

FURTADO-GALLAGHER MOVEMENT SKILL ASSESSMENT: COLLECTING EVIDENCE
FOR CONTENT-RELATED VALIDITY.

by

Ovande Furtado Junior

BS, Federal University of Parana, 1997

Submitted to the Graduate Faculty of

School of Education in partial fulfillment

of the requirements for the degree of

Masters of Science

Health, Physical, and Recreation Education

University of Pittsburgh

2004

UNIVERSITY OF PITTSBURGH
FACULTY OF SCHOOL OF EDUCATION

This thesis was presented

by

Ovande Furtado Jr.

It was defended on

November 19, 2004

and approved by

Jere Dee Gallagher, PhD, Associate Professor

Louis Pingel, PhD, Associate Professor

Mary Duquin, PhD, Associate Professor

FURTADO-GALLAGHER MOVEMENT SKILL ASSESSMENT: COLLECTING EVIDENCE FOR CONTENT-RELATED VALIDITY.

Ovande Furtado Jr.

University of Pittsburgh, 2004

The National Standards for Physical Education emphasizes the importance of linking instruction and assessment in physical education (NASPE, 2004). This link is even more beneficial to students and teachers alike when assessment is aligned with the NASPE content standards. Therefore, the purpose of this study was to develop a criterion-referenced assessment tool to test movement concepts and fundamental movement skills of children ages 3 through 10 and provide evidence for content-related validity. An attempt was made to develop an assessment tool that is tied closely to the postulates of the National Standards for Physical Education (NASPE, 2004). Content-related validity was collected through judgments from 20 content experts (professors = 8; and PE teachers = 12). An Internet-based item review form was used for data collection. Experts rated content at item level using a four point likert-type scale where 1 being 'not important at all' and 4 'very important'. Content was also rated at test level addressing the match between the test content and the test general characteristics. Questions were rated on a five point likert-type scale with 1 being 'poor' and 5 being 'excellent'. Descriptive statistics (percentage of responses and median) were used along with qualitative procedures for data analysis. The analysis of content at the item level yielded revision of 11 items. Also, three items were considered not appropriate for either 3-6 or 7-10 age groups. Four items were included based on experts written comments. Finally, two items were dropped out of the test. The analysis of the content at the test level showed that questions were rated as 'very good' or 'excellent' by 80% or

more of judges. The information collected from content experts served to validate and confirm the content of this assessment tool while providing helpful feedback with which to make improvements in content and format.

Results of this study provided an initial support for the content-related validity of the Furtado-Gallagher Movement Skill Assessment. Content validity is an ongoing process throughout the development of any instrument and further analysis will be needed to confirm the content validity of this assessment tool.

TABLE OF CONTENTS

PREFACE.....	viii
1.0 INTRODUCTION AND LITERATURE REVIEW	1
1.1 The Alignment of Assessment and Instruction in Physical Education.....	1
1.2 Domain Specifications	3
1.2.1 Movement Concepts	3
1.2.2 Fundamental Movement Skills	5
1.3 Test Development	7
1.3.1 Collecting Content-Related Validity.....	10
1.3.1.1 Definitions.....	10
1.3.1.2 Approaches for collecting content-related validity evidence.....	12
2.0 METHODOLOGY	15
2.1 Instrumentation	15
2.1.1 Item Pool Generation	15
2.1.2 Criteria for item selection	16
2.1.3 Item generation for movement concepts.....	17
2.1.4 Item Generation for Fundamental Movement Skills	21
2.1.5 Table of specifications	24
2.1.6 Item Content Review Form.....	25
2.1.6.1 Part A – Guidelines and Background Information	26
2.1.6.2 Part B – Item Level and Test Level Ratings.....	27
2.1.7 Pilot Study.....	28
2.2 Participants.....	29
2.2.1 Procedures for contacting the panel of judges	30
2.3 Data Analysis.....	32
3.0 RESULTS	33
3.1 Explanation of Group Differences	33
3.2 Content evidence at the item level.....	34
3.2.1 Behavioral Objective 1: Locomotor Skills	35
3.2.2 Behavioral Objective 2: Manipulative Skills.....	36
3.2.3 Behavioral Objective 3: Stability Skills.....	37
3.2.4 Behavioral Objective 4: Body Awareness	38
3.2.5 Behavioral Objective 5: Space Awareness	40
3.2.6 Behavioral Objective 6: Effort Awareness	41
3.2.7 Behavioral objective 7: Relationship Awareness	41
3.3 Content evidence at the item level.....	42
4.0 DISCUSSION	45
4.1 Content-related validity at the item level.....	45
4.2 Content-related validity at the test level	49
4.3 Suggested Test Format.....	50
4.3.1 Fundamental Movement Skill.....	51
4.3.2 Movement concepts	54
4.4 Conclusion	57

REFERENCES	58
APPENDIX A	60
APPENDIX B	61
APPENDIX C	77
APPENDIX D	78
APPENDIX E	81
APPENDIX F	83
APPENDIX G	87
5.0 INTRODUCTION	87
6.0 LITERATURE REVIEW	92
6.1 Assessment tools available for testing FMS	92
6.1.1 The Developmental Sequence of Motor Skill Inventory (DSFMSI)	92
6.1.2 The Ohio State University Scale of Intra-Gross Motor Assessment (SIGMA)....	93
6.1.3 Test of Gross Motor Development (TGMD)	94
6.1.4 The Fundamental Movement Pattern Assessment Instrument (FMPAI).....	95
6.2 The “No Child Left Behind” act and its tie with assessment.....	97
6.3 Domain specifications.....	99
6.3.1 Learning movement concepts	100
6.3.1.1 Cognitive vs. functional understanding of movement concepts	101
6.3.1.2 Combination with other movement concepts and skills	102
6.3.1.3 Body Awareness: What the body can do	104
6.3.1.4 Space awareness: where the body moves	105
6.3.1.5 Effort awareness: How the body can move	108
6.3.1.6 Relationship awareness: moving in relation to objects and people	110
6.3.2 Fundamental Movement Skills	111
6.3.2.1 Fundamental Stability Movement Skills.....	113
6.3.2.2 Fundamental Locomotor Movement Skills	113
6.3.2.3 Fundamental Manipulative Movement Skills.....	114
6.3.3 The importance of the fundamental movement skills.....	114
6.3.4 Movement concepts, fundamental movement skills and the Standards	116
6.4 Assessment.....	118
6.4.1 Why do physical education teachers assess?	119
6.4.2 The link between assessment and instruction	121
6.4.3 Formal/ Standardized and informal /non-standardized tests.....	122
6.4.4 Norm-Referenced Approach vs. Domain-Referenced Approach	124
6.4.5 Intended purposes and interpretations	125
6.4.6 Scores and their interpretation	126
6.4.7 Beyond norm-referenced approach.....	127
6.4.8 Approaches used to test, describe and interpret movement changes	129
6.4.9 Product-Oriented Assessment:.....	129
6.4.10 Process-Oriented Assessment	130
6.4.11 Total Body Approach (TBA) vs. Body Component Approach (BCA)	131
6.5 Test Validity.....	133
6.5.1 Validity evidence based on test content.....	134
6.5.2 Measuring content-related evidence through expert judgment.....	135
REFERENCES	139

LIST OF TABLES

Table 1: Terms used in the item selection criteria	17
Table 2: Items for the body awareness subdomain	18
Table 3: Items for the space awareness subdomain	19
Table 4: Items for the effort awareness subdomain	20
Table 5: Items for the relationship awareness subdomain	21
Table 6: Items for the locomotor subdomain	22
Table 7: Items for the manipulative subdomain	23
Table 8: Items for the stability subdomain	24
Table 9: Table of specifications containing all proposed for analysis.....	25
Table 10: Table showing number of years of teaching PE	31
Table 11: Table showing the level's of education for both teachers and professors	31
Table 12: Percentage of responses and median values for locomotor skills.....	36
Table 13: Percentage of responses and median values for manipulative skills	37
Table 14: Percentage of responses and median values for stability skills	38
Table 15: Percentage of responses and median values for the Body Awareness subdomain.....	39
Table 16: Percentage of responses and median values for the Space Awareness subdomain.....	40
Table 17: Percentage of responses and median values for the Effort Awareness subdomain.....	41
Table 18: Percentage of responses and median values for the Relationship Awareness subdomain	42
Table 19: Percentage of responses and median values for evidence validity at the test level.....	44
Table 20: Final list of items for fundamental movement skills	52
Table 21: Developmental sequence for the low beam wall (Gallahue and Donnelly, 2003)	53
Table 22: Final list of items for movement concepts.....	54
Table 23: Most common assessment tools to assess FMS.....	89
Table 24: Selected fundamental movement skills.	113
Table 25: National Standards for Physical Education	117

PREFACE

First thanks must go to my advisor Dr. Jere D. Gallagher, whose constant support, patience and guidance led me through the end of this project. Also warmest thanks must go to my parents, Geny Ferreira Furtado and Ovande Furtado, whose support, although from distance, allowed me to write this thesis.

My committee members, Jere D. Gallaher, Louis Pingel, and Mary Duquin, likewise assisted greatly in this undertaking, showing me places where I had strayed from the straight path and gently leading me back in the proper direction. For all their guidance, both with the mechanics of the written text and with the implications that lay therein, I wish to express my sincerest appreciation.

Also, I would like to thank Marilyn Ross and Louis Fabian for supporting me with the financial needs that allowed me to study in the U.S.

Finally, I must thank my co-researchers Oldemar Mazzardo, Fabio Fontana and Comfort Mokgothu at the Motor Behavior Laboratory for their assistance throughout the development of my thesis.

1.0 INTRODUCTION AND LITERATURE REVIEW

Movement is part of everyone's life. The interaction with the world in which we live occurs through movement. There is some evidence that motor skills in childhood may develop through informal play (NASPE, 2004). However, this is becoming less probable in today's society. Children today have fewer opportunities to develop skills through play. Thus, instruction in physical education can serve as a way to "help students develop the knowledge, attitudes, motor skills, behavioral skills, and confidence need to adopt and maintain physically active lifestyles" (National Center for Chronic Disease Prevention and Health Promotion, 1997, p. 205). A key component of instruction is assessment. Without assessment the teacher does not know whether children improve their movement.

The purpose of this research is to develop a criterion-referenced assessment tool that will test movement concepts and fundamental movement skills of children ages 3 through 10 years. In order to develop a test, several components must first be covered: assessment's alignment with instruction, existing movement tests, domains to be tested, and parameters of test construction including reliability and validity. Each of these items are reviewed in order.

1.1 THE ALIGNMENT OF ASSESSMENT AND INSTRUCTION IN PHYSICAL EDUCATION

Assessment in physical education is important for the improvement of instructional practices and students' movement. The justification of the link between assessment and instruction can be

found in two documents: No Child Left Behind and the National Standards for Physical Education. Throughout the “No Child Left Behind” act (2002) the issue of linking assessment tools and instructional practices is emphasized. The act urges the development of coherent assessment tools so that student progress can be tracked from year to year, and that assessment results can be more easily shared with parents and students. The second document also emphasizes the importance of linking instruction and assessment in physical education (NASPE, 2004). The document reflects what students should know and be able to do as a result of a quality physical education program (NASPE, 2004). The NASPE document goes further stating that students and teachers may benefit substantially when assessment is aligned with the NASPE content standards.

Of particular importance for the development of this test are standards 1 and 2. Standard 1 refers to the development of skills needed to enjoy participation in physical activities. These skills are widely known as fundamental movement skills and they comprise one of the two domains tested by the assessment tool that is being proposed in the current study. Standard 2 relates to the second domain being tested here, which is the domain of movement concepts.

Can the tests that are currently used for assessing quality of movement be used to improve individual performance and instruction? Four of the most commonly used tests for assessing process motor skills are briefly addressed next.

Existing movement test batteries are the Developmental Sequence of Motor Skill Development (Seefeldt & Haubenstricker, 1976), the Ohio State University Scale of Intra-Gross Motor Skill (Loovis & Ersing, 1979), the Test of Gross Motor Development (Ulrich, 2000), and The Fundamental Movement Pattern Assessment Instrument (McClenaghan & Gallahue, 1978). Each of these instruments has limitations in testing FMS. For example, none of these instruments

addresses the issue of movement concepts and only one test covers stability skills, both of which are important constructs to be assessed in physical education.

Therefore, this assessment tool is being developed to fill the gaps of other assessment tools, as well as to reflect the national standards for physical education. This will help in the process of linking assessment of motor skills to instructional practices. The items that form the current test come from two distinct domains: movement concepts and fundamental movement skills.

1.2 DOMAIN SPECIFICATIONS

The process of gathering evidence for content-related validity requires a discussion of the domain(s) being measured under the assessment being developed. The items that form the current test come from two distinct domains: (1) movement concepts, and (2) fundamental movement skill.

1.2.1 Movement Concepts

Movement concept learning relates to how the body *can* move (Gallahue & Donnelly, 2003). Participation in sports or game-like activities is dynamic in nature and can be quite demanding. However, one cannot teach or assess every possible variation of a movement used in a given task. Instead, students are required to learn and be able to demonstrate their cognitive and functional understanding of movement concepts that can be applied to various situations (Graham, Holt/Hale, & Parker, 2004). Therefore, teaching and testing movement concepts in physical education are important.

Taking into consideration the importance of movement concepts when learning, it is surprising that there has not been an attempt to develop and validate an assessment tool to test such concepts in physical education. A few checklists are available in some physical education textbooks (Gallahue & Donnelly, 2003; Graham et al., 2004; Kogut & National Association for Sport and Physical Education., 2003).

Movement concepts are commonly divided into four categories (behavioral objectives¹): body awareness, space awareness, effort awareness, and relationship awareness. Body awareness refers to ‘what the body can do’. Examples of items for *body awareness* are body shapes and nonlocomotor actions.

Space awareness is the second category of movement concepts that deals with ‘where the body moves’. Children who understand and apply (functional understanding) the concept of space awareness are more likely to move safely as they travel through the environment (Graham et al., 2004). Examples of items from this category are location (self and general space) and direction (left-right, forward-backward, up-down, etc.).

The third category that falls under the movement concept domain is *effort awareness*. This category includes ‘how the body can move’. Examples of items for this category are force (strong-light), time (fast-slow), and flow (bound-free). Understanding the concept of effort awareness is crucial because while performing movement skills, movers must apply the appropriate amount and degree of time, force and flow to complete the task successfully.

The last category under the movement concept domain is *relationship awareness*. This concept deals with both “*how* and *where*” the body moves in relation to objects and people. The developers of the current test subdivided this category into two measurement dimensions. These

¹ Behavioral objective is a term that describes specific areas within a domain having particular attributes (skills) that measure its content. This term will be used interchangeably with the term “subdomain” throughout this research.

are *Relationship with Objects and/or People* and *Relationship with People Only*. Examples of relationship with objects and/or people are over-under, in front-behind, meeting-parting, etc, whereas relationship with people only could be mirroring-matching, leading-following, etc. It is important to notice that the concept of relationship awareness gives meaning to the interaction between individuals and their environment (Graham et al., 2004).

Therefore, testing movement concepts is important because in real-life situations individuals have to perform in an environment that is unpredictable and full of variables. Thus, having a functional understanding of movement concepts should facilitate the relationship of the mover and the environment. Next, the second domain that is investigated in the current assessment tool is addressed.

1.2.2 Fundamental Movement Skills

Fundamental movement skills (FMS) are defined as “common motor activities with specific patterns. They are the general skills that form the bases for the more advanced and more specific motor activities, such as sport skills” (Wickstrom, 1983, p. 7). Paralleling movement concepts, FMS are divided into categories or subdomains. A category of movement, in the case of FMS, is a classification scheme based on common underlying principles of movement (Gallahue & Ozmun, 1998). Thus, fundamental movement skills that share the same characteristics are placed in the same category of movement. For example, the fundamental movement skills of catching and striking are placed in the movement category of manipulation. Both skills involve the manipulation of an object. Researchers agree in subdividing the domain of fundamental movement skill into three categories: stability, locomotor, and manipulation. Once the rationale

for dividing fundamental movement skills into different categories is provided specific information concerning each one category can be addressed.

Stability skills form one the categories that fall under FMS. They are believed to form the basis for all locomotor and manipulative skill since all movement involves an element of stability (Gallahue & Donnelly, 2003). This category can be subdivided even further into static balance, dynamic balance and non-locomotor movements. Although some text books in physical education place this last subcategory into the body awareness subdomain of movement concepts. Examples of stability skills are dodging, walking on the beam balance, tripod and one-foot balance. These skills require dynamic or static balance in which the main goal is to gain or maintain one's equilibrium against the force of gravity (Gallahue & Donnelly, 2003). According to Gallahue and Ozmun (1998), it is important that individuals master fundamental aspects of stability before efficient forms of locomotion occur. Surprisingly, only a few assessment tools have stability skills as part of their testing battery.

Locomotor movements form the second category that falls under FMS. They consist of any skill in which the body is transported in a horizontal or vertical direction from one point to another (Graham et al., 2004). Locomotor skills are sometimes subdivided into basic and combinations. Basic refers to skills that lack combination of elements, whereas combination skills have the combination of 2 or more elements embedded in its form. An example will help to clarify. The skill of running would fall under the 'basic' subcategory, whereas sliding under the combination category. This is because running consists of one single action that happens continuously. On the other hand, sliding requires the combination of a sideways hop. Examples of other fundamental movement skills are leaping, horizontal jump, hopping, galloping, and skipping.

Finally, manipulative skills form the third category of FMS. They refer to any skill in which the individual interacts with some kind of object during its performance (Gallahue & Donnelly, 2003). Manipulation occurs either upon giving (propulsion) or receiving (absorption) force to or from objects. Throwing, kicking, and striking are some examples of propulsive manipulative skills, whereas catching and trapping are examples of absorptive skills.

In summary, the movement concept and the fundamental movement skill domains will be included in the current test. The next section addresses the procedure required for test development.

1.3 TEST DEVELOPMENT

Test development involves a few steps that must be followed if the goal is to develop a test that: (1) actually measure what it is proposing to measure, and (2) is consistent over time, over occasion and over raters. The first item relates to collection of evidence for establishing validity, while the second item deals with reliability of test scores and/or test interpretations. Validity evidence based on test content is gathered prior to reliability evidence. However, three steps must be followed before evidence for content-related validity can be collected. First, the general purpose and the intended uses of the test must be delineated. Second, how test results will be used and interpreted must be defined. This will lead to the third step. The selection of the approach the test will be based on must be determined. These three steps are discussed in detail next. Then validity is addressed followed by a discussion of content-related validity which is the scope of this research.

The first step in test development is the definition of the test's general purpose and intended uses. The current assessment is being developed to be used by physical education teachers working with preschool and early elementary students. In these settings, assessments are employed for many reasons, and the use of information obtained from assessment varies. For many physical education teachers assessment exists with the exclusive purpose of assigning grades to students at the end of a unit/theme (Holt/Hale, 1999). However, assessments are conducted for other purposes as well. Physical education teachers can use assessment for *screening*, to determine the *status* of students, and for *placement* reasons. Assessments can also be used for *program evaluation*, analysis of *program content*, and even as a *motivational factor*. The second purpose of this test is to improve instruction.

Once the purpose of the test is defined the next step is to choose an approach in which the test development will be based. At least in educational and performance test development the most common approaches are 'norm-referenced' (NRA) and criterion-referenced' (CRA).

Tests developed under the norm-referenced approach allow for comparisons of an individual's performance to the performance of a normative group, whereas in criterion-referenced tests the performance is compared to some predetermined criterion (Burton & Miller, 1998). Because criterion-referenced tests are more linked to instruction, they are primarily used to evaluate progress, for evaluations of programs, and to plan instruction (Nitko, 2001). Embedded in norm-referenced test is the capability of determination of an individual's standing relative to his or her peers, this is most often useful for screening, placement and evaluating programs. Depending on the purpose of the test being developed one approach is more appropriate than the other. But simply establishing the purpose and the intended uses of a test is

not sufficient for deciding the most appropriate approach to select. One needs to know how test results will be used and interpreted.

Tests developed under NRA or CRA differ in terms of their purposes and thus allow for different types of interpretation of student results. There are some occasions in which teachers want to compare students' performances (between-individual comparison). For example, if a teacher wants to know whether Oldemar is better than Fabio on the performance of the fundamental movement skill of kicking or a group of skills (e.g., manipulative skills), a test developed under the NRA is preferred. In addition to between-individual comparisons one might want to interpret assessment results in relation to the performance of a large group of similar students (norm group) – similar age, sex, etc., - who happened to take the same assessment. Here is another example that illustrates this situation. Maggie has received a percentile rank score of 34 on the total test score of the TGMD (Ulrich, 2000). This means that she performed as well or better than 34% of the students in the normative group. This type of information can be useful for deciding whether or not a student needs remedial assistance or is a candidate for a gifted program (Nitko, 2001). However this score gives little information about Maggie's deficiencies that she needs to improve in the performance of gross motor skills. If one wants to know more about Maggie's performance, then a test developed under the CRA is preferred.

On the other hand, if the purpose is to collect information for a more complete interpretation of student's scores, then one needs more than just percentile ranks or other types of relative scores which are the core of norm-referenced tests. When there is a need for absolute interpretation of students' assessment results, criterion-referenced tests are preferred. A good example is when a teacher wants to know which of the fundamental movement skills students are having trouble mastering. Rank order scores may be helpful selecting the best performers in

locomotive movement skills; however, if the ultimate goal is to plan appropriate instruction, or track students improvement over a period of time one needs to be aware of the kinds of performances each student can do or the specific kinds of difficulties each student is experiencing in each skill tested (Nitko, 2001). Thus, a test that allows a teacher to infer about one's status in relation to the domain of fundamental movement skills, rather than simply in relation to how well others students performed, is necessary.

In short, when developing a test, attention must be given to establishing the general purpose and the intended uses of the test before any addition step is taken. The types of score interpretations must be delineated beforehand. These two pieces of information are necessary for the selection of the approach (NRA or CRA) under which the test is to be developed.

1.3.1 Collecting Content-Related Validity

The definition of the test's purpose and intended uses, the statement of how results will be interpreted, and the selection of the test approach, are all necessary steps prior actual collection of data/evidence for content-related validity. The next step is to determine test validity.

1.3.1.1 Definitions

“Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (AERA, APA, & NCME, 1999, p. 9). In the past, test developers emphasized the existence of three types of validity: (1) content validity; (2) criterion-related validity and (3) construct validity. Further, other terms have been used including concurrent validity, face validity, predictive validity, etc. However, in general, these are specific procedures for evaluating validity, rather than new kinds of interpretative inferences for test scores (Messick, 1989). Currently, the Standards for Educational and Psychological Testing

(AERA et al., 1999) considers validity a unitary concept. Thus, there are sources of validity that support the intended interpretation of test scores for the proposed purpose of the test. They must be viewed interdependent and complementary forms of validity evidence and not viewed as separate and substitutable types of validity (Rudner & Schafer, 2002). These sources of validity evidence are based on: (1) test content; (2) response processes; (3) internal structure; (3) relation to other variables; and (5) consequences of testing. Of special interest for this research is the source of validity evidence based on test content.

Validity based on test content is one of these sources of validity. In content-related validity the question is whether the items proposed by the test developer do, in fact, constitute a representative sample of the wider domain about which one wishes to make inferences (Thorn & Deitz, 1989). In addition, items are judge based on their importance (Anastasi & Urbina, 1997). Importance/relevance focuses on whether assessment tasks or items are included in the test user's domain which is defined by the test developer (Nitko, 2001). The process of collecting evidence for content-related validity is an essential requirement of test development that begins with the purpose of the test being clearly stated (Haley, Coster, & Faas, 2001). Next, the content domain being measured is fully described (Thorn & Deitz, 1989). Concurrently with the content domain analysis is the development of the behavioral objectives (Berk, 1984). Each objective should reflect a specific area of the entire domain. Specific literature is then researched and a pool of items is generated. Here, all major aspects of the domains are tapped in the correct proportion and items assigned to one of the behavioral objectives previously developed. The next question is what are the approaches used to collect evidence based on test content.

