guidelines for evaluating criterion-referenced tests and test manuals. Journal of Educational Measurement, 1978, 15, 321-327.

Hambleton, R. K., & Gifford, J. A. Development and use of criterion-referenced tests to evaluate program effectiveness. (Report No. 52. Amherst, Mass.: School of Education, Laboratory of Psychometric and Evaluation Research, University of Massachusetts, 1977).

Hambleton, R. K. & Novick, M. R. Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 1973, 10, 159-170.

Hambleton, R. K., Swaminathan, H., Algina, J., & Coulson, D. B. Criterion-referenced testing and measurement: A review of technical issues and developments. Review of Educational Research, 1978, 48, 1-47.

Harris, M. L., & Stewart, D. M. Application of classical strategies to criterion-referenced test construction. A paper presented at the annual meeting of the American Educational Research Association, 1971.

Haubenstricker, J., Henn, J., & Seefeldt, V. Developmental stages of hopping. Unpublished paper, Michigan State University, 1975.

Hayden, F. Physical fitness for the mentally retarded. Ontario, Canada: Metropolitan Toronto Association for Retarded Children, 1964.

Heathers, G. Overview of innovations in organization for learning. Interchange, 1972, 3, 47-68.

Hellebrandt, F., Lawrence, G., Glassow, R., & Carns, M. Physiological analysis of basic motor skills. American Journal of Physical Medicine, 1961, 40, 14-25.

Henry, R. M. Coordination and motor learning. Proceedings College Physical Education Association, 1956, 68-75.

Herkowitz, J. Developmental task analysis: The design of movement experiences and evaluation of motor development status. In M. V. Ridenour (Ed.), Motor development: Issues and applications. Princeton, New Jersey: Princeton Book Co., 1978.

Hively, E., Maxwell, G., Rabehl, G., Senison, D., & Lundin, S. Domain-referenced curriculum evaluation: Minnemast Project. CSE monograph series in evaluation. No. 1. Los Angeles: Center for the Study of Evaluation, University of California, 1973.

Hofmeister, A. Trends in testing the handicapped. Utah State University, Exceptional Child Center, October, 1974.

Hofmeister, A. Integrating criterion-referenced testing and instruction. In W. Hively & M. Reynolds (Eds.), Domain-referenced testing in special education. Minneapolis: Leadership Training Institute/Special Education, University of Minnesota, 1975.

Hunsicker, P., & Reiff, G. AAHPER youth fitness test manual. Washington, D.C.: American Alliance for Health, Physical Education, and Recreation, 1976.

Ivens, S. H. An investigation of item analysis, reliability, and validity in relation to criterion-referenced tests. Unpublished doctoral dissertation, Florida State University, 1970.

Johnson, L., & Londeree, B. Motor fitness testing manual for the moderately mentally retarded. Washington, D.C.: American Alliance for Health, Physical Education, and Recreation, 1976.

Johnson, R. Measurements of achievement in fundamental skills of elementary school children. Research Quarterly, 1962, 32, 94-103

Kane, R. J., & Meredith, N. V. Ability in the standing broad jump of elementary school children 7, 9, and 11 years of age. Research Quarterly, 1953, 23, 198-208.

Keesling, J. W. Identification of differing intended outcomes and their implications for evaluation. Paper presented at the Annual Meeting, American Educational Research Association, Washington, D.C., 1974.

Kosecoff, J., & Fink, A. The appropriateness of criterion-referenced tests for evaluation studies. (TM Report 60) Princeton, New Jersey: Educational Testing Service, 1976.

Kosecoff, J., Fink, A., & Klein, S. P. A system for describing and evaluating criterion-referenced tests. (TM Report 57) Princeton, New Jersey: Educational Testing Service, 1976.

Kraus, H., & Hirschland, R. Minimum muscular fitness tests in school children. Research Quarterly, 1954, 25, 178-188.

Kriewall, T. E. Applications of information theory and acceptance sampling principles to the management of mathematics instruction. Unpublished doctoral dissertation, University of Wisconsin, 1969.

Latchaw, M. Measuring selected motor skills in fourth, fifth, and sixth grades. Research Quarterly, 1954, 25, 439-449.