1.3.1.2 Approaches for collecting content-related validity evidence

Because the process used to generate the item pool is not sufficient to provide content-related evidence, further item analysis is necessary. Thorn and Deitz (1989) cite two approaches to establish evidence for content-related validity once the items have been generated. The first approach uses empirical methods such as factor analysis and the use of statistics to measure item difficulty and discrimination. Perhaps the major problem of such an approach is that it can result in a final selection of items that are not representative of the domain being measure by the test (Hambleton et al., 1978 cited in Thorn & Deitz, 1989). Therefore, an approach based on judgment of content experts is usually preferred. In this approach, the quality of the items is judged by the extent to which each reflects, in terms of its content, the domain it intends to measure (Thorn & Deitz, 1989). This judgment is done by at least two content experts who rate the relevance and representativeness of items. Collecting evidence for content-related validity through judgments from content experts are mainly based on content representativeness and relevance which focus on whether the assessment tasks are a representative and/or relevant sample from a larger domain of performance (Nitko, 2001). The development of an item review form (IRF) and selection of a panel of content experts are two necessary procedures when using this approach.

The IRF is intended to collect information from content experts regarding quality of the set of items proposed by developers of the test (Nitko, 2001). Some aspects that may be included in the IRF are: (a) item importance/relevance; (b) item representativeness; (c) link between group of items and specific behavioral objectives (subdomains); and (d) link between a set of items and the purpose(s) of a test; etc.

When developing an IRF test developers use a layout that best fits their needs. Generally, content experts are asked to rate items using ordinal scales (e.g., Likert-type scales). Also, YES or NO-type questions are often included in these forms. Further, content experts are asked to

provide written comments regarding items or a set of items (Haley et al., 2001). Written comments provided by raters are especially important in deciding whether a particular item should be included in the final list of items that will composed the final version of the test. Moreover, it is acceptable to ask content experts whether they feel it is necessary to include and/or remove items to/from the proposed list of items. Procedures commonly used to select the panel of judges are addressed next.

The process of selecting experts that rate the quality of proposed items is a critical step in collecting evidence for content-related validity. To rate the quality of items in their study, Haley et al. (2001) used a panel of judges that combined experts with theoretical/research-related backgrounds and experts with more practical/applied experience. This might be the best approach when selecting content experts. This combination may produce a more diversified set of data than if only a select category of experts was used. This approach can yield a broader set of opinions that may better help test developers in the interpretation of data. Also, selecting experts that are familiar with the test should be avoided. This might cause potential judgment bias, thus jeopardizing the final selection of items (Haley et al., 2001). Further it is recommended that experts rate items independently. The interaction of experts rating a set of items may also cause problems (Thorn & Deitz, 1989). Some judges may be influenced by the opinion of another judge while rating the quality of items, thus causing potential bias.

These procedures solely do not guarantee experts will rate test items with quality. The use of incentives may help motivate experts to complete the IRF (Thorn & Deitz, 1989). The authors recommended listing the names of the experts in the test manual. Experts may work harder when they are acknowledged for their input. Additionally, test developers may offer money in exchange to experts' voluntary participation.

In short, test validity is an important step in test development; as well as, the process of gathering evidence for content-related validity. The use of a panel of content experts is the most common procedure for collecting such evidence. The collection of data is usually done through item review forms that are comprised of rating questions that assess the quality of the proposed set of items.

This brief review addressed several issues regarding test construction and the steps necessary prior collection of content-related validity. First, four instrument used to measure children's motor skill development were addressed along with their majors shortcomings. Specifically, it was pointed out the failure of such instruments cover items related to movement concepts and stability skills. The need to develop instruments in physical education that test process characteristics of movement skills that are aligned with instructional practices was also emphasized. Then, issues regarding the two domains were addressed. The domains of Movement concept and fundamental movement skill were briefly described. After the domains were described issued regarding test development were be addressed. Finally, the major issues in collecting evidence based on test content were discussed. Next, the methodology used in the current research is addressed.

2.0 METHODOLOGY

2.1 INSTRUMENTATION

2.1.1 Item Pool Generation

The intended uses (purposes) of a test are commonly defined at the beginning of the development of a test. The three intended uses for the current test are to: (1) monitor individual progress during and/or following instruction, (2) evaluate effectiveness of the instructional program with the intent of adjusting the curriculum in accordance with student needs, and (3) to detect eventual deficits in movement concept understanding, as well as, problems in fundamental skill development.

After the purpose of the test was stated and the content domain specified, the next step was to generate a pool of items that, presumably, would sample the domains under investigation. Item generation was done by a systematic analysis of the content domain. Sixteen books (see Appendix A) were selected for such analysis. These books were chosen based on their relevance to the field of physical education. A movement skill/concept matrix was generated first by determining the items and then by counting each time an item (fundamental movement skill or a movement concept) appeared in each of the sixteen books. The same pattern of categories and subcategories used in Chapter 1 to discuss both domains was used in the development of the matrix. These categories were: (1) fundamental movement skill domain: Stability (dynamic and static balance), locomotion (basic and combinations of skills), and manipulation (propulsive and absorptive); and (2) movement concept domain: body awareness (body shapes and non-

locomotor), space awareness (location, directions, levels pathways and extensions); effort awareness (time, force, flow and space), and relationship awareness (with objects and/or people and partners). The total number of appearances for each skill and its percentage in relation to the total number of books was provided in the matrix. This information was used to assist in the decision of which items would be selected for further analysis and constitute the item review list². Described next is the criteria used to select items that comprised the item review list.

2.1.2 Criteria for item selection

After each book was carefully analyzed and items entered into the movement skill/concept matrix, the next step was to develop a criterion to help in the process of generation of the item review list that would be further analyzed by content experts.

The criterion used to select the items was both quantitative and qualitative. First, each measurement dimension was treated individually by selecting all items that appeared in at least eight (50% or above) of the books researched. This is the quantitative aspect of the procedure. The qualitative aspect of the procedure used for item selection, relates to personal judgments by the developers of the current test. Personal judgments are important for item selection at this stage of test development, since all behaviors in each domain may not be relevant, important, or widely applicable. Thus, test items are also selected qualitatively on the basis of the test developer's value judgments (Nunnally, 1978 cited in Thorn & Deitz, 1989).

A total of 82 items were selected to compose the matrix. Forty-three of these items appeared on eight or more books. This constituted the quantitative procedure for item selection. However, not all of these items were included in the final item review list. A qualitative analysis

² List of items included in the Item Review Form that was further sent to content experts for content analysis.

was used that helped in the final decision of items to be included. The labels in [Table 1](#) were used to describe whether or not an item would be part of the item review list. After the qualitative analyses, a total of 31 items were selected to be included in the item review form for the content expert analysis. The procedures used for such selection is described next for each domain under investigation in the Furtado-Gallagher Movement Skill Assessment.

Table 1: Terms used in the item selection criteria

Labels	Explanation
NOT INCLUDED	Item scored 50% or above, but not included
DISCARDED	Item did not score 50% or above: left out of the item review list
KEPT	Scored 50% or above: item included in the item review list
INCLUDED	Item did not score 50% or above, but included

2.1.3 Item generation for movement concepts

The combination of the quantitative and qualitative procedures yielded 15 items (5 body awareness, 5 space awareness, 3 effort awareness, and 2 relationship awareness) that were included in the item review list under the domain of movement concepts. Each of these items are covered next.

Of nine items entered on the skill/concept matrix for the body awareness subdomain, five were selected to be part of the item review list (see [Table 2](#)). Four items (round, narrow, wide and twisted) were collapsed into a single item titled body shapes. The other four items that comprised the body awareness subdomain were symmetrical / nonsymmetrical, twisting, bending, and front swing.

Table 2: Items for the body awareness subdomain

Skills	Item information		
	Total	%	Final Status
Body Shapes <i>Narrow, Wide, Round, Twisted</i>	8	50%	KEPT
Symmetrical / Nonsymmetrical	9	56%	KEPT
Stretching / Extending	7	44%	DISCARDED
Twisting	11	69%	KEPT
Bending	10	63%	KEPT
Front Swaying	9	56%	KEPT
Twirling	5	31%	DISCARDED
Sinking	5	31%	DISCARDED
Push/pull	6	38%	DISCARDED

The list of items for the space awareness subdomain is shown in [Table 3](#). All items entered on the skill/concept matrix were selected for the item review list. Previously, this subdomain was separated into categories. These categories were location, directions, levels, pathways, and extensions. After conducting a pilot study these categories became items to be assessed under the subdomain of space awareness. For example, general space and self-space were both considered as separate items. After the pilot study was conducted these concepts were considered dimensions to be tested for the item called ‘location’. The same procedure was used for the subsequent movement concepts subdomains of effort awareness and relationship awareness. This was done because the original items (now dimensions) were too specific.

Table 3: Items for the space awareness subdomain

Skills	Item information		
	Total	%	Final Status
Location <i>General Space, Self Space</i>	12	75%	KEPT
Directions <i>Up / Down, Forward / Backward, Clockwise / Counterclockwise, Right / Left</i>	12	75%	KEPT
Levels <i>Low / Middle / High</i>	13	81%	KEPT
Pathways <i>Straight, Curved, Zigzag</i>	12	75%	KEPT
Extensions <i>Large / Small, Far / Near</i>	6	38%	INCLUDED

The items from the effort awareness subdomain are presented in [Table 4](#). As mentioned before, the items for the ‘effort awareness’ were also modified after a pilot study was conducted. Thus, ‘time’, ‘force’, ‘flow’, and ‘space’ were changed to items. Three of these items were selected to be part of the item review list. The item ‘space’ was discarded since it was found in only 2 out of the 16 books researched. The item refers to the direct and flexible use of space. The developers of the current test agreed that the skill of general space (under space awareness) would be sufficient to measure such attributes.

Table 4: Items for the effort awareness subdomain

Skills	Item information		
	Total	%	Final Status
Time <i>Fast / Slow</i>	12	75%	KEPT
Force <i>Strong / Light</i>	11	69%	KEPT
Flow <i>Bound / Free</i>	10	63%	KEPT
Space <i>Direct / Flexible</i>	2	13%	DISCARDED

Finally, items representing the subdomain of relationship awareness are displayed in [Table 5](#). Both items made the item review list. One item representing the relationship with objects and/or people along with a second item representing the relationship with people only were included. Next, the complete list of items selected to be part of the item review list is provided in a table of item specifications.

Table 5: Items for the relationship awareness subdomain

Skills	Item information		
	Total	%	Final Status
With Objects and/or People <i>Over / under, In front / Behind, Meeting / Parting, Around / Through / Along</i>	8	50%	KEPT
With People Only <i>Leading / Following, Mirroring / Matching</i>	9	56%	KEPT

2.1.4 Item Generation for Fundamental Movement Skills

After careful revision, a total of 16 items (Locomotor=6, Manipulative=7, and Stability=3) from the fundamental movement skill domain were included in the item review list to be analyzed by content experts. Qualitative analysis was employed and two items that scored 50% or above in the ‘Locomotor’ subdomain (see [Table 6](#)) were left out of the item review list (not included). Despite the fact that the skill of walking appeared in 11 out 16 books, it was also excluded from the item review list. Walking was considered too easy to perform to be included in the current test. Items that are either too easy or too difficult do not discriminate well and consequently tell test examiners little about students’ performance status on a given attribute (Nitko, 2001). The skill of sliding/sidestepping was also not included even though it was found in 12 out of 16 books. The developers of the current test argued that the skill of sliding has basically the same biomechanical features as the skill of galloping which was found in 15 books researched.

Table 6: Items for the locomotor subdomain

Skills	Item information		
	Total	%	Final Status
Basic (one element)			
Walking	11	69%	NOT INCLUDED
Running	15	94%	KEPT
Leaping	11	69%	KEPT
Horizontal jumping	14	88%	KEPT
Jumping from a height	4	25%	DISCARDED
Vertical jumping	7	44%	DISCARDED
Hopping	14	88%	KEPT
Combinations (2 + elements)			
Climbing	3	19%	DISCARDED
Galloping	15	94%	KEPT
Sliding/sidestepping	12	75%	NOT INCLUDED
Skipping	15	94%	KEPT

The list of items for the manipulative subdomain is shown in [Table 7](#). A total of seven items (5 propulsive, and 2 absorptive) are included in the item review list. In addition, one item (trapping) that did not score over 50% was included. This skill was included because it measures manipulative absorptive aspects in the lower body which is different from catching that involves upper body measures.

Table 7: Items for the manipulative subdomain

Skills	Item information		
	Total	%	Final Status
Propulsive			
Underhand ball rolling	5	31%	DISCARDED
Overhand Throwing	14	87%	KEPT
Underhand Throwing	3	18%	DISCARDED
Kicking (stationary ball)	14	87%	KEPT
Punting	10	62%	NOT INCLUDED
Sidearm Striking (moving ball)	7	43%	INCLUDED
Batting (moving ball)	8	50%	KEPT
Overhand striking	4	25%	DISCARDED
Underhand striking	4	25%	DISCARDED
Volleying	5	31%	DISCARDED
Jumping rope	2	12%	DISCARDED
Stationary dribbling	13	81%	KEPT
Vertical Throw	1	6%	DISCARDED
Absorptive			
Catching	15	93%	KEPT
Trapping	7	43%	INCLUDED
Collecting	1	6%	DISCARDED
Propelling	1	6%	DISCARDED
Carrying	1	6%	DISCARDED

The list of items for the stability subdomain is showed in [Table 8](#). Surprisingly, only three out 12 items comprising the subdomain of stability appeared in at least eight or more books researched. The skill of ‘forward roll’ was not included due to safety issues. Also, all three items belonged to the measurement dimension of dynamic balance and none from static balance. To compensate for this imbalance it was decided to include the skill of one-foot balance which appeared in 7 out 16 books in the item review list. Aspects related to the selection of items for

FMS was addressed under this section. Presented next is the discussion of how items for the movement concept domain were selected.

Table 8: Items for the stability subdomain

Skills	Item information		
	Total	%	Final Status
Dynamic Balance			
Walking on low beam	8	50%	KEPT
Walking on line	3	19%	DISCARDED
Forward roll	10	63%	NOT INCLUDED
Dodging	9	56%	KEPT
Backward roll	6	38%	DISCARDED
Fleeing	5	31%	DISCARDED
Chasing	5	31%	DISCARDED
Cartwheel	2	13%	DISCARDED
Static Balance			
One-Foot balance	7	44%	INCLUDED
Tripod	4	25%	DISCARDED
Elbow-Knee balance	2	13%	DISCARDED
Headstand	2	13%	DISCARDED

2.1.5 Table of specifications

It is common for test developers to define the domain assessed by making use of what is known as a *table of specifications*. This table is developed by including the major content categories and skills that are to be assessed. For the proposed assessment tool, the main content categories (subdomains) are (1) body awareness; (2) space awareness; (3) effort awareness; (4) relationship awareness; (5) fundamental locomotor skills; (6) fundamental manipulative skills, and (7) fundamental stability skills. Each one of these categories is subdivided into at least two other categories which are called measurement dimensions (Table 9).

Table 9: Table of specifications containing all proposed for analysis

Subdomains	Number of Items
1. Body Awareness	
A. Body Shapes	1
B. Nonlocomotor	4
	Subtotal = 5
2. Space Awareness	
A. Locations	1
B. Directions	1
C. Levels	1
D. Pathways	1
E. Extensions	1
	Subtotal = 5
3. Effort Awareness	
A. Time	1
B. Force	1
C. Flow	1
	Subtotal = 3
4. Relationship Awareness	
A. Obstacle and/or people	1
B. Partners	1
	Subtotal = 2
5. Basic Locomotor Skills	
A. Basic (one element)	4
B. Combinations (two elements)	2
	Subtotal = 6
6. Basic Manipulative Skills	
A. Propulsive	5
B. Absorptive	2
	Subtotal = 7
7. Stability/Balance Skills	
B. Dynamic postures	2
C. Static postures	1
	Subtotal = 3
	Total test items = 31

2.1.6 Item Content Review Form

Item review forms are used to collect information from experts who make judgments regarding test content representativeness and relevance (Nitko, 2001). The judgments are made at both the item and test level. At the item level, evidence is collected to verify the extent to which each item is a measure of the content domain; whereas, at the test level, information is gathered to verify

representatives of the total collection of test items as a measure of the content domain (Thorn & Deitz, 1989). The item review form (IRF) developed to collect evidence for content-related validity for the Furtado-Gallagher MSA was comprised of two parts (see Appendix B): Part A – Guidelines and Background Information and Part B – Item and Test Level Ratings.

2.1.6.1 Part A – Guidelines and Background Information

Part A of the form included two distinct sections: Section A1 – Form Guidelines and Section A2 – Expert’s Background Information. It was decided to have experts to go through an introductory section (A1) that served to familiarize them with the format of the IRF. This section did not intend to cover every single aspect of the form, but provide important information to guide reviewers in the process of rating the items. Section A1 included an example of one item of each domain and their respective performance criteria (FMS) or measurable characteristics (MC). It also provided information on technical issues such as the system required for filling in the form online and Internet browser helpful hints. Experts were warned not to consult anyone to help in the process of rating the items. Information regarding compensation for participation was provided. Finally, the overall purpose of the test and its three intended uses were provided at the end of the section. Before clicking to continue onto Section A2 experts were asked to print out Section A1 since its information could be helpful during item rating. Section A1 was available after experts clicked ‘next’ on the bottom of the form.

Section A2 collected demographic information from experts. Email, number of years teaching physical education, and area of expertise were some of the things asked in this section. After completing section A2, the reviewers continued with Part B.

2.1.6.2 Part B – Item Level and Test Level Ratings

Part B of the item review form was also divided into two sections: Section B1 – Item Level Ratings (relevance / importance), and Section B2 – Test Level Ratings (match between the set of items and the test’s general characteristics).

Section B1 included a list of items that depicted both the fundamental movement skill and movement concept domains. These items were grouped into seven behavioral objectives: Locomotor, Manipulation, Stability (FMS), and Body Awareness, Space Awareness, Effort Awareness, and Relationship Awareness (MC). Items were rated on a likert-type scale from 1 to 4, with 4 being ‘very important’ and 1 ‘not important at all’. Item importance/relevance was rated separately for both 3-6 and 7-10-year age groups. Further, following each group of items for each behavioral objective, respondents were asked whether they felt any additional items related to that behavioral objective needed to be included. Finally, respondents were encouraged to provide any additional comments on each behavioral objective. In section B1, evidence for content-related validity was collected at the item level. This section intended to collect information regarding item level validity. Test level content validity was addressed in section B2.

Section B2 dealt with the representativeness of the total collection of test items as a measure of the content domain. Content experts were asked to rate the quality of the set of items on six separate questions. Questions 1 and 2 asked whether there is a match between the main purpose of the test and the set of items proposed for both fundamental movement skills and movement concept domains. In question 3, experts are asked to rate the extent to which the proposed set of items represent skills/concepts that are frequently taught in physical education. The appropriateness of the set of items for the two domains to measure separate but complimentary aspects of gross motor skills of children ages 3-10 was addressed in question 4.

Finally, questions 5, 6, and 7 asked the potential of the proposed set of items to measure the three intended uses of the current test.

The item review form was developed so that information could be collected over the World Wide Web. This was done to avoid mailing paper-based forms to participants and for ease of measurement. It was expected that collecting information over the Internet would expedite the process of data collection. The Internet-based item review form can be advantageous compared to the paper-based form in the sense that more interactive content is allowed with the former. For example, respondents were able to see a pictorial illustration of most items to be rated. This was done using the feature of pop-up windows which is used to save visual space in webpage materials. Including a picture of each item on a paper-based form, for example, would make the form too 'busy' since written information and pictures would be displayed together. Also, having experts submit information online allowed for tracking how much time was used to rate each section of the form. The form was developed with the application computer package Microsoft FrontPage version 2002. The forms were then hosted on an Apache Linux server with 99.99% uptime. The item review form discussed here went through several modifications since it was first developed. Presented next is the description of the pilot study that was conducted with the intent to detect potential problems that could result in systematic errors in data collection.

2.1.7 Pilot Study

A pilot study was conducted prior to the actual collection of data to look for potential systematic problems with the Internet-based item review form. Six experts in the field of motor development participated in the pilot study. Experts were asked to complete the form online and provide their feedback. A feedback form (see Appendix B) was prompted once experts finished

filling in the form. Experts were asked the degree to which the form was user-friendly. A likert scale (from strongly agree to strongly disagree) was used to collect ratings from experts. Four reviewers responded “somewhat agree” and two responded “strongly agree” when presented with the following statement: *I think the Item Review Form is user-friendly and content experts will not have major problems to complete it.* Also, experts were asked to provide their feedback on both Part A and Part B of the form. They were asked to comment on each section in terms of the amount of information provided, wording, and if the information was presented in a clear way. Finally, experts were given the chance to make general comments about the form in a separate input field. All experts made constructive comments that helped further improvements of the document. Several modifications were made including wording clarification, and item categories rearrangement.

2.2 PARTICIPANTS

Judgments from content experts were used as part of the process to collect evidence for content-related validity. First, a list of 29 potential participants was developed. This list was comprised of a combination of research-oriented experts (theoretical experience) - 11 and school teachers (practical experience) - 18. Research-oriented experts were selected based on their experience and level of publication in the field of motor development. School teachers were chosen based on whether they received some type of recognition (local or national) for teaching physical education in pre- and elementary school. The list of physical education teachers who had received teaching excellence awards was taken from the American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD) website.

2.2.1 Procedures for contacting the panel of judges

The procedure used to contact experts is described next. First, each potential content expert was contacted by phone. This first contact was used to ask their willingness to voluntarily participate as a content expert in the current research. The content of the phone call was standardized so that all potential participants would receive the same information. Once the expert agreed to participate, he/she was asked to provide a current email address. Also, content experts were notified that they would receive a compensation of \$50.00 for their voluntary participation upon completion of the study. Further, they were told that their names would be listed in the user's test manual. It was assumed that these two incentives would motivate content experts to rate the items. Following the phone call an email message was sent to the expert to confirm the contact. This was done specifically to ensure that the respondent did not have the sender's email address blocked by automatic email filters. Once the respondents replied to the email confirmation, a second email message was sent which contained information on how to access the forms over the Internet. An example of the content of this second email message is provided in Appendix C.

The forms were developed so that the information submitted by content experts would be sent to a secure database protected by an encrypted password. This was done to ensure the anonymity of respondents. Also, no names were linked to sections B1 and B2 of the form. Respondents were given identification numbers included in the content of the second email message.

A total of twenty participants (12= teachers, 8= professors) successfully completed the forms over the Internet (see [Table 10](#)). Of the 18 teachers from the original list, 1 teacher refused to participate upon contact, 2 did not respond to the second email, and 3 could not be contacted by phone. From the original list of 11 professors, 2 refused to participate, and 1 could not be

contacted either by phone or email. The mean teaching experience for PE for teachers was 19.33 years (SD= 8.07 years), whereas for professors was 28.13 years (SD= 10.21 years).

Table 10: Table showing number of years of teaching PE

Expertise	N	Mean	SD	Minimum	Maximum
Teachers	12	19.33	8.07	11	38
Professors	8	28.13	10.21	10	42

Information was also collected regarding participant’s level of education (see [Table 11](#)). Eleven out of 12 teachers had masters’ degree and 1 completed bachelor’s degree in physical education. All eight professors had doctoral degrees (either PhD or EdD). In addition, all professors have had several publications in the area of motor skill development including articles in research journals, book chapters and text books used in physical education settings. All the teachers that participated as content experts of the current study have received local and/or excellence teaching awards (see Appendix D for a list of awards).

Table 11: Table showing the level’s of education for both teachers and professors

			Highest education			
			BS	Master	Ph.D	Total
Expertise	Teachers	Count	1	11	0	12
		% within Expertise	8.3%	91.7%	.0%	100.0%
	Professors	Count	0	0	8	8
		% within Expertise	.0%	.0%	100.0%	100.0%

2.3 DATA ANALYSIS

The computer package SPSS for Windows version 12.0 was used for data analysis. Percentage of responses and the median for each question in sections 1 and 2 of Part B of the form were computed. A Pearson Chi-Square test was performed to verify how different teachers and professors rated the items. It was decided not to establish cut scores prior to data collection for section 1 of Part B which concerned collection of evidence for content-related validity at the item level. Here, specific questions were asked for single items. The decision whether or not an item would be included in the test relied heavily on the test developers' judgment combined with the data collected from the content experts. It was decided to use the percentage of 67% (2/3 of the overall responses) for considering an item for revision or exclusion of the final list. For section 2 of Part B the developers of the test were interested to see if the set of items, not single items, would match with the general characteristics of the test. Thus, the percentage of responses for each question was analyzed and if responses were high (over 75 percent) then it could be said that evidence was collected to support initial content validity of the test (Haley et al., 2001).

3.0 RESULTS

The results of the content-related validity for the FG-MSA are organized by the type of content evidence collected through experts' judgments. These are content evidence collected at the item and the test level. The results of the content validity collected at the item level are presented for each behavioral objective. At the test level, results are presented within a single section. Percentage of responses and median values are presented for each item (item level evidence validity) and each question (test level evidence validity). It should be noted that some items were modified based on the information collected at the item level. The details of how each item is modified are presented in the discussion section. Before content validity results for both item and test level are presented, there is a need to address a separate analysis to verify the extent to which respondents for two distinct groups would differ while rating the items.

3.1 EXPLANATION OF GROUP DIFFERENCES

The content of the current test was analyzed by 20 experts. An attempt was made to have experts from different backgrounds rate the content at both the item and test level. Information was collected from college/university professors and physical education teachers. Professors are believed to have more of a theoretical experience, whereas teachers have more of practical experience with regards to the content of this assessment. Having two distinct groups rate the items can represent a problem if respondents from each group rate items differently. To account for this potential problem Pearson Chi-Square test was performed to verify whether professors and teachers differ in rating items. The idea is that if significant difference exists, then there is

reason to analyze and discuss the results separately for professors and teachers. The result of the analysis (see Appendix E for complete tests) demonstrated that only the skill of 'Batting' (chi-square= 9.831, df= 3, p= .020) yielded significance. That is, for all items, except for Batting, professors and teachers were consistent across the 4 point scale used in the form. Since only one out of thirty-one items yielded significance the developers of the current test speculate this to be a result of a type I error. The same analysis was calculated for Section 2 of Part B of the form which had to do with collection of evidence for content-related validity at the test level. The Pearson Chi-Square analysis yielded no differences between professors and teachers in their ratings. Therefore, based on these results ratings from professors and teachers are collapsed.

3.2 CONTENT EVIDENCE AT THE ITEM LEVEL

Next, evidence for content-related validity is presented at the item level. The following results are presented grouped by behavioral objectives. This is done to follow the same pattern in which information was presented to content experts at the time of data collection. Further, a general rule was used to identify items for possible exclusion or revision. Based on this rule, items that received a total of 67% (2/3 of responses) or less when adding 3 and 4 of the 4-point likert-type scale were considered for revision or exclusion. In addition, a percentage of 40% or less on a given item for option 4 would also 'flag' that item for possible revision or exclusion.

3.2.1 Behavioral Objective 1: Locomotor Skills

Table 12 shows the percentage of responses and median values of all ‘Locomotor’ items. With the exception of ‘Leaping’ for 3-6 year-olds, and ‘Gallop’, all items that comprised the ‘Locomotor’ subdomain were included in the test for both age groups. ‘Leaping’ was removed from the younger group because only 65% of the respondents rated this item as either ‘very important’ or ‘moderately important’. ‘Gallop’ was removed for both age groups. Gallop was replaced by the skills of sliding (sideways) and included for both age groups. Sliding was recommended by six experts. ‘Jumping from height’ was also included under the locomotor subdomain. Two other skills (walking, and vertical jumping) were also suggested by experts. However, they were not included in the final list of items.

Table 12: Percentage of responses and median values for locomotor skills

Items	Age	Percentages				Median	Inclusion
		(Not important at all)		(Very important)			
		1	2	3	4		
Running	3-6	0.0	0.0	25.00	75.00	4	Y
	7-10	0.0	20.0	30.0	50.0	3.5	Y
Horizontal Jumping	3-6	0.0	15.0	5.0	80.0	4	Y
	7-10	0.0	15.0	20.0	65.0	4	Y
Leaping	3-6	5.0	30.0	25.0	40.0	3	N
	7-10	0.0	20.0	35.0	45.0	3	Y
Hopping	3-6	0.0	5.0	15.0	80.0	4	Y
	7-10	0.0	15.0	30.0	55.0	4	Y
Galloping	3-6	0.0	10.0	20.0	70.0	4	N
	7-10	0.0	35.0	20.0	45.0	3	N
Skipping	3-6	5.0	10.0	10.0	75.0	4	Y
	7-10	0.0	25.0	15.0	60.0	4	Y

Y= item included; N= not included

3.2.2 Behavioral Objective 2: Manipulative Skills

The analysis of the responses for the manipulative skills both from ratings and written comments yielded nine items to be included in the test. Five of these items were revised (see [Table 13](#)). In addition, the skill of ‘Batting’ was not included in the test for the younger group. Further, ‘Approach Kicking’ (7-10), ‘Side-Arm Striking’ (3-6), ‘Hand Dribbling’ (7-10), ‘Catching’ (7-10), and ‘Trapping’ (3-6 and 7-10) were all revised to fit age group requirements. In addition to the modifications done on these items, two other items were included under the manipulative skill category. These were the skills of ‘Underhand Throwing’ and ‘Soccer Dribbling’. The skill

of ‘Underhand Throwing’ is included for both age groups, while ‘Soccer Dribbling’ is limited to the 7-10 year-olds.

Table 13: Percentage of responses and median values for manipulative skills

Items	Age	Percentages				Median	Inclusion
		(Not important at all)		(Very important)			
		1	2	3	4		
Overhand Throwing	3-6	5.0	5.0	35.0	55.0	4	Y
	7-10	0.0	0.0	15.0	85.0	4	Y
Approach Kicking (stationary ball)	3-6	0.0	5.0	20.0	75.0	4	Y
	7-10	0.0	20.0	35.0	45.0	3	YR
Side-arm Striking (moving ball)	3-6	25.0	30.0	20.0	25.0	2	YR
	7-10	0.0	10.0	25.0	65.0	4	Y
Batting (moving ball)	3-6	35.0	30.0	15.0	20.0	2	N
	7-10	0.0	0.0	25.0	75.0	4	Y
Hand Dribbling (stationary)	3-6	5.0	20.0	20.0	55.0	4	Y
	7-10	0.0	15.0	20.0	65.0	4	YR
Catching (stationary)	3-6	0.0	20.0	20.0	60.0	4	Y
	7-10	0.0	5.0	15.0	80.0	4	YR
Trapping (stationary ball)	3-6	15.0	15.0	30.0	40.0	3	YR
	7-10	5.0	5.0	40.0	50.0	3.5	YR

Y= item included; N= not included; YR= included with revision

3.2.3 Behavioral Objective 3: Stability Skills

All three items proposed for the subdomain of ‘Stability Skills’ were included in the test (see [Table 14](#)). The skill of ‘Walk on Low Beam’ was modified for the 7-10 year-olds. Although only 40.0% of respondents rated ‘Dodging’ as ‘very important’, it was included without any modification. In addition, the skills of ‘Hand-and-Knee balance and ‘Single-Knee balance were

included in both groups. These are stability skills in which different body parts are used for balance and similar to foot balance they are important to be tested.

Table 14: Percentage of responses and median values for stability skills

Items	Age	Percentages				Median	Inclusion
		(Not important at all)		(Very important)			
		1	2	3	4		
Walk on low beam	3-6	0.0	15.0	10.0	75.0	4	Y
	7-10	0.0	20.0	25.0	55.0	4	YR
Dodging	3-6	5.0	25.0	30.0	40.0	3	Y
	7-10	0.0	5.0	20.0	75.0	4	Y
One-Foot balance	3-6	5.0	10.0	20.0	65.0	4	Y
	7-10	0.0	25.0	30.0	45.0	3	Y

Y= item included; YR= included with revision

3.2.4 Behavioral Objective 4: Body Awareness

Items from the ‘Body Awareness’ subdomain received the lowest ratings from experts (see [Table 15](#)). Based on written comments, it was decided to accommodate the items from this subdomain into two different subdomains. First, ‘Symmetrical / Nonsymmetrical’ was found to be too specific to be considered as an item. Hence, it was moved under ‘Body Shapes’. There, ‘Symmetrical / Nonsymmetrical’ is used as a dimension to be measured along with the three existing dimensions of narrow-wide, round (curved), and twisted (see [Table 15](#)). In addition, the item ‘Body Shapes’ was moved under the ‘Relationship Awareness’ subdomain (Behavioral Objective # 7). In order to fit within that subdomain, ‘Body Awareness’ was renamed to ‘Relationship of body parts’. Because 40% of respondents reported either ‘not important at all’ or ‘somewhat important’ for using the item of ‘Body Shapes’ to test 7-10 year-olds, it is tested in

the younger group only. ‘Symmetrical/Nonsymmetrical’ was modified to fit to the younger group. In addition, Both ‘Body Bending’ and ‘Body Twisting’ were also found too specific to be used as separate items. Thus, they collapsed to form a new item called ‘Non-Locomotor’. This new item was placed under the ‘Stability’ subdomain (Behavioral Objective # 2). Because ‘Body Twisting’ and ‘Body Bending’ was considered inappropriate for the older group (see ratings on the Table 15), ‘Non-Locomotor’ is tested in the younger group only. Finally, the item “Front Swing” was not included in the test. Only 65% and 45% of respondents judged this item as either, ‘very important’ or ‘moderately important’ to be tested among 3-6 and 7-10 year-olds, respectively.

Table 15: Percentage of responses and median values for the Body Awareness subdomain

Items	Age	Percentages				Median	Inclusion
		(Not important at all)		(Very important)			
		1	2	3	4		
Body shapes Narrow-Wide, Round (Curved), and Twisted	3-6	0.0	0.0	10.0	90.0	4	Y
	7-10	5.0	35.0	20.0	40.0	3	N
Symmetrical / Nonsymmetrical	3-6	15.0	25.0	25.0	35.0	3	N
	7-10	0.0	15.0	30.0	55.0	4	YR
Body Twisting	3-6	0.0	10.0	25.0	65.0	4	YR
	7-10	5.0	30.0	35.0	30.0	3	N
Body Bending	3-6	0.0	0.0	30.0	70.0	4	YR
	7-10	5.0	40.0	20.0	35.0	3	N
Front Swing	3-6	15.0	20.0	35.0	30.0	3	N
	7-10	20.0	35.0	20.0	25.0	2	N

Y= item included; N= not included; YR= included with revision

3.2.5 Behavioral Objective 5: Space Awareness

Unlike the preceding subdomains, items from the ‘Space Awareness’ movement concept category received considerably high ratings from experts (see Table 16). All items from the ‘Space Awareness’ subdomain are included in the test. On average, 92% of the respondents judged the items as either ‘very important’ or ‘moderately important’ for the age group of 3-6 and 83% for the older group.

Table 16: Percentage of responses and median values for the Space Awareness subdomain

Items	Age	Percentages				Median	Inclusion
		(Not important at all)		(Very important)			
		1	2	3	4		
SA Location	3-6	0.0	5.0	0.0	95.0	4	Y
	7-10	0.0	15.0	25.0	60.0	4	Y
SA Directions	3-6	0.0	5.0	20.0	75.0	4	Y
	7-10	0.0	15.0	25.0	60.0	4	Y
SA Levels	3-6	0.0	5.0	10.0	85.0	4	Y
	7-10	5.0	15.0	25.0	55.0	4	Y
SA Pathways	3-6	0.0	0.0	30.0	70.0	4	Y
	7-10	15.0	0.0	30.0	55.0	4	Y
SA Extensions	3-6	5.0	20.0	15.0	60.0	4	Y
	7-10	5.0	15.0	15.0	65.0	4	Y

Y= item included

3.2.6 Behavioral Objective 6: Effort Awareness

Two of the three items proposed for the concept of ‘Effort Awareness’ were included in the test (see Table 17). The items included were ‘Time’ and ‘Force’. The item ‘Flow’ was not included. Only 55% of respondents judged ‘Flow’ as either ‘very important’ or ‘moderately important’ to be included in the test for the younger group. Written comments from respondents revealed that testing ‘Flow’ can be very difficult to be accomplished for either age group. In addition, the concept of rhythm was included.

Table 17: Percentage of responses and median values for the Effort Awareness subdomain

Items	Age	Percentages				Median	Inclusion
		(Not important at all)		(Very important)			
		1	2	3	4		
EA Time	3-6	0.0	5.0	5.0	90.0	4	Y
	7-10	5.0	20.0	10.0	65.0	4	Y
EA Force	3-6	5.0	20.0	5.0	70.0	4	Y
	7-10	5.0	20.0	10.0	65.0	4	Y
EA Flow	3-6	15.0	30.0	5.0	50.0	3.5	N
	7-10	0.0	15.0	40.0	45.0	3	N

Y= item included; N= not included

3.2.7 Behavioral objective 7: Relationship Awareness

Items under the ‘Relationship Awareness’ subdomain also received high ratings from experts (see Table 18). On average, 90% of respondents judged items as either ‘very important’ or ‘moderately important’ for the younger group and 82.5% for 7-10 year-olds. Recall that due to the elimination of the ‘Body Awareness’ subdomain, ‘Body Shapes’ was moved to the current

subdomain and renamed to ‘Relationship of body parts’. Thus, behavioral objective # 7 has now a total of three items. Two items were originally proposed for this category and one item that was taken from the extinct ‘Body Awareness’ subdomain.

Table 18: Percentage of responses and median values for the Relationship Awareness subdomain

Items	Age	Percentages				Median	Inclusion
		(Not important at all)		(Very important)			
		1	2	3	4		
RA with objects and/or people	3-6	0.0	5.0	25.0	70.0	4	Y
	7-10	0.0	20.0	15.0	65.0	4	Y
RA with people only	3-6	0.0	15.0	25.0	60.0	4	Y
	7-10	0.0	15.0	25.0	60.0	4	Y

Y= item included

3.3 CONTENT EVIDENCE AT THE ITEM LEVEL

Results for validity at the item level were presented above. Results for validity at the test level are reviewed next. Two major content validity questions concerns the match between the main purpose of the set of items for the Fundamental Movement Skill domain (question 1) and Movement concept domain (question 2). Eighty-five percent of respondents on either question judged the match as either ‘very good’ or ‘excellent’ (see [Table 19](#)). Question three dealt with the extent to which the set of items proposed for the FG-MSA constituted items that are frequently taught to children in preschool and early elementary grades. Again, 85% of participants judged the representativeness as either ‘very good’ or ‘excellent’ on this major content validity issue.

The appropriateness of the set of items was asked on question 4. Eighty percent of the experts gave either ‘excellent’ or ‘very good’ on the appropriateness of the two measures (FMS

and MC) to measure separate but complimentary aspects of gross motor skills of children ages 3-6. For the older group (7-10) the ratings were slightly lower (75%). The potential of the test to be used to track individual progress over time was asked in question 5. Eighty-five percent of respondents judged the FG-MSA as either 'excellent' or 'very good' on intended use #1 for the younger group. The ratings for the older group was slightly lower (75%). Question 6 was concerned with the potential of the test to be used as a tool to evaluate physical education programs. This is the second intended use of the test. Eighty percent of respondents gave either 'excellent' or 'very good' on this validity issue for the younger group. The ratings for the older group were 15% lower than that for the younger group. Finally, question 7 asked participants to rate the potential of the set of items to be used as a tool to detect deficits in movement concept understanding, as well as problems in fundamental movement skill development. Again, 85% of the experts judged as either 'excellent' or 'very good' on this issue for the younger group. A slightly lower rating (75%) was given for the 7-10 year-olds.

Table 19: Percentage of responses and median values for evidence validity at the test level.

Questions	Age	Percentages					Median
		(Poor)				(Excellent)	
		1	2	3	4	5	
1. Judge the extent to which there is a match between the main purpose of this test and the set of items proposed for the domain of FUNDAMENTAL MOVEMENT SKILLS.		0.0	5.0	10.0	45.0	40.0	4
2. Judge the extent to which there is a match between the main purpose of this test and the set of items proposed for the domain of MOVEMENT CONCEPTS.		0.0	5.0	10.0	10.0	75.0	5
3. Judge the extent to which the set of items proposed by the developers of the current instrument represent skills/concepts that are frequently taught to children in preschool and early elementary grades.		0.0	0.0	15.0	40.0	45.0	4
4. Rate the appropriateness of the set of items proposed for the two domains (Fundamental Movement Skill and Movement Concepts) to measure separate but complimentary aspects of gross motor skills of children ages 3-6 and 7-10.	3-6	0.0	15.0	5.0	40.0	40.0	4
	7-10	5.0	0.0	20.0	40.0	35.0	4
5. Rate the potential of this set of items to be used as a tool to monitor/track individual progress during and/or following instructions (intended used 1)	3-6	5.0	5.0	5.0	35.0	50.0	4.5
	7-10	10.0	5.0	10.0	40.0	35.0	4
6. Rate the potential of this set of items to be used as a tool to evaluate effectiveness of the instructional program to improve children's movement skills (intended use 2)	3-6	5.0	5.0	10.0	35.0	45.0	4
	7-10	10.0	5.0	20.0	35.0	30.0	4
7. Rate the potential of this set of items to be used as a tool to detect eventual deficits in movement concept understanding, as well, problems in fundamental movement skill development (intended use 3).	3-6	5.0	0.0	10.0	40.0	45.0	4
	7-10	5.0	5.0	15.0	45.0	35.0	4

4.0 DISCUSSION

In the preceding section results were presented for evidence validity at both item and test levels. Little information was provided regarding how to modify the items and include or exclude items for the item level analysis. The current section deals with these issues in more detail. In addition, results related to the test level analysis are discussed in this section. Besides, this section includes the list of items that resulted from the current study and the discussion of how these items will be used in the next phase of the development of the FG-MSA.

4.1 CONTENT-RELATED VALIDITY AT THE ITEM LEVEL

A total of seven items (running, horizontal jump, leaping, hopping, galloping, skipping, and sliding) were selected for the Locomotor subdomain. With the exception of 'Leaping' and 'Galloping' all items were included in the test for both age groups. The skill of 'Leaping' might be considered too difficult for younger children to perform due to taking-off on one foot and landing on the other foot. The strength, balance, and coordination required are at an advance level. 'Galloping' was not included in the test. 'Galloping' was rated 'very important' for the younger group by 70% of judges and for the older group by only 45%. Thus, 'Galloping' was replaced by sliding for both age groups. Sliding was suggested by several experts through written comments. 'Jumping from height' was also added for the younger group only to parallel 'Leaping' for the older group. Gallahue and Donnelly (2003) hypothesize a developmental sequence of this skill in their text book. 'Walking' and 'Vertical Jump' were also suggested for inclusion. 'Walking' may be considered too easy even for the younger group and for that reason

it was not included in the test. The skill of 'Vertical Jump' was not included because it supposedly has similar characteristics as both 'Horizontal Jump' and 'Leaping'.

A total of nine items were selected for the Manipulative subdomain. With the exception of 'Batting (moving ball)' and 'Soccer Dribbling' all items will be included in the test for both age groups. 'Batting' is only included for the older group. A moving ball can be too hard to track by the younger group because the concept of time space is not fully developed at this level. 'Soccer Dribbling' is also a complex skill that requires perception and coordination abilities that may not be well developed in the younger group. Based on those reasons these two items are not included for the younger group. Five of the nine items included for the subdomain of Manipulation were revised from their original form. First, the skill of 'Approach Kicking' was originally proposed so that examinees would kick a stationary ball to a target. A modification was done for the 7-10 year-olds so that the ball will be rolled towards the examinee to be kicked. Side-Arm Striking is the second item modified to fit age group requirements. Only 25% of the respondents judged this item as 'very important' to be tested among the younger group. Thus, it was changed so that instead of a moving ball, examinees will be tested on striking a suspended ball. This should make this task more appropriate for 3-6 year-olds. A third item was modified to fit to the older group. The item of 'Hand Dribbling' was included for the older group, but examinees will be walking while performing the task. The younger group is still being tested on a stationary version of the skill. 'Catching' was the fourth item revised. In order to fit the requirements of the older group, examinees will be tested on their ability to perform the task while moving, rather than standing still. Finally, the skill of 'Trapping' was modified before its inclusion in the test. Previously, 'Trapping' was proposed to test examinees' ability to use their shin to stop a ball rolled towards them. The skill was modified so that examinees will stop the

ball using either foot, instead of their shin. In addition, while 3-6 year-olds will be standing and waiting for the ball, the older group will be moving towards the ball before attempting to stop it.

Behavioral objective three is ‘Stability Skills’. The three items originally proposed were included in the test. The skill of ‘Walk on Low Beam’ was modified for the older group. In order to make it more challenging, it was suggested to have three obstacles on the top of the beam. ‘Dodging’ was included for both age groups even though only 40.0% judged this item as ‘very important’ to be included for the younger group. It should be noted that experts were not given the developmental sequence for ‘Dodging’ while rating this item online. The lower rating seen here could be due to the fact that its developmental sequence is fairly unknown among teachers and professors. Gallahue and Donnelley (2003) were the first to publish a hypothesized development sequence for ‘Dodging’. In addition to these changes, a new item called ‘Non- Locomotor’ was added under the category of Stability Skills. The reasons for its inclusion are addressed in details next.

Three reasons were used to eliminate the subdomain of ‘Body Awareness’ and reclassify ‘Bending’ and ‘Twisting’ as ‘Non- Locomotor’ in the ‘Stability’ subdomain. ‘Symmetrical/Nonsymmetrical’ was moved to the ‘Body Shapes’ item which was placed under the ‘Relationship Awareness’ subdomain. First, the items under this category received the lowest ratings from content experts. This reason alone warned the developers of the current test to carefully analyze possible problems with the set of items for this category. The second reason was based on input from experts. Written comments were analyzed to identify the source of the low ratings. Some experts reported they were not sure whether the subdomain of ‘Body Awareness’ should be a single category since the items under this category were too specific. Some text books (Doherty & Bailey, 2003; Kirchner & Fishburne, 1998) recommended the

original organization while others placed the items under the 'Body Awareness' subdomain. For example, 'Body Shapes' is placed under the subdomain of 'Relationship Awareness' in Graham's text book (Graham et al., 2004). There, it is named 'Relationship of body parts'. In addition, while Symmetrical/Nonsymmetrical is considered as a single item in some books (Allison & Barrett, 2000), it is placed under the 'Relationship of body parts' in Graham's text and treated as a measure dimension of that category. Regarding the other two items ('Body Twisting' and 'Body Bending') proposed in the current test, it was decided to move them to the Stability subdomain. In addition, these items now measure dimensions under a single item called Non-Locomotor under the Stability subdomain. The term Non-Locomotor is sometimes used interchangeably with Axial Movements and represents a subset of the movement category of Stability (Gallahue & Donnelly, 2003). Therefore, the unexpected low item ratings, the experts' comments along with the support from the literature consulted led the developers of the current test to make such radical changes for the movement concept category of 'Body Awareness'.