Latchaw, M., & Egstrom, G. Human movement. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1969.

Livingston, S. A. Criterion-referenced applications of classical test theory. Journal of Educational Measurement, 1972, 9, 13-26.

Loovis, E. M. Model for individualizing physical education experiences for the preschool moderately retarded child. Unpublished doctoral dissertation, The Ohio State University, 1975.

Loovis, E. M., & Ersing, W. F. Assessing and programming gross motor development for children. Columbus: The Ohio State University, 1975.

Mann, L. Psychometric phrenology and the new faculty psychology: The case against ability assessment and training. Journal of Special Education, 1971, 5, 3-14.

Martuza, V. R. Applying norm referenced and criterion-referenced measurement in education. Boston: Allyn and Bacon, 1977.

Mehrens, W. A., & Lehmann, I. J. Measurement and evaluation in education and psychology. New York: Holt, Rinehart and Winston, 1978.

Millman, J. Criterion-referenced measurement. In W. J. Popham (Ed.), Evaluation in education: Current applications. Berkeley, California: McCutchan Publishing Company, 1974.

Milne, C. Fundamental motor skills sequences. Unpublished paper, University of Western Ontario, London, Ontario, 1972.

Moyer, J. E. A comparison of the Kuder-Richardson formula and kappa as estimates of the reliability of criterion-referenced tests. Unpublished doctoral dissertation, Western Michigan University, 1976.

Moyer, J. E. Alternative reliability indices. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York, 1977.

Nester, G. M. Development and evaluation of visual descriptor models for the assessment of selected fundamental locomotor skills in trainable mentally impaired children. Unpublished doctoral dissertation, Michigan State University, 1977.

Nineth-fourth Congress, Federal Public Law 94-142. Congressional Record, November 29, 1975, 12, 773-796.

Oakland, T., & Matuszek, P. Using tests in nondiscriminatory assessment. In T. Oakland (Ed.), Psychological and educational assessment of minority children. New York: Brunner/Mazel, 1977.

Orpet, R. E. Frostig movement skills test battery. Palo Alto, California: Consulting Psychologists Press, Inc., 1972.

Osborn, H. G. Item sampling for achievement testing. Educational and Psychological Measurements, 1968, 28, 95-104.

Popham, W. J. Educational evaluation. Englewood cliffs, New Jersey: Prentice-Hall, Inc., 1975.

Popham, W. J. Normative data for criterion-referenced tests? Phi Delta Kappan, 1976, 58, 593-594.

Popham, W. J. Criterion-referenced measurement. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1978(a).

Popham, W. J. Well-crafted criterion-referenced tests. Educational Leadership, 1978, 36, 91-95 (b).

Popham, W. J., & Baker, E. L. Writing tests which measure objectives. The Prentice-Hall Teacher Competency System. Englewood Cliffs: Prentice-Hall, 1973.

Popham, W. J., & Husek, T. R. Implications of criterion-referenced measurement. Journal of Educational Measurement. 1969, 6, 1-9.

Rarick, G. L. Research in physical education: Issues and concerns. Unpublished paper presented at the Annual Meeting of the National Consortium on Physical Education and Recreation for the Handicapped, Rochester, New York, 1979.

Rarick, G. L., & Dobbins, D. A. Basic components in the motor performance of educable mentally retarded children: Implications for curriculum development. Department of Physical Education, University of California, Berkeley, 1972.

Rarick, G. L., & Dobbins, D. A. Basic components in the motor performance of children six to nine years of age. Medicine and Science in Sports, 1975, 7, 105-110.

Rarick, G. L., & McQuillan, J. The factor structure of motor abilities of trainable mentally retarded children: Implications for curriculum development. Final Report USOE-BEH, 1977.

Reynolds, M. C., & Birch, J. W. Teaching exceptional children in all America's schools: A first course for teachers and principals. Reston, Virginia: Council for Exceptional Children, 1977.

Rovinelli, R. J., & Hambleton, R. K. On the use of content specialists in the assessment of criterion-referenced test item validity. Dutch Journal for Educational Research, 1977, 2, 49-60.