Behavioral objectives, 5, 6, and 7 received ratings that were fairly consistent compared to the previous movement concept category (Body Awareness) and the three fundamental movement skill categories (Locomotor, Manipulative, and Stability). This consistency and high item ratings within these three subdomains yielded the exclusion of only one item (Flow), and the inclusion two other items (Rhythm and Relationship of body parts). 'Flow' was proposed along with two other items under the movement concept category of 'Effort Awareness'. It was removed because of its low ratings among experts. This was confirmed by written comments from experts that reported this item as too difficult to be measure even for older children. It was decided to include the concept of 'Rhythm' under 'Effort Awareness'. This is an important concept since it is used in almost all forms of movement skills. The second concept included was

‘Relationship of body parts’. As discussed earlier, the item ‘Body Awareness’ was renamed ‘Relationship of body parts’ and included under the subdomain of ‘Relationship Awareness. These were the only two changes when combining all three movement concept categories of Space, Effort and Relationship Awareness. This consistency seen for items under these three categories might have been due to the fact that there is little disagreement in the literature regarding the importance of these items for both age groups. Still, this has to be confirmed when actual testing is carried out with examinees from different age groups. This is what the next phase of this project intends to accomplish.

4.2 CONTENT-RELATED VALIDITY AT THE TEST LEVEL

The results of the set of seven questions provided initial support for the content-related validity at the test level. In addition, the data indicated that the set of items proposed for both domains of the FG-MSA were found to be appropriate and representative of a larger set of items. Extra evidence for representativeness of the set of items was collected by asking experts to suggest items whenever they felt necessary for each behavioral objective. In addition, written comments from experts were used to modify items which in turn helped to support the appropriateness of the set of items.

Fewer experts (75.0%) felt that the set of items was appropriate to measure separate but complimentary aspects of gross motor skills for the older group compared to 85.0% for the younger group. The difference might have been due to the fact that some items were rated as too easy for the older group. Several items were revised or excluded to correct for this imbalance between the two age groups.

Respondents found the FG-MSA to be more appropriate to assess individual progress over time and detect deficits in movement concept understanding and/or problems in fundamental movement skill development (intended uses 1 and 3) than to be used as a tool to evaluate effectiveness of the instructional program (intended use 2). Written comments suggested concerns regarding this intended use since they did not see how the items would actually be tested. It should be noted that only two complete items were given as an example in the introduction of the forms that experts completed online. The developmental sequences for the fundamental movement skill items and the target criteria for the movement concept items will be developed in the next phase of this project. In addition, for all three intended uses, the ratings for the older group were slightly lower compared to the younger group. Use for evaluation of the effectiveness of the instructional program was even lower with only 65% of respondents giving either 4 or 5 for 7-10 year olds. The same issue discussed earlier regarding the level of simplicity of the items for the older group in some of the behavioral objectives applies in the case of the three intended uses of the FG-MSA. Despite some differences in ratings for each question these results provided initial evidence for content-related validity at the test level.

4.3 SUGGESTED TEST FORMAT

The current instrument is being developed to be used as an informal observational assessment tool for making within-individual comparisons, tracking individual changes in fundamental movement skill performance and movement concept understanding over time. The final version of test will have two subtests which are fundamental movement skill and movement concept.

Next, the list of items for the subtest of fundamental movement skill is presented along with an explanation of how these items will be tested.

4.3.1 Fundamental Movement Skill

The complete list of items for the subtest of fundamental movement skills is shown in [Table 20](#).

This subtest has 21 items that will test performance on locomotor movement skills in children 3-10 years of age. The observation of each skill will be compared to its developmental sequence (DS).

Table 20: Final list of items for fundamental movement skills

Items	Age Groups	
	3-4	7-10
Locomotor		
Running	✓	✓
Horizontal Jumping	✓	✓
Leaping		✓
Hopping	✓	✓
Jumping from height	✓	
Skipping	✓	✓
Sliding	✓	✓
Manipulative		
Overhand Throwing	✓	✓
Approach Kicking	Stationary ball	Moving ball
Side-Arm Striking	Suspended ball	Moving ball
Batting		Moving ball
Hand Dribbling	Stationary	Walking btw cones
Catching	Stationary	Moving
Trapping	Stationary	Moving
Underhand Throwing	✓	✓
Soccer Dribbling		✓
Stability		
Walk on low beam	Without obstacle	With obstacle
Dodging	✓	✓
One-foot balance	✓	✓
Single-Knee balance	✓	✓
Hand-and-Knee balance	✓	✓
Non-Locomotor	✓	✓

According to Gallahue and Donnelly (2003) a developmental sequence is ‘the orderly, predictable sequence of increased motor control and movement competence in the normally developing individual’ (Gallahue & Donnelly, 2003, p. 690). Developmental sequences have

verbal and visual descriptions to help examiners to evaluate examinees' stage of development. Most of the fundamental movement skills selected for the current test have validated developmental sequences. Some skills (e.g., walking on low balance beam) have hypothesized, yet not validated, developmental sequences that may be used, as well. Gallahue and Donnelly (2003) have hypothesized a developmental sequence for beam walk. [Table 21](#) shows verbal descriptions of the developmental sequence of beam walk.

Table 21: Developmental sequence for the low beam wall (Gallahue and Donnelly, 2003)

Initial Stage	Elementary Stage	Mature Stage
1. Balances with support	1. Can't walk a 2-inch low beam	1. Can walk a 1-inch low beam
2. Walks forward while holding for support	2. Uses follow-step w/ dominant foot leading	2. Uses alternating stepping action
3. Uses follow-step w/ dominant foot leading	3. Eyes focus on apparatus	3. Eyes focus beyond beam
4. Eyes focus on feet	4. May press one arm to trunk while trying to balance w/ the other	4. Both arms use at will to aid balance
5. Body rigid	5. Loses balance easily	5. Can move forward, backward, and sideways w/ assurance and ease
6. No compensating movements	6. Limited compensating movements	6. Movements are fluid, relaxed, and in control
	7. Can move forward, backward, and sideways but w/ considerable concentration and effort	7. may lose balance occasionally

Students are tested and compared with the verbal descriptions of each developmental sequence for each fundamental movement skill. Then, examinees are assigned to one of the following stages of development: I= initial; E= elementary; M= mature based on his/her performance. Since one of the intended uses of the current assessment tool is to track individuals' progress over time, students could be retested after a period of instruction, for example, and once again compared to the stages of development for each skill. This would allow for verification of levels of progression between the two testing sessions.

4.3.2 Movement concepts

The list of items for the subtest of movement concepts is shown on [Table 22](#). This subtest has 10 items that will test functional understanding of movement concepts in children 3-10 years of age.

The observation of each concept will be compared to its target criterion. Movement concepts are not like FMS skills that can be measure as a continuum process (Gallahue & Donnelly, 2003). While it would be appropriate to say that Cullen’s level of performance in the skill of hopping is ‘initial’, the same cannot be said in terms of a concept (e.g., pathways).

Table 22: Final list of items for movement concepts

Items	Age Groups	
	3-6	7-10
Space Awareness		
Location	✓	✓
Directions	✓	✓
Levels	✓	✓
Pathways	✓	✓
Extensions	✓	✓
Effort Awareness		
Time	✓	✓
Force	✓	✓
Rhythm	✓	✓
Relationship Awareness		
With objects and/or people	✓	✓
With people only	✓	✓
Of body parts	✓	✓

In assessing a concept individuals receive either 0 or 1. There is no less or more understanding. Further, while in FMS individuals' performance are compared to developmental sequences for each skill, in movement concepts individuals' understanding of each concept is compared to what is called here 'target criteria'. The target criteria for a concept refer to the characteristics associated to that concept that can be measured. The concept of pathways (straight, curved, zigzag) will be used as an example. The target criteria for this concept could be: (1) change pathways upon signal without hesitation; (2) Speed does change as change pathways, etc. Note that the term 'functional' is used before the word 'understanding'. This is to differentiate from cognitive movement concept understanding which is limited to asking individuals to simply describe what is to move in straight, curved or zigzag pathways. Functional understanding means that the examinee is required to show his/her understanding of the concept being presented by actually performing it.

Therefore, the next phase of this project will consist of assigning performance criteria for fundamental movement skills and movement concepts, developing test administration guidelines, and collecting further evidence for validity and reliability. The literature will be reviewed for validated and hypothesized developmental sequences for fundamental movement skills. Then these developmental sequences will be assigned to each skill of the test. Movement concepts are not skills to be mastered but concepts to be understood by students. Thus, target criteria will be developed for each movement concept tested.

The second important issue addressed in the next phase of this project is the development of the test administration guidelines. Test administration procedures will be developed for each skill and movement concept. These guidelines will cover material to be used for test administration, instructions for data collection and score interpretation. Finally, by testing

students further validity evidence (internal validity) as well as reliability evidence will be collected. Analysis of internal validity examines consistency in the relationships among the items relative to the conceptual framework for the construct being measured (Nitko, 2001). Items within dimensions (domains) are more homogeneous or more related to one another than items across dimensions. The current test is being developed to test two separate domains (FMS and MC). Thus, items within domains are expected to correlate moderately high, but lower across domains. Reliability analysis will verify the consistency of scores across items measuring approximately the same thing, over time, and over scorers (raters).

4.4 CONCLUSION

Evidence for content-related validity was collected at both item and test level through judgments of experts. Experts were eight college/university professors and twelve physical education teachers. Evidence for content-related validity at item level was accomplished by asking experts to judge the importance a list of 31 items (Fundamental Movement Skills=16, Movement Concepts=15) on a 4 points likert-type scale from 1 through 4, where 1 being 'not important at all and 4 'very important'. The combination of item ratings and written comments from experts resulted in the modification of 9 items. In addition, two items were deleted from the test. Further, four items were included based on suggestions from experts. Three other items were collapsed to form two single items. This resulted in a total of 31 items (see Tables 22 and 23 above). Evidence was also collected for content-related validity at the test level. This was done by asking experts to judge in a set of seven questions, the match item pool-purpose and item pool-intended uses, in addition to representativeness and appropriateness of the set of item proposed. A 5 points likert-type scale was used, with 1 being 'poor' and 5 being 'excellent'. Results of this study provided an initial support for the content-related validity of the Furtado-Gallagher Movement Skill Assessment. Content validity is an ongoing process throughout the development of any instrument and further analysis will be needed to confirm the content validity of this assessment tool.

REFERENCES

- AERA, APA, & NCME. (1999). *Standards for educational and psychological testing*. Washington D.C.: American Educational Research Association.
- Allison, P. C., & Barrett, K. R. (2000). *Constructing children's physical education experiences: understanding the content for teaching*. Needham Heights: Allyn and Bacon.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper Saddle River: Prentice-Hall.
- Berk, R. A. (1984). Conducting the item analysis. In R. A. Berk (Ed.), *A guide to criterion-referenced test construction*. Baltimore: The Johns Hopkins University Press.
- Burton, A. W., & Miller, D. E. (1998). *Movement skill assessment*. Champaign: Human Kinetics.
- Doherty, J., & Bailey, R. (2003). *Supporting physical development and physical education in the early years*. Buckingham: Open University Press.
- Gallahue, D. L., & Donnelly, F. C. (2003). *Developmental physical education for all children* (4th ed.). Champaign, IL: Human Kinetics.
- Gallahue, D. L., & Ozmun, J. C. (1998). *Understanding motor development : infants, children, adolescents, adults* (4th ed.). Dubuque, Iowa: McGraw-Hill.
- Graham, G., Holt/Hale, S. A., & Parker, M. (2004). *Children moving: a reflective approach to teaching physical education* (6th ed.). Mountain View: Mayfield.
- Haley, S. M., Coster, W. J., & Faas, R. M. (2001). A content validity study of the Pediatric Evaluation of Disability Inventory. *Pediatric Physical Therapy*, 177-184.
- Holt/Hale, S. A. (1999). *Assessing motor skills in elementary physical education*. Oak Ridge: NASPE Publications.
- Kirchner, G., & Fishburne, G. J. (1998). *Physical education for elementary school children* (10th ed.). Boston, Mass.: WCB/McGraw Hill.
- Kogut, S. P., & National Association for Sport and Physical Education. (2003). *Beyond activities : learning experiences to support the National Physical Education Standards : elementary*. Reston, VA: National Association for Sport and Physical Education.
- Loovis, E. M., & Ersing, W. F. (1979). *The Ohio State University Scale of Intra-Gross Motor Assessment*

(*OSU Sigma*). Cleveland Heights.

McClenaghan, B. A., & Gallahue, D. L. (1978). *Fundamental movement : a developmental and remedial approach*. Philadelphia: Saunders.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd. ed.). New York: American Council on Education/Macmillian Publishing Co.

NASPE. (2004). *Moving into the future: National standards for physical education* (2nd ed.). Reston: McGraw Hill.

Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Columbus: Prentice-Hall.

Rudner, L. M., & Schafer, W. D. (2002). *What teachers need to know about assessment*. Washington DC: National Education Association.

Thorn, D. W., & Deitz, J. C. (1989). Examining content validity through the use of content experts. *The Occupational Therapy Journal of Research*, 9(6), 334-347.

Ulrich, D. A. (2000). *Test of gross motor development-second edition* (2nd ed.). Austin: Pro-ed.

Wickstrom, R. L. (1983). *Fundamental motor patterns* (3rd ed.). Philadelphia: Lea & Febiger.

APPENDIX A

LIST OF BOOKS USED FOR SELECTION OF ITEMS

1	Gallahue, D. L., & Donnelly, F. C. (2003). Developmental physical education for all children (4th ed.). Champaign, IL: Human Kinetics.
2	Thomas, K. T., Lee, A. M., & Thomas, J. R. (2003). Physical education methods for elementary teachers : 70 proven lesson plans (2 nd ed.). Champaign, IL: Human Kinetics.
3	Pangrazi, R. P. (2001). Dynamic physical education for elementary school children. Needham Heights: Allyn and Bacon
4	Haywood, K. M. (2001). Learning activities for life span motor development. Champaign: Human Kinetics.
5	Gabbard, C. (2004). Lifelong motor development (4th ed.). Needham Heights: Allyn & Bacon.
6	Payne, V. G. and Isaacs, L. D. (1999). Human motor development: a lifespan approach. Mayfield Publishing: Mountain View.
7	Graham, G., Holt/Hale, S. A., & Parker, M. (2004). Children moving: a reflective approach to teaching physical education (6th ed.). Mountain View: Mayfield.
8	Doherty, J. and R. Bailey (2003). Supporting physical development and physical education in the early years. Buckingham, Open University Press.
9	Kirchner, G. and G. J. Fishburne (1998). Physical education for elementary school children. Boston, Mass., WCB/McGraw Hill.
10	Kogut, S. P. and National Association for Sport and Physical Education. (2003). Beyond activities: learning experiences to support the National Physical Education Standards: elementary. Reston, VA, National Association for Sport and Physical Education.
11	Colvin, A. V., N. J. E. Markos, et al. (2000). Teaching the nuts and bolts of physical education: building basic movement skills. Champaign, IL, Human Kinetics.
12	Human Kinetics (Organization) and B. Pettifor (1999). Physical education methods for classroom teachers. Champaign, IL, Human Kinetics.
13	Chepko, S., & Arnold, R. K. (2000). Guidelines for physical education programs: grades standards, objectives, and assessments. Needham Heights: Allyn and Bacon.
14	Allison, P. C., & Barrett, K. R. (2000). Constructing children's physical education experiences: understanding the content for teaching. Needham Heights: Allyn and Bacon.
15	Kogan, S. (2004). Step by step: a complete movement education curriculum (2nd ed.). Champaign: Human Kinetics.
16	Nichols, B. (1994). Moving and learning: the elementary school physical education experience (3rd ed.). St. Louis: Mosby.

APPENDIX B

ITEM REVIEW FORM USED TO COLLECT INFORMATION OVER THE INTERNET

Part A – Section 1: Form Guidelines

Dear content expert,

If you have reached to this page, then you have successfully followed the instructions sent via email. Please take a minute to read the following information since it will help you filling in the item review form.

PRIOR TO BEGINNING YOUR RATING, BELOW YOU WILL FIND:

1. **Example of the final version of the test performance record**
 - 1.1 Fundamental Movement Skill Subtest
 - 1.2 Movement Concept Subtest
2. **Useful considerations**

1. EXAMPLE OF THE FINAL VERSION OF TEST PERFORMANCE RECORD:

The final version of the test that teachers use will look something like the following. Notice that only one skill and one concept are presented below. All test items will follow the same pattern.

1.1 FUNDAMENTAL MOVEMENT SKILL SUBTEST: Skill and its performance criteria:

Horizontal Jump	Performance Criteria	1 st attempt	2 nd attempt	Score			
	1. Preparatory movement includes flexion of both knees with arms extended behind body.						
	2. Arms extended forcefully forward and upward reaching full extension above head.						
	3. Take off and land on both feet simultaneously.						
	4. Arms are thrust downward during landing.						
Total Score							
Skill proficiency level				<table border="1" style="width: 100%; text-align: center;"> <tr> <td>B</td> <td>I</td> <td>A</td> </tr> </table>	B	I	A
B	I	A					
				(B) Beginner (I) Intermediate (A) Advanced			

1.2 MOVEMENT CONCEPT SUBTEST: Concept and its measurable characteristics:

Pathways	Characteristics Tested	Score	
Straight / Curved / Zigzag	1. Change pathways upon signal without hesitation.	0	1
	2. Speed does not change as change pathways.	0	1
Total Score:			

2. USEFUL CONSIDERATIONS

Please read the information below as it will help you while rating the items.

• SYSTEM REQUIRED

- Internet connection
- Microsoft Internet Explorer™ (recommended). You can download Microsoft Internet Explorer for free by [clicking here](#).

• INTERNET BROWSER HELPFUL HINTS

- **Required** fields have to be selected before moving to next page.
- Clicking question marks (?) that follow each item will pop up the image associated to it - e.g., running. Try it now! >> [\(?\)](#)
- Once you begin rating the items, please follow through to the end WITHOUT going back (clicking 'back' button on Internet browser).

• CONSULTATION AND TIME

- Please do not consult anyone to help you rate the items on sections 1 and 2.

Simply use your experience in the field of physical education to answer each question.

- It is estimated that the entire process of rating the items should take approximately 30 to 45 minutes.

• COMPENSATION

- We thank you for your cooperation and for that reason we will list your name in the user manual of the current assessment tool.
- Also, as an exchange for your voluntary participation, we will send you \$50.00 check upon receiving the information submitted through this online form.

• PURPOSE OF THE TEST

- The purpose of the FURTADO-GALLAGHER MSA will be:

... to assess children's skill proficiency on fundamental movement skills and understanding of movement concepts. Students' performance will be compared to pre-established performance criteria for verification of levels of mastery in the attributes tested.

• THE INTENDED USES OF THE TEST

- **The 3 main intended uses of the FURTADO-GALLAGHER MSA will be to:**
 - a. monitor individual progress during and/or following instruction;
 - b. evaluate effectiveness of instructional programs with the intent of adjusting it according to students' needs;
 - c. detect deficits in movement concept understanding, as well as, problems in fundamental movement skill development.

QUICK CHECKLIST

- I have read the information above.
- I have printed out this page for future reference - [click here to print \(PDF format\)](#)
- I have about 40 minutes of free time to spend filling in the online form (form will be displayed when you click link below)

IF YOU CHECKED ALL THREE CHECKBOXES ABOVE, YOU ARE READY TO BEGIN RATING THE ITEMS.

[PLEASE CLICK HERE TO START](#) >>>>

* Required Fields

> Background Information	
* ID number:	<input type="text"/> (The ID # was sent to you by email along with the general instructions that helped you to reach this page. Please DO NOT continue unless you have this piece of information. Sent me an email if you cannot find your ID #).
* Email:	<input type="text"/>
* First name:	<input type="text"/>
* Last name:	<input type="text"/>
Middle Initial:	<input type="text"/>
Years teaching PE:	<input type="text"/>
Highest education:	<input type="text"/>
Area of expertise:	<input type="text"/>
List major awards received (if any):	<input type="text"/>
> Address Information	
Check will be sent to the address provided below:	
Address 1:	<input type="text"/>
Address 2:	<input type="text"/>
City:	<input type="text"/>
State:	<input type="text"/>
Zip Code:	<input type="text"/>

Part B – Section 1: Item Level Ratings (Fundamental Movement Skills)

(1/3)

Directions for Fundamental Movement Skill Items

First, read carefully each behavioral objective and its list of items. Next, please indicate the extent to which each item should be included in the current test.

Judge a test item solely on the basis of its importance/relevance for each age group (3-6 and 7-10). Please use the 4-point rating scale shown below:

For example:

Rating **walking** as 'not important at all' for 7- 10 year olds indicates that you feel this item should not be included in that age group test. Checking 'not important at all' for both age groups indicates you feel the item 'walking' SHOULD NOT be included in the test AT ALL.

- 4= VERY IMPORTANT** (to be included in the test)
- 3= MODERATELY IMPORTANT** (to be included in the test)
- 2= SOMEWHAT IMPORTANT** (to be included in the test)
- 1= NOT IMPORTANT AT ALL** (to be included in the test)

FUNDAMENTAL MOVEMENT SKILLS			
BEHAVIORAL OBJECTIVE 01: LOCOMOTOR			
To develop student's gross locomotive movement skills that involve transporting the body in a horizontal or vertical direction from one place to another.			
(All Fields Required)			
USE THE SCALE BELOW FOR ITEMS 01-06			
4=very important, 3=moderately important, 2=somewhat important, 1=not important at all			
#	Item	Rate the degree of importance for 3-6 Year-Olds	Rate the degree of importance for 7-10 Year-Olds
01	Running (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
02	Horizontal Jumping (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
03	Leaping (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
04	Hopping (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
05	Galloping (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
06	Skipping (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>

Based on the set of items proposed ABOVE (01-06) would you:

ADD items to the BEHAVIORAL OBJECTIVE 01?

If you think so, then please use field below to suggest your items.

Comments for items **01 through 06** (if any):

FUNDAMENTAL MOVEMENT SKILLS			
BEHAVIORAL OBJECTIVE 02: MANIPULATION			
To develop students' gross manipulative movement skills in which force is imparted to or received from objects. (All Fields Required)			
USE THE SCALE BELOW FOR ITEMS 07-13			
4=very important, 3=moderately important, 2=somewhat important, 1=not important at all			
#	Item	Rate the degree of importance for 3-6 Year-Olds	Rate the degree of importance for 7-10 Year-Olds
07	Overhand Throwing (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
08	Kicking (stationary ball) (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
09	Side-Arm Striking (moving ball) (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
10	Batting (moving ball) (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
11	Stationary Dribbling (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
12	Catching (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
13	Trapping (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>

Based on the set of items proposed ABOVE (07-13) would you:

ADD items to the BEHAVIORAL OBJECTIVE 02?

If you think so, then please use field below to suggest your items.

Comments for items **07 through 13** (if any):

FUNDAMENTAL MOVEMENT SKILLS			
BEHAVIORAL OBJECTIVE 03: STABILITY			
To develop students' static and dynamic stability skills that place a premium on gaining and/or maintaining one's equilibrium in relationship to the force of gravity.			
(All Fields Required)			
USE THE SCALE BELOW FOR ITEMS 14-16			
4=very important, 3=moderately important, 2=somewhat important, 1=not important at all			
#	Item	Rate the degree of importance for 3-6 Year-Olds	Rate the degree of importance for 7-10 Year-Olds
14	Walk on Low Beam (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
15	Dodging (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
16	One-Foot Balance (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>

⤴ Based on the set of items proposed ABOVE (14-16) would you:

ADD items to the BEHAVIORAL OBJECTIVE 03?

If you think so, then please use field below to suggest your items.

Comments for items **14 through 16** (if any):

Part B – Section 1: Item Level Ratings (Movement Concepts)

(2/3)

Directions for Movement Concept Items

First, read carefully each behavioral objective and its list of items. Next, please indicate the extent to which each item should be included in the current test.

Judge a test item solely on the basis of its importance/relevance for each age group (3-6 and 7-10). Please use the 4-point rating scale shown below:

For example:

Rating **walking** as 'not important at all' for 7- 10 year olds indicates that you feel this item should not be included in that age group test. Checking 'not important at all' for both age groups indicates you feel the item 'walking' SHOULD NOT be included in the test AT ALL.

- 4= VERY IMPORTANT** (to be included in the test)
- 3= MODERATELY IMPORTANT** (to be included in the test)
- 2= SOMEWHAT IMPORTANT** (to be included in the test)
- 1= NOT IMPORTANT AT ALL** (to be included in the test)

MOVEMENT CONCEPTS			
BEHAVIORAL OBJECTIVE 04: BODY AWARENESS			
To develop in children the ability to know what actions the body is capable of doing and understand how these actions are performed.			
(All Fields Required)			
USE THE SCALE BELOW FOR ITEMS 17-21			
4=very important, 3=moderately important, 2=somewhat important, 1=not important at all			
#	Item	Rate the degree of importance for 3-6 Year-Olds	Rate the degree of importance for 7-10 Year-Olds
17	Body Shapes (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
18	Symmetrical / Nonsymmetrical (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
19	Body Twisting	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
20	Body Bending	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
21	Front Swing (?)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>

Based on the set of items proposed ABOVE (17-21) would you:

ADD items to the BEHAVIORAL OBJECTIVE 04?

If you think so, then please use field below to suggest your items.

Comments for items **17 through 21** (if any):

MOVEMENT CONCEPTS			
BEHAVIORAL OBJECTIVE 05: SPACE AWARENESS			
To develop, through movement, students' functional understanding of concepts that relate to <i>WHERE</i> the body moves in the space.			
(All Fields Required)			
USE THE SCALE BELOW FOR ITEMS 22-26			
4=very important, 3=moderately important, 2=somewhat important, 1=not important at all			
#	Item	Rate the degree of importance for 3-6 Year-Olds	Rate the degree of importance for 7-10 Year-Olds
22	Location General Space - Self Space	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
23	Directions Up / Down - Forward / Backward - Right / Left Clockwise / Counterclockwise	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
24	Levels Low / Middle / High	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
25	Pathways Straight / Curved / Zigzag	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
26	Extensions Large / Small - Far / Near	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>

Based on the set of items proposed ABOVE (22-26) would you:

ADD items to the BEHAVIORAL OBJECTIVE 05?

If you think so, then please use field below to suggest your items.

Comments for items **22 through 26** (if any):

MOVEMENT CONCEPTS					
BEHAVIORAL OBJECTIVE 06: EFFORT AWARENESS					
To develop, through movement, students' functional understanding of concepts that relate to <i>HOW</i> the body moves in the space.					
(All Fields Required)					
USE THE SCALE BELOW FOR ITEMS 27-29					
4=very important, 3=moderately important, 2=somewhat important, 1=not important at all					
#	Item	Rate the degree of importance for 3-6 Year-Olds		Rate the degree of importance for 7-10 Year-Olds	
27	Time Fast / Slow	4= <input type="checkbox"/>	3= <input type="checkbox"/>	2= <input type="checkbox"/>	1= <input type="checkbox"/>
28	Force Strong / Light - Hard / Soft	4= <input type="checkbox"/>	3= <input type="checkbox"/>	2= <input type="checkbox"/>	1= <input type="checkbox"/>
29	Flow Bound / Free	4= <input type="checkbox"/>	3= <input type="checkbox"/>	2= <input type="checkbox"/>	1= <input type="checkbox"/>

Based on the set of items proposed ABOVE (27-29) would you:

ADD items to the BEHAVIORAL OBJECTIVE 06?

If you think so, then please use field below to suggest your items.

Comments for items **27 through 29** (if any):

MOVEMENT CONCEPTS			
BEHAVIORAL OBJECTIVE 07: RELATIONSHIP AWARENESS			
To develop, through movement, students' functional understanding of concepts that relate to one's relationship with objects and/or people.			
(All Fields Required)			
USE THE SCALE BELOW FOR ITEMS 30 and 31			
4=very important, 3=moderately important, 2=somewhat important, 1=not important at all			
#	Item	Rate the degree of importance for 3-6 Year-Olds	Rate the degree of importance for 7-10 Year-Olds
30	With objects and/or people (Over / Under), (In front / Behind), (Meeting / Parting), (Along / Through)	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>
31	With people only (Mirroring / Matching), (Leading / Following),	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>	4= <input type="checkbox"/> 3= <input type="checkbox"/> 2= <input type="checkbox"/> 1= <input type="checkbox"/>

⚡ Based on the set of items proposed ABOVE (30 and 31) would you:

ADD items to the BEHAVIORAL OBJECTIVE 07?

If you think so, then please use field below to suggest your items.

Comments for items **30 and 31** (if any):

Part B – Section 2: Test Level Ratings

(3/3)

Match between the set of items and the test's general characteristics

Directions

First, read carefully each question. Next, please rate each question using the five-point rating scale shown below: (Notice that the scale below is different from that used in the previous section).

- 5= EXCELLENT
- 4= VERY GOOD
- 3= GOOD
- 2= FAIR
- 1= POOR

01 - Judge the extent to which there is a match between the main purpose of this test and the set of items proposed for the domain of **FUNDAMENTAL MOVEMENT SKILLS:**

> [Click here to see comparison chart](#)

Check the option button corresponding to your rating for question 01 using the scale below **(Required)**

Poor	Fair	Good	Very Good	Excellent
1= <input type="checkbox"/>	2= <input type="checkbox"/>	3= <input type="checkbox"/>	4= <input type="checkbox"/>	5= <input type="checkbox"/>

02 - Judge the extent to which there is a match between the main purpose of this test and the set of items proposed for the domain of **MOVEMENT CONCEPTS:**

> [Click here to see comparison chart](#)

Check the option button corresponding to your rating for question 02 using the scale below **(Required)**

Poor	Fair	Good	Very Good	Excellent
1= <input type="checkbox"/>	2= <input type="checkbox"/>	3= <input type="checkbox"/>	4= <input type="checkbox"/>	5= <input type="checkbox"/>

03 - Judge the extent to which the set of items proposed by the developers of the current instrument represent skills/concepts that are frequently taught to children in preschool and early elementary grades.

> [Click here to see the list of all items](#)

Check the option button corresponding to your rating for question 03 using the scale below **(Required)**

Poor	Fair	Good	Very Good	Excellent
1= <input type="checkbox"/>	2= <input type="checkbox"/>	3= <input type="checkbox"/>	4= <input type="checkbox"/>	5= <input type="checkbox"/>

04 - Rate the appropriateness of the [set of items](#) proposed for the two domains (movement concepts and fundamental movement skills) to measure separate but complimentary aspects of gross motor skills of children ages 3-6 and 7-10.

Check the option button corresponding to your rating for question 04 using the scale below **(Both Required)**

3-6 Year Olds	Poor	Fair	Good	Very Good	Excellent
	1 = <input type="checkbox"/>	2 = <input type="checkbox"/>	3 = <input type="checkbox"/>	4 = <input type="checkbox"/>	5 = <input type="checkbox"/>

7-10 Year Olds	Poor	Fair	Good	Very Good	Excellent
	1 = <input type="checkbox"/>	2 = <input type="checkbox"/>	3 = <input type="checkbox"/>	4 = <input type="checkbox"/>	5 = <input type="checkbox"/>

05 - Rate the potential of this [set of items](#) to be used as a tool to monitor/track individual progress during and/or following instructions (test intended use # 1)

Check the option button corresponding to your rating for question 05 using the scale below **(Both Required)**

3-6 Year Olds	Poor	Fair	Good	Very Good	Excellent
	1 = <input type="checkbox"/>	2 = <input type="checkbox"/>	3 = <input type="checkbox"/>	4 = <input type="checkbox"/>	5 = <input type="checkbox"/>

7-10 Year Olds	Poor	Fair	Good	Very Good	Excellent
	1 = <input type="checkbox"/>	2 = <input type="checkbox"/>	3 = <input type="checkbox"/>	4 = <input type="checkbox"/>	5 = <input type="checkbox"/>

06 - Rate the potential of this [set of items](#) to be used as a tool to evaluate effectiveness of the instructional program to improve children's movements/motor skills (intended use # 2).

Check the option button corresponding to your rating for question 06 using the scale below **(Both Required)**

3-6 Year Olds	Poor	Fair	Good	Very Good	Excellent
	1 = <input type="checkbox"/>	2 = <input type="checkbox"/>	3 = <input type="checkbox"/>	4 = <input type="checkbox"/>	5 = <input type="checkbox"/>

7-10 Year Olds	Poor	Fair	Good	Very Good	Excellent
	1 = <input type="checkbox"/>	2 = <input type="checkbox"/>	3 = <input type="checkbox"/>	4 = <input type="checkbox"/>	5 = <input type="checkbox"/>

07- Rate the potential of this [set of items](#) to be used as a tool to detect eventual deficits in movement concept understanding, as well as, problems in fundamental movement skill development (intended use # 3).

Check the option button corresponding to your rating for question 07 using the scale below **(Both Required)**

3-6 Year Olds	Poor	Fair	Good	Very Good	Excellent
	1= <input type="checkbox"/>	2= <input type="checkbox"/>	3= <input type="checkbox"/>	4= <input type="checkbox"/>	5= <input type="checkbox"/>

7-10 Year Olds	Poor	Fair	Good	Very Good	Excellent
	1= <input type="checkbox"/>	2= <input type="checkbox"/>	3= <input type="checkbox"/>	4= <input type="checkbox"/>	5= <input type="checkbox"/>

General Comments

Use the field below to provide general comments.

APPENDIX C

LOCAL AND/OR NATIONAL AWARDS RECEIVED BY THE TEACHERS THAT PARTICIPATED AS CONTENT EXPERTS

Teacher id #	Year	Teaching Excellence Awards
	1988	South Carolina Elementary Physical Education Teacher of the Year
T01	2000	NASPE Southern District Elementary Physical Education Teacher of the Year
	2002	National Board for Professional Teaching Standards Certification in Elementary Physical Education
T02*	1995	Southwest District Elementary PE Teacher of the Year
	2000	Idaho's Elementary PE Teacher of the year
T03	2001	Utah State Alumni K-12 PE Teacher of the year
	2002	N.W. District Elementary PE Teacher of the year
	2003	NASPE N.W. District Elementary Teacher of the year
T04	2001	Iowa Elementary Physical Education Teacher of the Year
	2003	AAHPERD National Mabel Lee Award Recipient
	----	Teacher of the Month – Malcolm Price Laboratory School
T05*	1997-98	Oakley Elementary School Teacher of the Year
T06*	2003	Georgia Elementary PE teacher of the year for GA-HPERD
	----	McHenry Primary teacher of the year (twice)
	----	CTAHPERD Elementary Teacher of Year
T07	----	A.H.M. Youth Services Friend of Children Award
	1998	Connecticut Elementary School Teacher of the Year
	1996	Teacher of the Year for Elementary PE for Maryland
T08	1998	Teacher of the Year for Elementary PE - Eastern District
	2003	National Board Certification
T09*	----	SC Governor's Physical Education – Hammond Hill Elementary
T10	2001	NASPE Teacher of the Year - Eastern District
T11	2000	Maryland Elementary PE Teacher of the Year
T12*	----	District Educator of the Year (Columbus, OH)
*		National Board Certified
---		Did not provide date

APPENDIX D

FEEDBACK FORM USED IN THE PILOT STUDY TO SEND COMMENTS ABOUT THE ITEM REVIEW FORM

Now that you had the opportunity to complete the Item Review Form, please use the feedback form given below to make comments/suggestions. The information provided here will help us to improve this Item Review Form before sending it to content experts.

Pilot Study

ITEM REVIEW FORM - Feedback Form

Your name: _____

I think the Item Review form is user-friendly and content experts will not have major problems to complete it.

- Strongly disagree
- Somewhat disagree
- Undecided
- Somewhat agree
- Strongly agree

1. Please use the field below to comment about the "Introductory Section"

You can comment about:
Amount of information provided, wording, information presented in a clear way, etc.

2. Please use the two following fields to comment about the " Section One" (Item importance and representativeness)

FUNDAMENTAL MOVEMENT SKILLS

You can comment about:
Wording, information presented in a clear way, etc.

MOVEMENT CONCEPTS

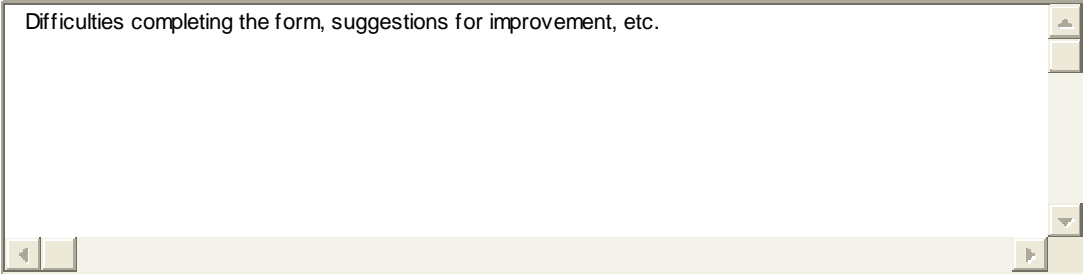
You can comment about:
Wording, information presented in a clear way, etc.

3. Please use the following field to comment about the " Section Two" (Item Pool Match)

You can comment about:
Wording, information presented in a clear way, etc.

4. Please use the following field to provide general comments that are not direct associated with the previous topics.

Difficulties completing the form, suggestions for improvement, etc.



APPENDIX E

SECOND EMAIL MESSAGE USED TO PROVIDE INFORMATION ON HOW TO RATE REACH THE ITEM REVIEW FORM ONLINE

November 1, 2004

University of Pittsburgh
Motor Behavior Laboratory
Subject: Furtado-Gallagher Movement Skill Assessment

URL (where the forms are hosted):
<http://www.virtualland.net/test/>

Your ID Number: <participants ID>

Dear, <participants name>

We would like to thank you for taking the time to participate in the process of collecting evidence for content-related validity of the **FURTADO-GALLAGHER Movement Skill Assessment** (FG-MSA).

The Furtado-Gallagher Movement Skill Assessment is a new criterion-referenced assessment tool that is being developed to test movement concepts and fundamental movement skills of children ages 3-10. The purpose of the current study is to collect validity evidence based on test content (content-related evidence). Information will be collected through an item review form (IRF). Item review forms are usually used to collect information from content experts on items being proposed for a given test.

The results of this study will help the developers of the FG-MSA to decide the list of items to be included in the final version of the test. In addition, information collected from this study will help us to measure the extent to which there is a match between the proposed items and the general characteristics of the test (e.g., test purpose, physical education content domain, test content domain, and test intended uses).

We particularly want your feedback because of your experience in the field of physical education. It is estimated that the entire process of rating these items will take approximately 30-45 minutes.

It would be appreciated if you will go online and fill in the forms prior to <date to be completed> (please use URL above to open the forms online). Other phases of this research cannot be carried out until we complete analysis of the item review form. We would welcome any comments that you may have concerning any aspect of the Furtado-Gallagher Movement Skill Assessment. Your responses will be held in strictest confidence. Also, there will not be any link of your name with the results of this study.

We will be pleased to send you a summary of this research's results if you desire. Thank you for your cooperation.

Sincerely yours,
Ovande Furtado Jr.

APPENDIX F

PEARSON CHI-SQUARE TEST

Locomotor Subdomain				
Items / Age Group	Value	df	p-value	
	3-6 year-olds	1.111	1	.292
	7-10 year olds	.833	2	.659
Horizontal Jump	3-6 year-olds	3.333	2	.189
	7-10 year olds	4.054	2	.132
Leaping	3-6 year-olds	3.194	3	.363
	7-10 year olds	.357	2	.836
Hopping	3-6 year-olds	3.333	2	.189
	7-10 year olds	2.576	2	.276
Galloping	3-6 year-olds	1.548	2	.461
	7-10 year olds	.357	2	.836
Skipping	3-6 year-olds	5.000	3	.172
	7-10 year olds	.069	2	.966

Manipulative Subdomain

Items / Age Group	Value	df	p-value
Overhand Throwing			
3-6 year-olds	5.823	3	.121
7-10 year olds	1.046	1	.306
Approach Kicking			
3-6 year-olds	1.319	2	.517
7-10 year olds	3.598	2	.165
Side-arm Striking			
3-6 year-olds	5.694	3	.127
7-10 year olds	2.179	2	.336
Batting			
3-6 year-olds	9.831	3	.020
7-10 year olds	1.111	1	.292
Stationary Dribbling			
3-6 year-olds	4.470	3	.215
7-10 year olds	1.277	2	.528
Catching			
3-6 year-olds	1.250	2	.535
7-10 year olds	3.594	2	.166
Trapping			
3-6 year-olds	.556	3	.907
7-10 year olds	5.938	3	.115

Stability Subdomain

Items / Age Group	Value	df	p-value
Walking on low beam			
3-6 year-olds	.139	2	.933
7-10 year olds	2.784	2	.249
Dodging			
3-6 year-olds	1.632	3	.652
7-10 year olds	1.875	2	.392
One-foot balance			
3-6 year-olds	4.295	3	.231
7-10 year olds	.185	2	.912

Body Awareness Subdomain

Items / Age Group	Value	df	p-value
Body Shapes			
3-6 year-olds	1.481	1	.224
7-10 year olds	5.045	3	.169
Symmetrical/Nonsymmetrical			
3-6 year-olds	3.413	3	.332
7-10 year olds	2.386	2	.303
Body Twisting			
3-6 year-olds	1.122	2	.571
7-10 year olds	2.242	3	.524
Body Bending			
3-6 year-olds	.159	1	.690
7-10 year olds	2.589	3	.459
Front Swing			
3-6 year-olds	5.367	3	.147
7-10 year olds	2.232	3	.526

Space Awareness Subdomain

Items / Age Group	Value	df	p-value
SA Location			
3-6 year-olds	1.579	1	.209
7-10 year olds	2.847	2	.241
SA Directions			
3-6 year-olds	.833	2	.659
7-10 year olds	2.847	2	.241
SA Levels			
3-6 year-olds	1.740	2	.419
7-10 year olds	3.131	3	.372
SA Pathways			
3-6 year-olds	.357	1	.550
7-10 year olds	2.576	2	.276
SA Extensions			
3-6 year-olds	1.944	3	.584
7-10 year olds	2.906	3	.406

Effort Awareness Subdomain

Items / Age Group	Value	df	p-value
EA Time			
3-6 year-olds	2.176	2	.337
7-10 year olds	2.212	3	.530
EA Force			
3-6 year-olds	5.714	3	.126
7-10 year olds	2.212	3	.530
EA Flow			
3-6 year-olds	1.667	3	.644
7-10 year olds	1.076	2	.584

Relationship Awareness Subdomain

Items / Age Group	Value	df	p-value
RA with objects and/or people			
3-6 year-olds	.714	2	.700
7-10 year olds	1.517	2	.468
RA with people only			
3-6 year-olds	1.389	2	.499
7-10 year olds	1.111	2	.574

Set of 7 Questions

Items / Age Group	Value	df	p-value
Question 1	.845	3	.839
Question 2	2.917	3	.405
Question 3	.150	2	.928
Question 4			
3-6 year-olds	2.639	3	.451
7-10 year olds	3.929	3	.269
Question 5			
3-6 year-olds	2.857	4	.582
7-10 year olds	4.152	4	.386
Question 6			
3-6 year-olds	4.292	4	.368
7-10 year olds	7.450	4	.114
Question 7			
3-6 year-olds	3.333	3	.343
7-10 year olds	4.491	4	.344

APPENDIX G

COMPLETE VERSION OF THE LITERATURE REVIEW

5.0 INTRODUCTION

The ultimate goal of every physical education program is to help children and youth acquire the knowledge, skills and self confidence they need to participate in a wide variety of physical activities during their schools years and beyond (Buschner, 1994 in Nikolay, Grady, & Stefonek, 1997, p.33). For years researchers have been linking a sedentary lifestyle to health problems such as high blood pressure, heart disease, and obesity. Despite these warnings, the last report of *The American Surgeon General* indicated that Americans become increasingly inactive with each year of age. The U.S. Centers for Disease Control and Prevention (CDC) recently released data showing that 15 % (almost 9 million) of individuals in the age group of 6-19 are overweight (CDC, 2004). This represents a 4% increase from the overweight estimates of 11 percent obtained from previous data collected in the period of 1988 to 1994. Even though this is mere speculation, one could relate such results with the estimation that, on average, 35% of youth age 10- to- 18 years choose to terminate their participation in sport each year (Gould, 1987 in Brustad, Babkes, & Smith, 2001).

If adopting an active lifestyle is so important, then why are so many youths not vigorously active on a regular basis? Sport psychologists (Brodkin & Weiss, 1990; Brustad et al.,

2001) point to many factors such as premature sport involvement; competitive stress; lack of self-confidence, etc. that can account for this; however, the lack of skills competence to participate in physical activities and sports is certainly important and it is the scope of this research. There is reason so believe that youngsters are particularly disposed to drop out of sports and participation in physical activity if it becomes apparent that they do not possess as much ability/skills as others (Roberts, 1984 in Brustad et al., 2001). In many papers youngsters cite reasons that are not direct related to ability for dropping out of sports. However, Roberts believe that the stated reasons may be superficial and social desirable explanations that mask underlying perception of low ability.

Presented next is a scenario that might help one understand the underlying issues associated with this lack of interest in physical activity. Cullen is a 6th grade student at Fall School who did not have the opportunity at an early age to master several fundamental movement skills (FMS) such as running, skipping, throwing, etc.. Neither had he acquired functional understanding of movement concepts (space awareness, relationship awareness, etc.) that is believed to help in the acquisition and performance of FMS. Now, Cullen has difficulties playing sports at school or even participating in game-like activities. The scenario presented above is a reality in many schools throughout America.

A combination of a number of factors may be associated with this lack of skill competence. Some of these factors are, lack of opportunity to experience (at school or home) a diversity of movement skills, feeling of incompetence, lack of enjoyment, and the failure by teachers to test students with the purpose of (1) monitoring individual progress; (2) better adjust instructional plans to fit students needs; and (3) detecting eventual deficits in movement concept understanding, as well as, problems in fundamental skill development. This last factor

(assessment) is the scope of this work and the reason why the current assessment tool is being proposed.

A wide variety of assessment tools have been developed over the past years with the intent of providing some degree of information regarding levels of motor skill proficiency of individuals. Among these instruments are those devoted to measure levels of performance of FMS which are defined as common motor activities (e.g., walk, run, skip, throw, stationary dribbling) having specific movement patterns (Gabbard, 2000). Further, these basic skills form the base upon which more complex (e.g., sport/specialized) skills develop (Wickstrom, 1983). If FMS are addressed in various assessment tools, the same is not true with respect to movement concepts. Just a few checklists (not validated) are provided in some recent text books (Chepko & Arnold, 2000; Gallahue & Donnelly, 2003; Graham, Holt/Hale, & Parker, 2004). [Table 23](#) shows some of the most common instruments developed to assess process (qualitative characteristics of movements) rather than product (quantitative characteristics of movements) characteristics of gross motor skills that include FMS as part of their assessment batteries.

Table 23: Most common assessment tools to assess FMS

Test name	Author(s)	Last edition
Developmental Sequence of Motor Skill Development	Seefeldt and Haubenstricker	1976
Ohio State University Scale of Intra-Gross Motor Skill	Loovis and Ersing	1979
Test of Gross Motor Development	Ulrich	2000
The Fundamental Movement Pattern Assessment Instrument	McClenaghan and Gallahue	2003

Each one of the assessment tools shown in Table 1 states the importance of helping the movement science or physical education professionals better understand changes in the motor development of individuals and assisting in learning skills. Despite their wide usage, these assessment tools are not free of criticism. Failure in providing items that test movement concepts and stability skills; as well as, items that are emphasized in the National Standards for Physical Education (NASPE, 2004) are some of the shortcomings of the assessment tools cited above.