Safrit, M. J. Evaluation in physical education. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1981.

Salvia, J., & Yesseldyke, J. E. Assessment in special and remedial education. Boston: Houghton Mifflin Co., 1978.

Sapp, M. M. Developmental sequences of galloping: Preliminary stages. Unpublished paper, Michigan State University, 1980.

Schurr, E. Movement experiences for children. New York: Appleton-Century-Crofts, 1967.

Seefeldt, V. Developmental sequence of catching. Unpublished paper, Michigan State University, 1976(a).

Seefeldt, V. Developmental sequence of running. Unpublished paper, Michigan State University, 1976(b).

Seefeldt, V. Developmental sequence of the standing long jump. Unpublished paper, Michigan State University, 1976(c).

Seefeldt, V., & Haubenstricker, J. Developmental sequence in skipping. Unpublished paper, Michigan State University, 1974.

Seefeldt, V., & Haubenstricker, J. Developmental sequence of kicking. Unpublished paper, Michigan State University, 1975.

Seefeldt, V., & Haubenstricker, J. Developmental sequence of striking. Unpublished paper, Michigan State University, 1976(a).

Seefeldt, V., & Haubenstricker, J. Developmental sequence of throwing. Unpublished paper, Michigan State University, 1976(b).

Seefeldt, V., Reuschlein, S., & Vogel, P. Sequencing motor skills within the physical education curriculum. Report to the national convention of American Alliance for Health, Physical Education, and Recreation, Houston, March, 1972.

Sharkey, B. J. Fitness and work capacity testing. Missoula, Montana: Forest Service, U.S. Department of Agriculture, 1977.

Sharkey, B. J. Physiology of fitness. Champaign, Illinois: Human Kinetics Publishers, 1979.

Sinclair, C. Movement and movement patterns of early childhood. Division of Educational Research and Statistics, State Department of Education, Richmond, Virginia, June 1973.

Strasler, G. M., & Raeth, P. G. An internal consistency estimate for criterion-referenced tests. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York, 1977.

Swaminathan, H., Hambleton, R. K., & Algina, J. Reliability of criterion-referenced tests: A decision-theoretic formulation. Journal of Educational Measurement, 1974, 11, 263-268.

Taylor, E. Achievement scales in physical education skills for children in grades I, II, III. The Elementary School Journal, May, 1941, 677-682.

Torgerson, W. S. Theory and methods of scaling. New York: Wiley, 1958.

Vandenberg, S. G. Factor analytic studies of the Lincoln-Oseretsky test of motor proficiency. Perceptual and Motor Skills, 1964, 19, 23-41.

Vogel, P. G. Diagnostic-prescriptive teaching: A comparison of norm and criterion-referenced approaches. Unpublished paper presented at the Annual Meeting of the American Alliance for Health, Physical Education, and Recreation, Seattle, 1977.

Walker, C. B. Standards for evaluating criterion-referenced tests. Los Angeles: California University, Center for the Study of Evaluation, 1978. (ERIC Document Reproduction Service No. ED 179 595).

Wessel, J. A. I CAN physical education curriculum. Northbrook, Illinois: Hubbard, 1979.

Wessel, J. A. I CAN preprimary motor and play skills program. East Lansing, Michigan: Marketing Division, Instructional Media Center, Michigan State University, 1980.

Wickstrom, R. L. Fundamental motor patterns. Philadelphia: Lea and Febiger, 1977.

Wild, M. R. The behavior pattern of throwing and some observations concerning its course of development in children. Research Quarterly, 1938, 9, 20-24.

Wilson, H. A. A humanistic approach to criterion-referenced testing. Paper presented at the Annual Meeting, American Educational Research Association, New Orleans, 1973.

Yesseldyke, J. E. Diagnostic-prescriptive teaching: The search for aptitude-treatment interactions. In L. Mann & D. A. Sagatina (Eds.), The first review of special education. Philadelphia: JSE Press, 1973.

Yesseldyke, J. E., & Salvia, J. Diagnostic-prescriptive teaching: Two models. Exceptional Children, 1974, 41, 181-185.