This last shortcoming, for example, makes it difficult for such assessment tools to be used aligned with instructional practices in physical education. The importance of this link between assessment and instructions is addressed in the “No Child Left Behind” (NCLB) act which became national law in January 8, 2002. One of the main issues discussed in this new law is the importance of assessment as a way to improve instructional programs and student’s learning. To accommodate the new law, schools will be required to assess students regularly and keep track of students’ progress from year to year.

Therefore, the purpose of this work is to develop a criterion-referenced assessment tool that will test movement concepts and fundamental movement skills of children ages 3 through 10 and provide evidence for content-related validity. An attempt is made to develop an assessment tool that is tied closely to the postulates of the National Standards for Physical Education (NASPE, 2004). In addition, to overcome the shortcomings of other tests, movement concepts and stability items are proposed as part of the assessment battery. The gathering of evidence for content-related validity will be accomplished through judgments of a panel of content experts and will set the stage for the development of a final list of items. Students’ performance will be compared to pre-established criteria for verification of levels of mastery in the attributes tested.

This assessment tool will be divided into 7 main areas (subdomains): (a) body awareness; (b) space awareness, (c) effort awareness, (d) relationship awareness, (e) locomotor skills, (f) manipulative skills, and (g) stability skills. Areas (a) through (d) will test students' general understanding on movement concepts, whereas areas (e) through (g) will test proficiency on fundamental movement skills.

The proposed intended uses of this test will be to (1) monitor individual progress during or following instruction, (2) evaluate effectiveness of the instructional program with the intent of adjust it according to students' needs, and (3) to detect eventual deficits in movement concept understanding, as well as, problems in fundamental skill development.

The following sections of the review will discuss previously published tests, provide a basis for the categories included in the new test and detail test development procedures.

6.0 LITERATURE REVIEW

To set the background for the development of the current test four tests that have been used to measure children's motor skill development will be evaluated. This historical review and reviews of the elementary physical education text books provide the foundation for the second section which provides the categories for the new test. The third, fourth and fifth sections review issues related to test development.

6.1 ASSESSMENT TOOLS AVAILABLE FOR TESTING FMS

A variety of assessment tools exist to assess the performance of FMS of children. Only those instruments developed to examine skill process³, rather than product, that can be applied within school settings are discussed in the current section. The most common of these assessment tools are: the Developmental Sequence of Motor Skill Development (Seefeldt & Haubenstricker, 1976); the Ohio State University Scale of Intra-Gross Motor Skill (Loovis & Ersing, 1979); the Test of Gross Motor Development (Ulrich, 2000); and The Fundamental Movement Pattern Assessment Instrument (McClenaghan & Gallahue, 2003).

6.1.1 The Developmental Sequence of Motor Skill Inventory (DSFMSI)

The authors of the DSFMSI (Seefeldt and Haubenstricker), who are proponents of the total body approach to testing fundamental motor skills, released this instrument in 1976. The instrument

³ Process-oriented assessment tools are designed to examine the process (quality) characteristics of movement; whereas, in the product-oriented tests the examiner is interested in the performance outcome (quantity) or the product of the behavior (Gabbard, 2004). More about these two approaches are discussed later in the chapter.

evolved from studies done at Michigan State University on the identification of developmental sequences within selected skills. In fact, many of the instruments developed today to assess fundamental motor skills use the developmental sequences suggested by Seefeldt and Haubenstricker. Each sequence consists of 3 to 5 stages of observable behaviors which can be used to classify students according to their level of development. The stages range from immature (stage 1) to mature (stage 4 or 5).

Validity and reliability have not been reported by the developer of the DSFMSI; however, the developmental sequences are based on longitudinal and cross-sectional data collected as part of the ongoing Michigan State University Motor Performance Study (Gabbard, 2004).

6.1.2 The Ohio State University Scale of Intra-Gross Motor Assessment (SIGMA)

The SIGMA (Loovis & Ersing, 1979) is a criterion-referenced assessment tool designed to assess eleven fundamental motor skills of children of preschool to 14 years of age. Some of the skills tested in the instrument are: skipping, striking, kicking, running, ladder and stair climbing. Each skill is presented in four developmental levels.

Gabbard (2004) points out that the instrument's psychometric properties have been questioned, except for face validity with the literature. Still, the authors report that the SIGMA can be administered with acceptable accuracy in formal testing situations or an informal free play setting. Content validity was determined by a panel of 11 experts who used a 5-point Likert-type scale to rate the test for understandability and usefulness and by documentary analysis of the literature (Payne & Isaacs, 1998).

One unique advantage of the SIGMA over other process-oriented instruments is that it provides a link to program intervention through the Performance Base Curriculum (PBC). The PBC creates a link between assessment and program intention in the sense that it provides activities to accommodate each developmental level within each skill tested.

6.1.3 Test of Gross Motor Development (TGMD)

Perhaps the most used process-oriented assessment tool to test FMS is the Test of Gross Motor Development – TGMD-2 (Ulrich, 2000). It is considered one of the most practical and simplest to administer instrument that examines the qualitative aspects of motor skill behavior. The 2000 edition was released to overcome some of the shortcomings raised by test reviewers from the previous edition (1985). The test is designed to assess children 3 to 11 years of age. The test assesses 12 locomotor and manipulative movement skills. Some of these skills are: running, leaping, galloping, kicking, ball rolling, two-handed striking, etc. The author states in the manual that there is no need for special training for test administration and each child can be tested in about 15 minutes. One characteristic that differentiates the TGMD-2 from other tests is that it provided both norm and criterion/domain interpretations. Several studies were conducted to collect evidence for both validity and reliability.

The content validity of the TGMD was established by having three content experts judge whether the specific gross motor skills selected represented skills that are frequently taught to children in preschool and early elementary grades (Ulrich, 2000). The author also, asked judges to rate the degree to which the skills were representative of the gross motor skill domain. The results showed that judges were unanimous in declaring the skills as representative of the gross motor skill domain and frequently taught to this age group.

Zhu and Cole (1996) used the many-faceted Rasch model to recalibrate and test the appropriateness of the items for the TGMD (Ulrich, 1985). Goodness-of-fit statistics⁴ demonstrated that only 8 items (performance criteria) were misfit. Thus, most of the items were well defined and were measuring a similar trait (gross motor skills). Another way to test appropriateness of items in a test is by analyzing items difficult level. This analysis indicated that overall, items were well spread within the categories and subtests, except for ball bouncing, skipping, sliding and leaping. The items (performance criteria) in these subtests (skills) were too close one to another in terms of difficulty level providing limited information to test users (Zhu & Cole, 1996). According to Nitko (2001) items that are too difficult or too easy provide little information about examinees' performance. Thus, interpretation of results might be extremely affected in these cases. Zhu and Cole suggested that the author of the TGMD review the items (performance criteria) for ball bouncing, skipping, sliding and leaping. Despite the problems raised by Zhu and Cole's study, the second edition of the TGMD was released in 2000 without addressing such issues.

6.1.4 The Fundamental Movement Pattern Assessment Instrument (FMPAI)

The FMPAI is also an easy to use assessment tool that can be applied to test the qualitative aspects of fundamental movement behavior of students in the school setting. The test was first published by McClenaghan and Gallahue in 1978. The newest version of the FMPAI was released in 2003 by Gallahue and Cleland. The first version of the instrument included five fundamental movement skills (running, horizontal jumping, throwing, catching, and kicking). Gallahue and Donnelly (2003) incorporated several additional items in their expanded version of

⁴ Provide an index to examine the extent to which examinees with low ability obtained lower scores on an item and examinees with high ability obtained high scores on any item.

the instrument. Some of these skills are: walking, vertical jumping, body rolling, dodging, one-foot balance, etc. The authors refer to the test as an informal instrument used to classify individuals at the initial, elementary, or mature stage of development in several fundamental movement skills. The instrument is best used to assess movement change over time since it compares student results to pre-established criteria rather than group norms.

Unlike the TGMD that provides norm and criterion interpretations, the FMPAI does not give quantitative scores, nor can it be used to compare one child or group of children to another child or group (e.g., between-individual or between-group comparisons). One advantage of the FMPAI over the TGMD is that the former provides a means for testing both FMS using the total and the component body approach. The total body assessment helps one getting the general picture of the group's level of ability and to identify children who are experiencing difficulty; whereas, the component assessment allows the teacher to pinpoint exactly where the problem lies (Gallahue & Donnelly, 2003). More about the differences between these two approaches is provided later in the chapter.

This section addressed several assessment tools used to test fundamental motor skill behavior of children. The purpose of the current test is to test not only fundamental movement skills, but also movement concepts. However, there are no valid assessment tools available for testing such important components of motor skill development. A few checklists are available in text books (Chepko & Arnold, 2000; Gallahue & Donnelly, 2003; Graham, A., & Parker, 1998; Graham et al., 2004) that cover some of the movement concepts present in the literature.

The tests discussed above provide the foundation for the assessment tool developed here. In addition to these tests a recent national legislation, the act "No Child Left Behind" is

addressed. Current trends are for the evaluation of quality physical education of which this act can provide guidance.

6.2 THE “NO CHILD LEFT BEHIND” ACT AND ITS TIE WITH ASSESSMENT

Signed into law on January 8, of 2002, the “No Child Left Behind” (NCLB) act intends to fill an achievement gap with regards to accountability, flexibility, and choice in the educational system in the U.S. in kindergarten through 8th grade, so that no child is left behind. The document that describes the law is long and addresses several issues that should be improved in education; however, only those related to assessment are addressed in this section. In several instances the NCLB act addresses the importance of improvement in assessment tools and procedures currently used in school settings. The issues concerning assessment in the NCLB document addresses two main topics: (1) the need for a more coherent assessment system so that student progress can be tracked from year-to-year, and (2) increasing the ways by which assessment results are shared with parents and students.

The NCLB act urges schools to develop an assessment system that is coherent in the sense that it allows tracking student progress from year-to-year. The accomplishment of this task depends on two main issues: the availability of assessment tools that offer easy and efficient recording, storage and retrieving of data, and the teachers’ willingness in keeping assessments up-to-date. To keep assessments up-to-date teachers have to be committed for the task making this a common practice in the school settings. There are several advantages for such practice including improvement of instructional programs, the possibility of detecting students that need learning reinforcement, etc. Therefore, in addition to their commitment to instruction physical

education teachers throughout the country have to be committed and willing to assess students' improvement. Besides the teachers' willingness, the availability of assessment tools carefully and exclusively designed for the purpose of tracking students' improvement from year-to-year is necessary. Assessment results should be stored in a way to facilitate retrieving and easy comparisons of data. One might think that Microsoft Excel⁵ will work. However, not only are specific technical skills required but the user would have to spend time preparing the spreadsheets in Excel in order for it to work as expected. But what if teachers could choose assessment tools that have such capabilities saving them from this extra work? Gallahue and Donnelly (2003) argue for the benefits of using computer applications for assessment in physical education. The problem is that computer software packages are limited in some subject areas in physical education. Fitness assessment has been receiving significant attention by computer testing developers. These computerized fitness assessment software allow for class data, generation of individual reports, tracking individuals over time, and for reporting and prescribing personalized fitness profiles. The most used ones are, the President's Challenge Software (www.msfitnesstracker.com); the Fitnessgram Software (www.americanfitness.net) and the Blackport Physical Fitness Test Software (www.americanfitness.net). If in the area of fitness there are several options in terms of computerized assessment tools, the same is not true in the area of motor skills assessment. Thus, there is a need for the development of computerized assessment tools with the intent of testing children's motor skills. Thus, one of the objectives of the current work is to fill this gap in fundamental movement skill assessment.

An important issue was discussed above that deals with the need for development of a more coherent assessment tool allowing for easy tracking of students' improvement from year to year. Another issue that is addressed in the NCLB act document is the improvement of the ways

⁵ This is a spreadsheet program from Microsoft that is part of the Microsoft Office suite.

assessment results are shared with parents and students. The most common and traditional way to share assessment results with parents and students is through written reports developed by the teacher responsible for administering the tests. Written reports may include, previous and current assessment results with charts showing students' improvement. Test objectives and target goals may also be part of these reports. The only problem is that it takes a significant amount of time for gathering data and making these reports available for parents and students. Thus, a computerized motor assessment tool may be the best solution for such a problem. Computerized assessment tools allow for easy data recording (when the software package is available for PDA Handhelds devices). This would provide safe storage of information, and most importantly a means to easily compile data in order to generate graphic reports containing information for parents and students. In addition, sharing of data can also occur over the World Wide Web making the interaction between teachers, students and parents more efficient.

In short, two main issues were discussed in this section that relates to the NCLB act: (1) the importance of creating assessment tools in education that can easily track students performance from year-to-year, and (2) the need for increasing the ways by which assessment results are shared with parents and students. Suggestions were made for possible solutions. Next, issues regarding domain specifications are discussed.

6.3 DOMAIN SPECIFICATIONS

The process of gathering evidence for content-related validity requires a detailed discussion of the domain(s) being measured. The items that form the current test come from two distinct domains: movement concepts and fundamental movement skills. These two domains are not

discussed individually, but rather within a framework called developmental movement approach. Such an approach reinforces the concept of individual differences rather than age-group appropriateness. This model is well defined by Gallahue and Ozmun (1998) when they state that;

“The developmental physical approach to physical education aims to instruct individuals in the use of their bodies so that they can execute a wide variety of fundamental movements efficiently and effectively throughout life, and apply these basic abilities to a wide range of specialized movement skills for daily living, recreational, or competitive sport needs” (Gallahue & Ozmun, 1998, p. 465).

This developmental curricular model views learning as the interaction of three things; biology of the individual, the goal of the task and the conditions of the environment. The movement analysis framework used in this approach is based on the three movement categories into which movement may be classified and the specific movement skills that may be extracted from these categories (Gallahue & Ozmun, 1998). The three movement categories are: fundamental stability movement skills, fundamental locomotor movement skills, and fundamental manipulative movement skills. An example of a fundamental movement skill is running which is placed under the category of the fundamental locomotor movement skills.

Movement concepts are also part of the movement analysis framework used in the developmental model. Movement concepts may be divided into four main categories: body, space, effort, and relationship awareness. Therefore, movement concepts and fundamental movement skills are the two domains assessed in the current assessment tool. A more complete discussion of each domain is provided below. Because movement concepts should be taught first it precedes the discussion of the fundamental movement skills.

6.3.1 Learning movement concepts

Movement concept learning relates to how the body *can* move (Gallahue & Donnelly, 2003). Participation in sports or game-like activities is dynamic in nature and can be quite demanding. Students are required to constantly change the way their bodies move. “The dynamic nature of movement, the ever changing environment, and the situational requirements of the task make it absolutely essential for children to learn how their body can move” (Gallahue & Donnelly, 2003’ p. 108). However, one cannot teach nor assess every possible variation of movement used in a given task. Instead, students are required to learn and be able to demonstrate their cognitive and functional understanding of movement concepts that can be applied to various situations (Graham et al., 2004).

Taking into consideration the importance of movement concepts when learning, it is quite a surprise that there has not been any attempt to develop and validate an assessment tool to test such concepts in physical education. A few checklists are available in some physical education textbooks (Gallahue & Donnelly, 2003; Graham et al., 2004; Kogut & National Association for Sport and Physical Education., 2003). The various aspects of movement concept leaning are discussed below.

6.3.1.1 Cognitive vs. functional understanding of movement concepts

Simply knowing the meaning of movement concepts (cognitive understating) is not enough. Students need to be able to demonstrate functional understanding of movement concepts through actual performance. A student might, in reality, know the meaning of a given concept, but not be able to apply that concept while performing motor skills. Knowing whether or not students have attained functional understanding of movement concepts can be accomplished through administration of valid tests carefully design for this purpose. Unfortunately, such tests are not available yet, and one of the objectives of the current test is to fill this gap.

Cognitive understanding can be tested by asking students to describe in words the aspects of movement concepts. The examiner could ask the difference between, self space and general space, or free and bound, etc., and evaluate students' responses. If assessing cognitive understanding of movement concepts might be considered easy, the same is not true for functional understanding.

Assessing functional understanding of movement concepts is more complicated. Functional understanding means the students know and are able use the concepts in their movement (Graham et al., 2004). Students are required to demonstrate, through movement, their functional understanding of the concepts. Often measurements in movement concepts are done using nominal scales (e.g., pass/fail or 0/1). This is because levels of skill proficiency are not part of movement concepts, since they are concepts to be understood, not skills that have to be mastered (Graham et al., 2004). For example, students are asked to perform a skill (e.g., basic movement skill of stationary ball dribbling) at different levels (high/medium/low). Because the emphasis is upon the students' functional understanding of the movement concept of "levels" the students' proficiency in the skill of stationary ball dribbling is of secondary importance. Examiners should be looking at whether or not the students are able to perform what is being asked in the task which in this case is bouncing the ball at the level of the head, waist and knees. Next a discussion is provided to address the issue of combination of various movement concepts with movement skills.

6.3.1.2 Combination with other movement concepts and skills

In real world situations a movement concept (e.g., general space) is used in combination with other movement concepts and/or movement skills. Two issues are important: (1) the way

movement concepts should be taught and assessed and (2) how movement concepts improve the quality of movement skill performance with such combination.

Teaching strategies are briefly discussed to make the link to assessment. First, practitioners should employ an easy-to-difficult approach when teaching movement concepts to students, especially to younger ones. Movement concepts can be taught individually and as students demonstrate both cognitive and functional understanding they can begin participating in activities that require some sort of combination of movement concepts and movement skills. For example, students are taught the notion of self and general space. Then, separately, students could be instructed with the concepts of pathways (e.g., curved/straight/zigzag, etc.), directions (e.g., forward/backwards, clockwise/counterclockwise, right/left, etc.) and levels (e.g., low/medium/high). Next, assessments could be administered to ensure students' cognitive and functional understanding of the concepts. Once a desirable level of understanding is reached, the teacher can begin to develop activities that combine different movement concepts with various types of movement skills. For example, students can be asked to run, gallop or skip in general space (location) in different directions and levels.

Another aspect of movement concepts is that they can be used to improve the quality of movement skill performance. One can ask a child to run in a straight line across the gym. However, this same skill could be performed backwards, sideways, or zigzag, either very slow or very fast. These are some examples of movement concepts that can be used to modify the skill of running, consequently enriching the performance of such skill. Further, in real world situations movements are rarely performed under the same conditions time after time (Gallahue & Donnelly, 2003). Basketball can be used as an example. When playing basketball students have to apply different concepts to perform the skill of dribbling. Certainly, it is not wise to simply

run straight while dribbling the ball in the court. The performer is required to dribble in zigzag, backwards, sideways, fast or slow in order to avoid losing the ball to the opponent. Next, the four subdomains (behavioral objectives⁶) of movement concepts (body awareness, space awareness, effort awareness, and relationship awareness) and their respective subcategories (measurement dimensions⁷) are discussed.

6.3.1.3 Body Awareness: What the body can do

Body awareness is the individual's ability to know what actions the body is capable of doing and understand how these actions are performed (Doherty & Bailey, 2003). It is critical that children understand what the body does while moving. This can help them to demonstrate, while moving, many ways to move individual body parts while controlling other body parts (Buschner in Human Kinetics (Organization) & Pettifor, 1999). The current test will measure two body awareness concepts: body shapes and nonlocomotor movements. Body shapes can be curved, twisted, narrow, wide, and symmetrical/asymmetrical. These are concepts that can be used in the performance of any locomotor, manipulative and stability skill. It is reasonable to assume that before children start understanding the concept of body shapes they need to be able to identify specific body parts. After they are comfortable differentiating among body parts, one can introduce the notion of body shapes and the relation to one another (Graham et al., 2004). For example, one could ask children to travel and then, upon the teacher's signal, stop in a wide shape. Next, rather than just stopping in a wide or narrow shape the students have to make a distinction between symmetrical and nonsymmetrical body forms. The quality in the

⁶ Behavioral objective is a term that describes specific areas within a domain having particular attributes (skills) that measure its content. This term will be used interchangeably with the term "subdomain" throughout this work.

⁷ A measurement dimension is a subcategory under a behavioral objective. It is used to specify all possible attributes measured under a subdomain.

performance of these skills depends strongly on the individual's ability to functionally understand the processes involving to control body shapes (Doherty & Bailey, 2003).

Nonlocomotor actions involve movements done while the body remains in one location (Thomas, Lee, & Thomas, 2003). Some examples of nonlocomotor actions are stretching/extending, twisting/turning, bending/curling/flexing, swaying/swinging, twirling, sinking, and pushing/pulling. Some authors (Gallahue & Donnelly, 2003; Graham et al., 2004) place the nonlocomotor category under the fundamental movement skills treating it as stability skills. The current test will adopt a pattern used by other authors (Human Kinetics (Organization) & Pettifor, 1999; Thomas et al., 2003) in which nonlocomotor actions are considered part of the subdomain of body awareness. Also the developers of the current test decided to keep static and dynamic balance, sometimes placed under the nonlocomotor category, as part of the stability skill under the fundamental movement skill domain. Next, concepts involving space awareness are addressed.

6.3.1.4 Space awareness: where the body moves

Understanding space awareness is crucial in physical education and consequently for the process of learning and practicing movement skills. Children that understand and apply (functional understanding) the concept of space awareness are more likely to move safely as they travel through the environment (Graham et al., 2004). Thus, assessment of space awareness is extremely important to ensure students are internalizing such concepts that will further be used in games, gymnastics and dance. The literature that discusses movement concepts is consistent with regards to the concepts included in the space awareness category. However, the same consistency is not observed in terms of how these concepts are organized into subcategories. For example, Gallahue and Donnelly (2003) use three subcategories under space awareness which

are *Levels, Directions* and *Ranges*. Yet Graham et al. (2004) are more specific in their attempt to explain space awareness breaking it into five subcategories including *Location, Directions, Levels, Pathways* and *Extensions*. Because the five-subcategory approach seems to be less confusing and more detailed, it will be used throughout the current work to describe space awareness. Next a discussion of each of these subcategories is provided.

Location can be divided into two components: self and general space. Self-space (personal space) is the area immediately surrounding one's body, whereas general space is defined as the total available area one has to perform a certain task (Gallahue & Donnelly, 2003). Additionally, one finds in the literature the term "restricted space" which is a specific area within a general space that is limited or prescribed in which one may move. Asking a child to extend his/her arms away from the body as far as possible while standing is an example of exploration of one's self space. Exploring general space means making use of the available area to move without bumping into someone or losing control of movements (Graham et al., 2004).

The second subcategory is *Directions*. "Directions in space are the dimensional possibilities into which the body or its parts move or aim to move-up and down, right and left, forward and backward, clockwise and counterclockwise"(Graham et al., 2004, p. 248). Understanding the concept of "directions" helps in the process of enhancement of awareness of the body as it is projected into space (Gallahue & Donnelly, 2003). Cognitive and physical maturation play an important role when assessing students' understanding about the concept of directions. Thus, it should not be a surprise to see pre-school children having difficulties understanding and putting into practice the directions of left and right and clockwise and counterclockwise as opposed to forward and backwards (Graham et al., 2004).

Levels is the third subcategory under spaces awareness. *Levels* can be seen as the horizontal layers in space where the body and its parts are situated or can move (Graham et al., 2004). If some concepts of Directions are strongly dependent on cognitive and physical maturational factors, the same may not be true for the concept of Levels. With proper instruction even pre-scholars are able to functionally understand the concept of Levels which are divided in low, medium and high. Conventionally, the space below the knees is considered low level. Crawling is an example of a locomotor action performed at a low level. The space over the shoulders is considered high level. Graham et al (2004) emphasizes that although one cannot move the whole body into high level, there are some actions in which body parts are brought to a high level. Stretching the arms up is an example of an action performed at high level. Finally, medium level consists of the space between the low and high level. In another words, the space between the knees and shoulders. The contact hand-ball while dribbling a basketball, for example, typically occurs in middle level.

Pathways is the fourth subcategory of space awareness. Pathways are designs made by someone's body moving in space – either in the air or on the floor (Allison & Barrett, 2000). Pathways can be either straight or curved (circle, zigzag, etc.). Besides the performer's body in space, one can think of pathways in terms of the trails of objects by performers (ball, hockey puck). It is important that performers understand that a curved or angular pathway may in some cases play a key role in avoiding collisions when traveling through a crowd (Graham et al., 2004).

Extensions is the fifth subcategory discussed and it has to do with spatial actions done in space with relation to the body and body parts (Graham et al., 2004). Children should be able to differentiate between near/far and small/large extensions (movements). Using our own hands to

explore all the space close or away from our body is an example of near/far extensions. Yet small/large “extensions” can be demonstrated by asking one to travel in a limited area. Upon the teacher’s signal, the performer uses a smaller extension to stop with all body parts as close to the body as possible. On the other hand, the performer demonstrates a larger extension by jumping , into the air, extending arms and legs as far away from the body as possible.

The issue of space awareness was addressed in this section. For easy understanding, space awareness was divided into five subcategories (*Location, Directions, Levels, Pathways* and *Extensions*). It is important that students develop early the ability to occupy an area or to travel in an area while maintaining awareness of others (Graham et al., 2004). For this reason, the concept of location (self-space and general space) should be taught first to children who have had little experience in formal physical education instruction. In the next section the second category of movement concepts, effort awareness, is discussed.

6.3.1.5 Effort awareness: How the body can move

If space awareness deals with *where* the body can move, effort awareness defines *how* the body can move (Gallahue & Ozmun, 1998). The concept of effort awareness is divided into four measurement dimensions: *Force* (strong/light); *Time* (fast/slow); *Flow* (bound/free); and *Space*. When observing proficient movers performing a certain sport activity one has the sense that most of the actions have the appropriate amount and degree of time, force and flow. As an example of effort awareness, Cameron is just 8 years old, but he strikes a baseball like a “professional”. Most of Cameron’s precision when striking a baseball has to do with his functional understanding of the concept of effort. His awareness of this concept gives him the ability to apply the correct level of force at the right time while swinging the bat to hit the ball. Next, a discussion of the categories of effort awareness is provided.

Time is one of the most important concepts that children must internalize to ensure effectiveness of their movements. Time is the speed at which movement takes place (Gallahue & Ozmun, 1998). Movements may be fast in nature (e.g., running, dodging, etc.); and sometimes movements take on a more slow characteristic (e.g., walking on a balance beam, leaping, etc.). Regardless of the situation students should be exposed to a variety of experiences and assessed periodically on the concept of time. Thus, changes in timing of movement will occur without forethought as students adapt to different situations (Graham et al., 2004).

Force is the second subcategory under the concept of effort awareness. In early learning children usually have a hard time applying the right amount of force while performing a certain task. Have you tried to ask a 6 year-old child to return a soccer ball by kicking it back to you? Too often you will notice that he/she either kicks the ball applying too much or too little force. This lack of consistency in kicking the ball with the appropriate force is partially due to the failure by the child in understanding (functionally) the concept of force. Gallahue and Ozmun (1998) define force as the degree of muscular tension required to move the body or its parts from place-to-place or to maintain its equilibrium (Gallahue & Ozmun, 1998). Controlling muscular tension can also be understood as the force applied to an object while performing a given task (e.g., the soccer example given earlier).

The third subcategory discussed under effort awareness is *flow*. Flow relates to the continuity or coordination of movements (Gallahue & Ozmun, 1998). Flow can be divided into “free” and “bound”. An example of a free flowing movement is when the performer is incapable of stopping a movement after it has been started. Hitting a tennis ball is an example of free flow. Once it is initiated little can be done to stop it before ball impact. On the other hand, bound movements are stoppable actions. Herein, the performer is in control of at all times (Graham et

al., 2004). Performing a slow cartwheel with a pause for a handstand before traveling is an example of bound action.

Finally, *Space* is the fourth measurement dimension of effort awareness. Space refers to the direct or flexible use of space. Individuals performing direct movements occupy less space when compared to flexible movements. Flexible movements are bendable, circuitous; whereas, direct movements are straight and undeviating (Allison & Barrett, 2000).

In the current section the concept of effort awareness was discussed. This category of movement concept has to do with *how* the body can move. The category of effort awareness was broken into four measurement dimensions (*Time, Force, Flow* and *Space*) and a discussion was provided for each of these dimensions. Next, the fourth and last category of movement concepts (relationship awareness) is addressed.

6.3.1.6 Relationship awareness: moving in relation to objects and people

Gallahue and Ozmun state that the concept of relationship awareness deals with both “how and where” the body moves in relation with objects and people (Gallahue & Ozmun, 1998). The developers of the current test subdivided this category into two measurement dimensions. These are Relationship with Objects and/or People and Relationship with People Only. It is important to notice that the concept of relationship awareness gives meaning to the interaction between individuals and their environment (Graham et al., 2004).

Having a functional understanding of the concepts of the Relationships with Objects and/or People is crucial for efficient movement. This is because in real life situations individuals have to perform in an environment that is unpredictable and full of variables. The concepts discussed in this section can be applied to objects and or people; however, some have more meaning when used specifically in relation to objects or people only. For example, the concepts

of on and off, along, through, over, under, around and surrounding make more sense when applied with relation to objects. On the other hand, the concepts of near and far, in front, behind, and alongside are generally applied to relationships with people (Graham et al., 2004). The concepts addressed previously may be used for objects and people. Concepts that are used exclusively with relation to people are addressed below.

Relationship with people only involve, for example, the action of meeting and parting. This is basically moving toward or away from a partner. Also, one could consider the concepts of matching (partners are side-by-side) and mirroring (partners face each other). The former has to do with duplication of a partner's movement in real time. Whereas, mirroring is the reverse reproduction of the partner's movements as if looking in a mirror (Graham et al., 2004). Lastly, leading and following are also considered concepts of the relationship with people.

This section provided a discussion about movement concepts. The section covered one of the domains that are assessed in the current assessment tool. For a better understanding, movement concepts were divided into 4 major categories (space awareness, effort awareness and relationship awareness. In addition, the characteristics of each of these categories were discussed. Next, the second domain assessed by the current test is discussed.

6.3.2 Fundamental Movement Skills

As mentioned before, this assessment tool is being developed to be used in practices that are aligned with the developmental movement approach, which uses a movement analysis framework based on movement concepts and fundamental movement skills. Issues regarding movement concepts were introduced above. Next, a discussion of the second domain assessed by the current test which is fundamental movement skills is provided.

One of the earliest works to study FMS was done by Ralph L. Wickstrom. He spent several years investigating various types of fundamental (basic) movement skills. The author's effort in collaborating for a better understanding of human motor behavior resulted in the release in 1970 of a book titled "Fundamental Motor Patterns". Although the third and last edition of this book was published in 1983, it is still a valuable source of information aiming researchers in the investigation process of motor skill development. Wickstrom defines FMS as "common motor activities with specific patterns. They are the general skills that form the bases for the more advanced and more specific motor activities, such as sport skills" (Wickstrom, 1983, p. 7)

The work of Wickstrom was further extended leading to the development of valuable research in the field of motor behavior analysis. The skills studied by Wickstrom were later divided in categories: Locomotive, Manipulative and Stability (Table 24). A category of movement is a classificatoion scheme based on common underlying principles of movement (Gallahue & Ozmun, 1998). Thus, motor skills that share the same characteristics are placed under the same category of movement. For example, the fundamental movement skill of catching and striking are placed under the movement category of manipulation. This is because they both involve the manipulation of some sort of object.

Below is a discussion of the movement categories and the respective fundamental movement skills that fit under each category.

Table 24: Selected fundamental movement skills.

Fundamental Stability Movement Skills	Fundamental Locomotor Movement Skills	Fundamental Manipulative Movement Skills
Cartwheel	Running	Throwing
One-foot balance	Jumping	Catching
Headstand	Hopping	Trapping
Body rolling	Skipping	Striking
Walking on line	Sliding	Volleying
Dodging	Galloping	Bouncing

6.3.2.1 Fundamental Stability Movement Skills

These skills are believed to form the basis for all locomotor and manipulative skill themes since all movement involves an element of stability (Gallahue & Donnelly, 2003). Any movement in which the body remains in place but moves around its own vertical or horizontal axis is considered a stability movement skill (Graham et al., 2004). Some of these skills are dodging, walking on the beam balance, tripod and one-foot balance. These skills require dynamic or static balance in which the main goal is to gain or maintain one's equilibrium against the force of gravity (Gallahue & Donnelly, 2003).

6.3.2.2 Fundamental Locomotor Movement Skills

Any skill in which the body is transported in a horizontal or vertical direction from one point to another is considered as locomotor movement skill (Graham et al., 2004). According to Gallahue and Ozmun (1998), it is important that individuals master fundamental aspects of stability before efficient forms of locomotion occur. Surprisingly, only a few assessment tools have stability skills as parts of their testing battery. Jumping, running, leaping, skipping, hopping are some of the skill themes that may be incorporated into this category of movement.

6.3.2.3 Fundamental Manipulative Movement Skills

Manipulative movement skills are regarded as skills in which individuals interact with some kind of object during its performance. Gallahue and Donnelly (2003) suggest dividing this category into gross and fine motor manipulation. In the former type of manipulative movements the emphasis is upon giving or receiving force to objects or receiving force from objects, whereas in the latter there is an emphasis on motor control, precision and accuracy. Since gross manipulative movements are the primary concern of physical educators they will be part of this assessment tool and referred to simply as fundamental manipulative movement skills. Further, it seems that proficiency in gross manipulative skills tend to occur later than both stability and locomotor skills (Gallahue & Ozmun, 1998). The authors argue that this may be partially because most manipulative movements have some elements of both stability and locomotion. Throwing, catching, kicking, and striking are some examples of such fundamental movement skills. A brief description of each movement category was given above. Next, a discussion about the importance of the fundamental movement skills for the field of motor skill development is provided.

6.3.3 The importance of the fundamental movement skills

Fundamental movement skills are assumed to provide a framework upon which more complex skills develop. For example, the mature pattern of the movement skill of stationary ball dribbling is necessary for the development of more complex skills such as dribbling a ball while moving forward or the lay up in basketball that requires a combination of dribbling a ball, running and jumping. Motor development specialists (Gabbard, 2000; Sayre & Gallagher, 2001; Thomas et al., 2003; Wickstrom, 1983) agree that it is crucial that such FMS are learned during the

“fundamental movement skill phase”. The phases of motor development are generally classified as the reflexive, rudimentary, fundamental, and specialized movement phases. During infancy and young childhood reflexive and rudimentary phases, respectively are observed. It is believed that these two phases of development represent the fundamental building blocks for the next two phases, that is, fundamental and specialized movement phases (Gabbard, 2004). The fundamental movement phase goes from about 2- to 7-years-of-age and it is believed to be one of the most important periods within the motor skill development (Gabbard, 2004; Gallahue & Donnelly, 2003). This is the period for children to master the various kinds of FMS that fall under the three categories cited earlier, that is, stability locomotive and manipulative.

This phase of development is the most prominent in the literature of motor development. Gallahue and Ozmum (1998) provide us with an interesting schematic representing a descriptive view of the phases and stages of motor development by using a representation of an hourglass. In the past, authors represented such schematic using the figure of a pyramid. In that representation the Specialized Movement Phase used to be placed on the top of the pyramid, giving a false idea that such phase represented the end of motor skill development. Thus, the use of the hourglass to represent the motor skill development gives a more meaningful representation. This is because specialized skills may be used for different purposes such as “lifelong recreational, daily living and competitive utilization”. Gallahue and Donnelly (2003) emphasize the importance of developing and refining fundamental and specialized movement skills during the crucial pre and elementary school years. According to the authors, failure to develop a mature pattern of such skills may affect participation later in life of even recreational activities. It should be noticed that skills can still be learned later in life; however, it makes it easier if these skills are mastered in childhood. Gallahue and Donnelly (2003) raise an interesting point related to this issue. The

authors claim that most of the skills that are not developed earlier in life remain unlearned. One main factor that might contribute to this is related to accumulation of bad habits from improper learning. According to the authors, it is much more difficult to “unlearn” faulty movements than to learn them in the first place. Another factor is fear. People often avoid participating in recreational sport activities simply because they are afraid of being injured or ridiculed by peers making it difficult for learning skills later in life. All this means that developing skills during the appropriate period is crucial for all individuals.

Issues related to movement concepts and fundamental movement skills were addressed in this section. Next, these two domains are discussed in relation to the National Standards for Physical Education.

6.3.4 Movement concepts, fundamental movement skills and the Standards

The National Standards for Physical Education (NASPE, 2004) was first published in 1995 and it comprised an effort to establish national content standards for physical education. The second edition of the document was released in May of 2004. This current edition of the standards reflects current thinking on what students should know and be able to do as a result of a quality physical education program (NASPE, 2004).

It was discussed in the introduction of this work about the importance of physical activity as a way of development and maintenance of good health. According to NASPE (2004), “the goal of physical education is to develop physically educated individuals who have the knowledge, skills, and confidence to enjoy a lifetime of healthful physical activity”(NASPE, 2004, p. 11). However, the question being asked is what is a physically educated person? The

answer to this question is demonstrated in 6 statements that comprise the actual national standards for physical education (Table 25).

Table 25: National Standards for Physical Education

Standard #	Description
Standard 1	Demonstrates competency in motor skills and movement patterns needed to perform a variety of physical activities.
Standard 2	Demonstrates understanding of movement concepts, principles, strategies, and tactics as they apply to the learning and performance of physical activities.
Standard 3	Participate regularly in physical activities.
Standard 4	Achieves and maintains a health-enhancing level of physical fitness.
Standard 5	Exhibits responsible personal and social behavior that respects self and others in physical activity settings.
Standard 6	Values physical activity for health, enjoyment, challenge, self-expression, and/or social interaction.

Of particular importance are standards 1 and 2. This is because the content of the current test reflects what students should know and be able to do in terms of those two standards. Standard 1 refers to the development of skills needed to enjoy participation in physical activities. These skills are widely known as fundamental movement skills and they comprise one of the two domains tested by the assessment tool that is being proposed in the current study. Standard 2 relates to the second domain being tested here, which is the domain of movement concepts.

Therefore, the current assessment tool is being developed to reflect two important national standards for physical education. This will help in the process of linking assessment of motor skills with instruction practices, which is the most current thinking in the area of physical education according to NASPE (2004).

6.4 ASSESSMENT

A brief discussion is provided next that addresses the various terms used in assessment. The first term is assessment itself. *Assessment* in education is a broad term and refers to the process of obtaining information (student's competence) that is used to make a particular educational decision (Nitko, 2001). *Measurement* is often incorrectly used interchangeably with the term assessment. It refers to a procedure for assigning scores to a particular attribute of a person. Nitko (2001) emphasizes that the process of assigning scores to individuals (measurement) maintain the order that exists in the real world among people being measured. This means that if John is better than Sally on the skill of running, then John's score should reflect this difference and consequently be higher than Sally's. *Test* should also be differentiated from the first two terms. It is defined as an instrument or systematic procedure for observing and describing characteristics of individuals (Nitko, 2001). Finally, the term *evaluation* is also sometimes confused with the terms assessment and measurement. It refers to "the process of making a value judgment about the worth of a student's product or performance" (Nitko, 2001, p. 7). Nitko also emphasizes that evaluation may occur even without measurements or test results. A teacher can make judgments (evaluate) about students based upon observation only, without using any systematic testing procedures.

The most common terms used in assessment were addressed in this section. The reasons assessments are used in physical education is addressed in the following section.

6.4.1 Why do physical education teachers assess?

Assessments are employed for many reasons in physical education, and the use of the information obtained varies. This section deals with the applicability of the information acquired from assessment procedures.

For many physical education teachers assessment exists with the exclusive purpose of assigning grades to students at the end of a unit/theme (Holt/Hale, 1999). However, assessments are conducted for other purposes, as well. Physical education teachers can use assessment for *screening*, to determine the *status* of students, and for *placement* reasons. Assessments can also be used for *program evaluation*, analysis of *program content*, and even as a *motivational* factor. Each one of these purposes is briefly addressed next.

Screening, sometimes called diagnosis, is one of the many reasons physical education teachers assess students and it is usually carried out at the beginning of the school year. Screening is the process by which students are assessed to determine if they should be referred for further testing and whether they require attending special/remedial learning programs (Payne & Isaacs, 1998). The main idea is to provide a means of differentiating normal-functional individuals from those who may not be developing normally (Gabbard, 2004).

Physical education teachers can also use assessments with the purpose of **determining status, progress or achievement** of students. This is a common practice in physical education and serves as valuable tool for providing feedback to students and teachers alike. Teachers can use assessments to verify whether learning targets are being reached (Gabbard, 2004). For example, to verify whether his students are progressing in the performance of fundamental manipulative movement skills, Iverson decides to use an assessment tool to measure the

students' proficiency at the beginning of the school year/unit and again at the end of it. Then, the two testing sessions are compared and Iverson evaluates the students' performance based on the information obtained. The information obtained can also be used as a form of feedback to the teacher.

Placement is another reason assessments are conducted in physical education. Placement or classification is the procedure used to place individuals in classes or groups according to their motor abilities (Gabbard, 2004). After a test is conducted, Iverson notices that his 4th graders are at different levels of skills performance in several fundamental movement skills. Hence, Iverson's decision is to form two separate groups. This decision can enhance teaching and learning processes of fundamental movement skills.

Iverson can also use the information obtained through assessments to **evaluate** the appropriateness of the **content** being taught. After testing his students' understanding in several movement concepts, Iverson may decide to spend less time teaching such concepts as was originally planned. Since students seemed to be very familiar with movement concepts, Iverson can begin teaching fundamental movement skills earlier than it is stated in the school program. Therefore, the teacher is using the information obtained from assessment to write program objectives that are more challenging for students. This may help to improve the motivation of students during instruction. This example illustrates a typical scenario from a developmental physical education program where the teacher is allowed to modified features of the school program to better accommodate the students needs.

Information obtained from assessment can also help program directors and administrators to **evaluate** the **program** used in schools and thereby determine the possible need for change (Gabbard, 2004). This is a common practice at the school district level but also may occur at a

national level. An example is the “No Child Left Act” in which schools have to accommodate their programs to nationally stated standards. The act was suggested because students have been scoring poorly for several years on various national achievement tests.

Finally, **motivation of students** is another reason assessments are conducted within school settings. Here, information obtained from assessing movement concepts can be shared with students. The idea is to make students aware of how well they are performing and how they can improve. Students receiving feedback about their performance are more likely to be motivated during instruction (Gabbard, 2004).

This section addressed the general issues related to assessment. Next, issues dealing with types of assessment are discussed.

6.4.2 The link between assessment and instruction

Among other things, the main goal of the current work is to develop an assessment tool that will help in the process of linking instructional practices with assessment. This will be possible since the proposed instrument is being developed based on two of the six national standards for physical education that was cited earlier.

The advantage of developing an assessment tool based on the standards is that it states what is worth teaching and learning in physical education. Learning can be enhanced when learning goals match both assessment and instructional practices (NASPE, 2004). The idea is to “tell students what is important for them to know and be able to do, teach them what you told them they would learn, design appropriate tasks that allow them to practice what you taught them, and assess them on what they have been practice” (Tannehill, 2001, p.19 cited in NASPE, 2004).

Assessments should no longer be seen as a mere way for documenting learning. Instead, both instruction and assessment should be considered two interlinked components of learning, yielding information about student progress toward the achievement of the content standards in physical education and facilitating their achievement (NASPE, 2004). The Furtado-Gallagher MSA attempts to make this link by basing its content on two of the six national standards for physical education. Next, a discussion of two types of tests (formal and informal) is provided.

6.4.3 Formal/ Standardized and informal /non-standardized tests

Tests can be divided into formal/standardized or informal/non-standardized. Formal tests provide standardized or uniform conditions and instructions to test examinees, whereas informal tests are less restrictive with regards to such characteristics (Burton & Miller, 1998). The advantages and disadvantages associated with formal and informal tests are discussed below.

Since formal tests are developed under standardized procedures, assessment results may be compared with a certain degree of confidence, even though different persons administer the same test on different occasions (Nitko, 2001). Perhaps, this is the main advantage of formal over informal tests. The TGMD (Ulrich, 2000) is an example of a formal test in which conditions and instructions were standardized during the development of the test, more specifically when the normative group was first tested. Thus, anyone willing to use the TGMD must follow all the standardized procedures available in the test's manual. If the test user does not use these standardized procedures, the test becomes an informal instrument since there is no adherence to all the manual's instructions.

Two major weaknesses of formal tests that are pointed out by Burton and Miller (1998) are discussed next. First, the main disadvantage of using a formal test according to the authors is that it may require examinees to use equipment that is standardized in terms of size and weight, even though body sizes and weights of different examinees are not similar. This may represent a problem if the age range of examinees is significantly large. An example of this is the TGMD which is designed to test children ages 3 to 10 years. In a previous version of the test (Ulrich, 1985), Ulrich collected data for the normative group by using only one ball size (8 to 10 in.) while testing the performance of several manipulative fundamental movement skills. Some reviewers addressed the problem and Ulrich made changes to accommodate specific equipment sizes and weights for the various age groups. The changes appeared in the newest version of the TGMD (2000). Instead of a standard ball size, examiners are now instructed to use an 8- to 10-inch playground ball for children ages 3- to 5-years and a basketball for children ages 6- to 10-years in the manipulative skill of stationary dribble. Therefore, test developers should standardize equipment in terms of body-scaled values in order to make test results comparable (Burton & Miller, 1998).

Another disadvantage of formal tests pointed out by Burton and Miller (1998) is the artificial nature of tasks and environment used for collecting data. One may claim that an examinee may not perform naturally when instructions are strictly standardized. The exploration of what an examinee may do in a more natural movement situation is not allowed (Burton & Miller, 1998). For example, asking a child to gallop between two cones may not capture an examinee's true performance level. Perhaps, an alternative solution for this problem is to standardize instructions that are more applied to the examinee's real world experiences. Thus, instead of instructions that ask examinees to gallop, skip or slide between two cones, a test could

be developed with standardized instructions based on short stories that resemble examinees real word situations. The idea is to stay away from artificial environment while keeping a certain level of standardization with regards to instruction and environment.

If formal tests fail in allowing for observation in more natural settings the same can not be said for informal or non-standardized tests. The lack of standardized constrains in informal tests offers examiners more flexibility in determining the actual skills of the persons being tested and this is considered the main advantage of these types of tests ((Burton & Miller, 1998). In such cases more trials are permitted and consequently longer observations are required. Thus, the time required for test administrations in informal assessments is generally longer compared to formal tests. A possible solution for this problem is to use informal tests more closely integrated with instruction, thus saving time that is spent on testing itself (Nitko, 2001). In addition to the time, directions for informal tests may be too vague for examiners to understand and information concerning validity and reliability is not provided in informal tests (Burton & Miller, 1998).

In short, tests can be either formal or informal. Advantages and disadvantages are associated with each test category. Possible solutions for some of the shortcomings of both formal and informal tests were provided.

6.4.4 Norm-Referenced Approach vs. Domain-Referenced Approach

The two most common approaches used in test development within the field of physical education are: Norm-Referenced Approach (NRA) and Criterion-Reference Approach (CRA). The current assessment tool will be developed using the CRA. The main focus of domain-referenced tests is on *what* test takers can do and what they *know*, not on how they compare with others (Anastasi & Urbina, 1997). Therefore, the current test is characterized mainly for use to

track individuals' performance over time rather than determine individual comparisons. In the following sections the differences between NRA and CRA will be reviewed. The intended purposes and results interpretation, as well as the scoring process which defines how the test results must be interpreted are discussed. Closing this section is a discussion on how domain-referenced tests may prevail over norm-referenced tests allowing for a more complete understanding of students performance on a given task.

6.4.5 Intended purposes and interpretations

Tests developed under NRA or CRA differ in terms of their purposes and thus allow for different types of interpretation of student performance. There are some occasions in which teachers want to compare students' performances (between-individual comparison). For example, if a teacher wants to know whether Oldemar is better than Fabio on the performance of the fundamental movement skill of kicking or a group of skills (e.g., manipulative skills), a test developed under the NRA is preferred. In addition to between-individual comparisons one might want to interpret assessment results in relation to the performance of a large group of similar students (norm group) – similar age, sex, etc., - who happened to take the same assessment. Here is another example that illustrates this situation. Maggie has received a percentile rank score of 34 on the total test score of the TGMD (Ulrich, 2000). This means that she performed as well or better than 34% of the students in the normative group. This type of information can be useful for deciding whether or not a student needs remedial assistance or is a candidate for a gifted program (Nitko, 2001). However this score gives little information about Maggie's deficiencies that she needs to improve in the performance of gross motor skills which are the focus of the TGMD. If one wants to know more about Maggie's performance, then a test developed under the CRA is preferred

(see section 2.4.4.3 for more details on this). In short, tests might have different purposes and allow for different interpretation based on the approach used in its construction. The next section deals with the different types of scores used in norm-referenced and domain-referenced tests that allows for interpretation of test results.

6.4.6 Scores and their interpretation

Norm-Referenced Approach and CRA differ also in terms of the scores used for interpretation of students' assessment results. A raw score is the most common one and refers to the number of points gained by students taking a given test (Nitko, 2001). Very little can be said about the meaning of raw scores and for this reason they are rarely used for interpretation of students' results. To overcome the shortcomings of raw scores both NRA and CRA use alternative scores making interpretation of test results more meaningful. In norm-referenced assessment tools performance scores (raw scores) are transformed into relative scores (percentile ranks, linear standard score, grade-equivalent score, etc.). The scores under this concept may be used to tell the percentage of individuals in a norm group scoring below a particular raw score (percentile rank); or to inform the location of a particular score in relation to the mean and standard deviation of a norm group (linear standard score); or to determine the grade placement at which a particular individual raw score is the average for a norm group - grade-equivalent score - (Nitko, 2001). Unlike the NRA the CRA uses absolute rather than relative scores to overcome the problems associated with raw scores. Nitko (2001) cites percentages, speed of performance, and quality ratings as some examples of such scores. A *Percentage score* is simply a number that indicates the percentage of maximum points earned by an individual in a given test. If a test has 10 items and a particular student gets 7 items right, then the student has 70% of the items correct.

The second type of score is *speed of performance* and refers to the amount of time an individual takes to complete a task or a number of tasks in a fixed amount of time. Running a marathon (about 43 miles) in 3 hours is an example of this type of score. The third score often used in domain-referenced tests is *quality ratings* which is basically the quality level at which an individual performs ('Excellent', 'Good' or Needs to Work; or 'Beginner Level', 'Intermediate Level' or 'Advanced Level'). The Fundamental Movement Pattern Assessment Instrument (Gallahue & Donnelly, 2003) is a domain-referenced instrument that uses quality rating scores. After being tested individuals are placed under one of the following stages of development (quality ratings): Initial Stage, Elementary Stage, Mature Stage or Sport Skill Stage. As seen herein, NRA and CRA use different types of scores instead of simple raw scores to make interpretation of test results more meaningful. Next, a discussion on how, in some cases, domain-referenced test may prevail over norm-referenced tests is presented.

6.4.7 Beyond norm-referenced approach

It is important to note that both NRA and CRA are important if the goal is to understand levels of learning of students (Nitko, 2001). However, there are some aspects embedded in the criterion-referenced tests that allows for a more complete understanding of students performance on a give task. These aspects that favor criterion-referenced over norm-referenced tests are discussed next.

First, to fully interpret a student's score from a given test one needs more than just percentile ranks or other types of relative scores which are the core of norm-referenced tests. When there is a need for absolute interpretation of students' assessment results, the domain-referenced tests are preferred. A good example is when a teacher wants to know which of the fundamental movement skills students are having trouble mastering. Rank order scores may be

helpful selecting the best performers in locomotive movement skills; however, if the ultimate goal is to plan appropriate instruction, or track students improvement over a period of time one needs to be aware of the kinds of performances each student can do or the specific kinds of difficulties each student is experiencing in each skill tested (Nitko, 2001). Thus, a test that allows a teacher to infer about one's status in relation to the domain of fundamental movement skills, rather than simply in relation to how well others students performed, is necessary.

Second, unlike norm-referenced tests, criterion-referenced tests happen to have a close relationship with developmentally-based physical education programs. This is a term used by Gallahue and Donnelly (2003) to describe programs in physical education that emphasize the concept of individual differences. These programs should be distinguished from those that rely on pre-established content and assessment procedures that take into consideration students' age and/or grade. The authors also claim that since norm-referenced tests are based on age appropriateness they are only of secondary importance to developmental physical education.

In developmental physical education programs, assessments are often carried out at the beginning of each teaching session for verification of a students' readiness for a specific content that is going to be taught. Following this verification is the decision of what and how the content should be taught. For example, a physical education teacher is administering a test to verify levels of proficiency on the performance of FMS of 4th graders. Generally, students in 4th grade are able to perform a variety of FMS with a good degree of proficiency. However, because many factors (experience, physical and cognitive development, etc.) may affect ones performance on these FMS, all students may not perform as expected at 4th grade. If after administering the test the results showed that several students performed below what is expected for 4th graders the decision may be that rather than teaching more complex skills or combinations of FMS,

emphasis should be placed first on the mastery of individual FMS. This is a common practice in developmental physical education and tests developed under the CRA rather than NRA can better help teachers on this process of decision making.

In short, tests developed under both NRA and CRA are important sources of reference to track student learning; however, student results interpretation under criterion-referenced approach offer test examiners a better understanding of what students are able or not able to do when tested in a given task. This type of information is crucial for developmental physical education programs that place emphasis on the concept of individual differences. Also, the types of scores used for interpretation of students' results differ for the two types of approach presented here. Next section deals with the two approaches used to test, describe and interpret movement changes.

6.4.8 Approaches used to test, describe and interpret movement changes

The previous section provided a discussion about NRA and CRA which are two types of approaches used in the development of tests in movement behavior. This section discusses the two approaches used to test, describe and interpret movement behavior changes. These two approaches are product-oriented or process-oriented. The former is more interested on the final outcome of the performance; that is, the quantitative aspects of movement behavior (quantifiable outcomes). On the other hand, process-oriented assessments deal with testing, description and interpretation of the qualitative aspects (form) of movement behavior.

6.4.9 Product-Oriented Assessment:

Product-oriented instruments are focused on the final outcome of movement behavior (Anastasi & Urbina, 1997). The movement form or technique used to perform the task is generally of little interest for those administering process-oriented assessments (Payne & Isaacs, 1998). How many goals a child scores, how many miles an adult runs and the number of yards a ball is thrown are all examples of product-oriented scores. Product values are quantifiable measures of performance (Gabbard, 2004). And this is by far the simplest form of assessment. In part, this is because the data collection process does not require major device aids such as a video camera or similar tools to record movement behavior in order to help students to improve their skills. This is because examiners are only interested in the final outcome, thus making the process of data collection less complicated.

An example would be evaluating free-throw shootings in basketball. While using a product-oriented approach one would count the number of success shootings students complete in 30 seconds. Recording sheets and pencils are the only tools needed for data collection in the example of free-throw shooting. However, there are some cases in which the examiner wants to collect more detailed information about an individual's movement behavior. If this is the case then process-oriented assessments are recommended.

6.4.10 Process-Oriented Assessment

Instruments developed to assess the quality aspects of movement behavior are called process-oriented assessments. Examiners using these kinds of tests are concerned with the technique or form of the behavior used to perform movement tasks (Payne & Isaacs, 1998). Much of the literature that deals with measurement and evaluation of changes in movement behavior, especially fundamental movement skills, is based on qualitative descriptions (Gabbard, 2004).

Administering such tests requires a greater deal of work from the examiner compared to product-oriented assessments. Usually, more trials are necessary for observation of the quality aspects of the performance in process-oriented assessments, which in turn takes more time for data collection when compared to product-oriented tests. The performance is frequently videotaped to increase accuracy of process-oriented tests. This process is useful when there is a need for more accurate and detailed movement behavior analysis of performers. Performance of rudimentary and fundamental movement skills are often assessed with use of process-oriented assessment tools. These skills call for qualitative type of data since one is interested on the quality/form of the movement performed (Gabbard, 2004). Next, the two approaches under which movement behavior, especially fundamental movement skills, can be qualitatively analyzed are discussed.

6.4.11 Total Body Approach (TBA) vs. Body Component Approach (BCA)

For years motor development specialists have been investigating ways to describe changes in the process of movement behavior. Two approaches have emerged from these studies: Total Body Approach and Body Component Approach. The former was developed to describe changes in the configuration of the total body; whereas, the intent of the latter is to describe changes in the configuration of body parts (Painter, 1994). Each one of these approaches is different in the way they describe how individuals move from one level of skill performance to the next.

Body Component Approach: Mary Ann Roberton devoted most of her career to study and support the BCA. The approach defines sequential changes in the configuration of body parts so that, for the same skill, a sequence describing, for example, arm action is differentiated from a sequence describing leg action (Painter, 1994). The approach was introduced to the study of motor skill development in 1977 when Roberton used two sets of body component categories

to describe the overhand throw for force (one for arm action and the other for pelvic-spinal action). The findings pointed to the premise that development of the two components appeared to occur at different rates. Subsequent to the 1977 study, more studies were conducted for different skills that expanded to examine three or more body components.

Total Body Approach: much of the research supporting the TBA has been done by Seefeldt and his colleagues at Michigan State University. The body in this approach is viewed as a whole when changes in fundamental movement skills are analyzed. Seefeldt and his colleagues (Branta, Seefeldt, & Haubenstricker, 1984; Seefeldt & Haubenstricker, 1982) have identified developmental sequences for several fundamental movement skills based on this approach such that each level, describes the movement of the arms, legs, trunk, and head in a composite format. These researchers admit that they do not believe that actions of all body parts develop as a unit (lockstep fashion), but argue that “there is sufficient cohesion among certain characteristics of a pattern to define those as stages of development” (Branta et al., 1984, p. 470). In turn, such an approach could be used to describe changes in movement patterns and to serve as a guide in motor development assessment (Painter, 1994). Perhaps, the biggest advantage of such an approach to describing movement changes is the feasibility of observing these changes even without video recording equipment. With proper training, examiners can reach a reasonable percentage of agreement when observing such changes in regular physical education settings. This is far more complicated when tests developed with the BCA are employed. This is because one needs to concentrate on specific body parts (e.g., legs, trunk, head, etc) at the time of observation. This would take considerable effort and time to be done and for this reason is practically unviable.

The current assessment tool is being developed under the criterion-referenced approach with process-oriented characteristics. It will allow for collection of qualitative type of data which will reflect the domains of “movement concepts” and fundamental movement skills. For this reason, between-individuals comparisons will not be possible. It will allow, though, interpretations that compare students with themselves in relation to pre-established criteria from the two domains under investigation.

6.5 TEST VALIDITY

According to the *American Educational Research Association*, the *American Psychological Association* and the *National Council on Measurement in Education* (1999) “Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (AERA, APA, & NCME, 1999, p. 9). This is the most current definition of validity which leads to the current thinking in test development. In the past, test developers emphasized the existence of three types of validity: (1) content validity; (2) criterion-related validity and (3) construct validity. Further, other terms have been used including concurrent validity, face validity, predictive validity, etc. However, in general, these are specific procedures for evaluating validity, rather than new kinds of interpretative inferences for test scores (Messick, 1989). Currently, the *Standards for Educational and Psychological Testing* (AERA et al., 1999) suggests validity to be used as an unitary concept. Thus, there are sources of validity that support the intended interpretation of test scores for the proposed purpose of the test. “These sources of evidence may illuminate different aspects of validity, but they do not represent distinct types of validity” (AERA et al., 1999). They must be viewed interdependent and

complementary forms of validity evidence and not viewed as separate and substitutable validity types (Rudner & Schafer, 2002). These sources of validity evidence are based on: (1) test content; (2) response processes; (3) internal structure; (3) relation to other variables; and (5) consequences of testing. Of special interest for this research is the source of validity evidence based on test content.

6.5.1 Validity evidence based on test content

As it was discussed above, the current thinking in test development considers validity as a unitary concept. The term content validity was used in 1974 standards as one type of validity. Then, the term was changed to content-related evidence in the 1985 standards emphasizing that it referred to one type of evidence within a unitary conception of validity. In the current standards (1999), this type of evidence is characterized as “evidence based on test content” (AERA et al., 1999). In this context, the question being asked is whether the items proposed by the test developer do, in fact, constitute a representative sample of the wider domain about which one wishes to make inferences (Thorn & Deitz, 1989). In addition, items are judge based on their importance (Anastasi & Urbina, 1997). Importance/relevance focuses on whether assessment tasks or items are included in the test user’s domain which is defined by the test developer (Nitko, 2001).

The process of collecting evidence for content-related validity is an essential requirement of test development that begins with the purpose of the test being clearly stated (Haley, Coster, & Faas, 2001). Next, the content domain being measured is fully described (Thorn & Deitz, 1989). Concurrently with the content domain analysis is the development of the behavioral objectives (Berk, 1984). Each objective should reflect a specific area of the entire domain. Specific

literature is then researched and a pool of items is generated. Here, all major aspects of the domains are tapped in the correct proportion and items assigned to one of the behavioral objectives previously developed.

Because the process used to generate the item pool is not sufficient to provide content-related evidence, further item analysis is necessary. Thorn and Deitz (1989) cite two approaches to establish evidence for content-related validity once the items have been generated. The first approach uses empirical methods such as factor analysis and the use of statistics to measure item difficulty and discrimination. Perhaps the major problem of such an approach is that it can result in a final selection of items that is not representative of the domain being measure by the test (Hambleton et al., 1978 cited in Thorn & Deitz, 1989). Therefore, an approach based on judgment of content experts is usually preferred. In this approach, the quality of the items is judged by the extent to which each reflects, in terms of its content, the domain it intends to measure (Thorn & Deitz, 1989). This judgment is done by at least two content experts who rate the relevance and representativeness of items.

Another point to be considered when collecting evidence for content-related validity is that evidence should be gathered for each intended proposed interpretation and uses of test scores. No test is valid for all purposes or in all situations (AERA et al., 1999). Therefore, each recommend use or interpretation requires validation. One way of doing this is to ask the panel of content experts to rate the degree of the relationship between the set of items and each intended interpretation or use of the test being proposed.

6.5.2 Measuring content-related evidence through expert judgment

The most common procedure for collecting evidence for content-related validity is based on judgments done by experts regarding the content of the tasks or items on the instrument. These judgments are mainly based on content representativeness and relevance which focuses on whether the assessment tasks are a representative and/or relevant sample from a larger domain of performance (Nitko, 2001). The process of measuring content-related evidence through expert judgment involves two major steps: (1) the development of an item review form; and (2) the selection of the panel of content experts.

Once a set of items is generated by the developers of the test, the next step is to develop what is called *Item Review Form (IRF)*. The IRF is intended to collect information from content experts regarding quality of the set of items proposed by developers of the test (Nitko, 2001). Some aspects that may be included in the IRF are: (a) item importance/relevance; (b) item representativeness; (c) link between group of items and specific behavioral objectives (subdomains); (d) link between a set of items and the purpose(s) of a test; etc. The main content of an IRF depends mostly on the nature of the content being investigated. For example, there are cases in which including matching tasks in an IRF is unnecessary. Content experts use matching tasks to link each item proposed by the test developer with its respective domain. This is not the case of the current assessment tool. This is because there is little doubt, if any, that the skill of running belongs to the domain of fundamental movement skills, or that general space is a concept placed under the movement concept domain. The same is true for the remaining group of item proposed in the current study. However, in some domain areas this is not as clear, thus matching tasks are required to ensure that each item in the test is, in fact, measuring what it is intended to measure.

The layout of IRF is another issue to be addressed. There is no “right” and “wrong” in terms of the layout used in item review forms. Test developers use a layout that best fit their needs. Generally, content experts are asked to rate items using ordinal scales (e.g., Likert-type scales). Also, YES or NO-type questions are often included in these forms. Further, content experts are asked to provide written comments regarding items or a set of items (Haley et al., 2001). Written comments provided by raters are especially important in deciding whether a particular item is worth to be included on a test. Moreover, it is common asking content experts whether they feel it is necessary to include and/or remove items to/from the proposed list of items. Procedures commonly used to select the panel of judges are addressed next.

The process of selecting experts that rate the quality of proposed items is a very important step in collecting evidence for content-related validity. To rate the quality of items in their study, Haley et al. (2001) used a panel of judges that combined experts with theoretical/research-related backgrounds and experts with more practical/applied experience. This may be the best approach when selecting content experts. This combination may produce a more diversified set of data than if only either category of experts was used. This approach can yield a broader set of opinions that may better help test developers in the interpretation of data.

Also, it is important that the experts selected for the review of the items are not familiar with the test and did not participate in any previous item review study involving that particular test. This might cause potential judgment bias, thus jeopardizing the final selection of items (Haley et al., 2001). Further it is recommended that experts rate items independently. The interaction of experts rating a set of items may also cause problems (Thorn & Deitz, 1989). Some judges may be influenced by the opinion of another judge while rating the quality of items, thus causing potential bias.

The combination of the procedures described above does not guarantee experts will rate test items with quality. Because experts rate test items in the absence of test developers, it is impossible to ensure that item raters do their best to complete the task. The use of incentives may help control such variable (Thorn & Deitz, 1989). The authors recommended the names of the experts to appear in the test manual. Experts may work harder when they are acknowledged to this fact. Additionally, test developers may offer money in exchange to experts' voluntary participation.

In conclusion, test validity is an important step in test development; as well as, the process of gathering evidence for content-related validity. The use of a panel of content experts is the most common procedure for collecting such evidence. The collection of data is usually done through item review forms that are comprised of rating questions that assess the quality of the proposed set of items. Next, the methodology used in the current research is addressed.

REFERENCES

- AERA, APA, & NCME. (1999). *Standards for educational and psychological testing*. Washington D.C.: American Educational Research Association.
- Allison, P. C., & Barrett, K. R. (2000). *Constructing children's physical education experiences: understanding the content for teaching*. Needham Heights: Allyn and Bacon.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper Saddle River: Prentice-Hall.
- Berk, R. A. (1984). Conducting the item analysis. In R. A. Berk (Ed.), *A guide to criterion-referenced test construction*. Baltimore: The Johns Hopkins University Press.
- Branta, C., Seefeldt, V., & Haubenstricker, J. (1984). Age changes in motor skills during childhood and adolescence. In R. L. Terjung (Ed.), *Exercise and sport sciences review*. New York: Macmillan.
- Brodkin, P., & Weiss, M. R. (1990). Developmental differences in motivation for participating in competitive swimming. *Journal of Sport & Exercise Psychology*, 12, 248-263.
- Brustad, R. J., Babkes, M. L., & Smith, A. L. (2001). Youth in sport: sport considerations. In R. N. Singer, H. A. Hausenblas & C. M. Janelle (Eds.), *Handbook of sport psychology* (2nd ed.). New York: John Wiley & Sons, Inc.
- Burton, A. W., & Miller, D. E. (1998). *Movement skill assessment*. Champaign: Human Kinetics.
- CDC, C. f. D. a. P. (2004). *Prevalence of overweight among children and adolescents: United States, 1999-2000*, CDC, from <http://www.cdc.gov/nchs/products/pubs/pubd/hestats/overwght99.htm>
- Chepko, S., & Arnold, R. K. (2000). *Guidelines for physical education programs: grades standards, objectives, and assessments*. Needham Heights: Allyn and Bacon.
- Doherty, J., & Bailey, R. (2003). *Supporting physical development and physical education in the early years*. Buckingham: Open University Press.
- Gabbard, C. (2000). *Lifelong motor development* (3rd ed.). Needham Heights: Allyn & Bacon.
- Gabbard, C. (2004). *Lifelong motor development* (4th ed.). Needham Heights: Allyn & Bacon.
- Gallahue, D. L., & Donnelly, F. C. (2003). *Developmental physical education for all children* (4th ed.). Champaign, IL: Human Kinetics.

- Gallahue, D. L., & Ozmun, J. C. (1998). *Understanding motor development : infants, children, adolescents, adults* (4th ed.). Dubuque, Iowa: McGraw-Hill.
- Graham, G., A., H. H. S., & Parker, M. (1998). *Children moving: a reflective approach to teaching physical education*. Mountain View: Mayfield.
- Graham, G., Holt/Hale, S. A., & Parker, M. (2004). *Children moving: a reflective approach to teaching physical education* (6th ed.). Mountain View: Mayfield.
- Haley, S. M., Coster, W. J., & Faas, R. M. (2001). A content validity study of the Pediatric Evaluation of Disability Inventory. *Pediatric Physical Therapy*, 177-184.
- Holt/Hale, S. A. (1999). *Assessing motor skills in elementary physical education*. Oak Ridge: NASPE Publications.
- Human Kinetics (Organization), & Pettifor, B. (1999). *Physical education methods for classroom teachers*. Champaign, IL: Human Kinetics.
- Kogut, S. P., & National Association for Sport and Physical Education. (2003). *Beyond activities : learning experiences to support the National Physical Education Standards : elementary*. Reston, VA: National Association for Sport and Physical Education.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.). New York: American Council on Education/Macmillian Publishing Co.
- NASPE. (2004). *Moving into the future: National standards for physical education* (2nd ed.). Reston: McGraw Hill.
- Nikolay, P., Grady, S., & Stefonek, T. (1997). *Wisconsin's model academic standards for physical education*. Unpublished manuscript, Madison.
- Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Columbus: Prentice-Hall.
- Painter, M. A. (1994). Developmental sequences for hopping as assessment instruments: a generalizability analysis. *Research Quarterly for Exercise and Sport*, 65(1), 1-110.
- Payne, G., & Isaacs, L. D. (1998). *Human motor development: a lifespan approach*. Mountain View: Mayfield Publishing Company.
- Rudner, L. M., & Schafer, W. D. (2002). *What teachers need to know about assessment*. Washington DC: National Education Association.
- Sayre, N. E., & Gallagher, J. D. (2001). *The young child and the environment: issues related to health, nutrition, safety, and physical activity*. (1st ed.). Needham Heights: Allyn & Bacon.

- Seefeldt, V., & Haubenstricker, J. (1982). Patterns, phases, or stages: an analytical model for the study of developmental movement. In J. A. S. Kelso & J. E. Clark (Eds.), *The development of movement control and co-ordination*. Chichester: John Wiley & Sons.
- Thomas, K. T., Lee, A. M., & Thomas, J. R. (2003). *Physical education methods for elementary teachers : 70 proven lesson plans* (2nd ed.). Champaign, IL: Human Kinetics.
- Thorn, D. W., & Deitz, J. C. (1989). Examining content validity through the use of content experts. *The Occupational Therapy Journal of Research*, 9(6), 334-347.
- Ulrich, D. A. (1985). *Test of gross motor development*. Austin: Pro-ed.
- Ulrich, D. A. (2000). *Test of gross motor development-second edition* (2nd ed.). Austin: Pro-ed.
- Wickstrom, R. L. (1983). *Fundamental motor patterns* (3rd ed.). Philadelphia: Lea & Febiger.
- Zhu, W., & Cole, E. L. (1996). Many-faceted rasch calibration of a gross motor instrument. *Research Quarterly for Exercise and Sport*, 67(1), 24-34.